Exploring adaptation and divergence in Berthelot's pipit (*Anthus berthelotii*) through contemporary and historical population genomics

Eleanor Claire Sheppard

A thesis submitted for the degree of Doctor of Philosophy
University of East Anglia, UK
School of Biological Sciences

March 2024

Abstract

Understanding the interplay between demographic and environmental factors, including anthropogenic influences, in driving evolutionary changes within small, isolated natural populations is fundamental to conservation and evolution. This thesis aims to deepen our understanding of these dynamics through genome-wide analyses of variation, adaptation, and divergence across populations of Berthelot's pipit (*Anthus berthelotii*), an endemic island bird species. Host genetic variation was examined in relation to pathogen infection prevalence within and across populations. While a candidate gene approach targeting known immune genes yielded no association with avian pox infection, a genotype-environment association approach, combined with RAD-seq data, strongly linked avian pox prevalence to variation at specific genes involved in cellular stress signalling and immune responses. Also, avian malaria and pox prevalence were strongly correlated across island populations, underlining the need to study multiple pathogens simultaneously for a comprehensive understanding of host-pathogen evolution. Subsequently, a second genotype-environment association analysis focusing on avian malaria prevalence identified multiple novel candidate genes involved in various physiological processes, some of which were also identified in the pox analysis. These findings indicate that pathogen-mediated selection plays an important role in shaping genetic variation across populations, and they identify a range of new candidate genes with which to explore natural immunogenetic evolution. Next, whole-genome analyses, incorporating temporal comparisons using museum DNA, indicated range-wide reductions in genetic variation in the pipit over the 20th century, exacerbating population divergence across archipelagos, with pronounced effects in the smallest, most recently colonised, population. These results highlight the lasting impacts of historical founder events on genetic diversity, while also raising the possibility that recent environmental changes, including anthropogenic pressures, may have influenced genetic decline. Overall, these findings advance our understanding of how neutral and adaptive processes contribute to genomic evolution and divergence across natural populations.

Contents

List of tables and figures

Chapter contributions

This thesis has resulted in the following manuscripts presented in Chapters 2, 3, and 4, reflecting collaborative efforts, with my significant contribution as the lead author. Two of the three data chapters have been published in peer-reviewed journals as of the current submission. The following section provides citations for each chapter, outlining my specific contributions and co-author information.

Chapter 2 | Genomic associations with poxvirus across divergent island populations in Berthelot's pipit

A version of this chapter has been published as:
Sheppard, E.C., Martin, C.A., Armstrong C., González-Quevedo, C., Illera, J.C., Suh, A., Spurgin, L.G. and Richardson D.S. (2022). Molecular Ecology, 31, 3154–3173. doi: 10.1111/mec.16461.

ECS and DSR conceived and designed the study with input from LGS and CAM. Fieldwork was undertaken by ECS, CAM, CA, CG-Q, JCI, LGS and DSR. CA and CG-Q developed the genetic datasets. ECS performed data analyses and wrote the manuscript with supervision from DSR, LGS and AS.

Chapter 3 | Genotype-environment associations reveal genes potentially linked to avian malaria infection in populations of an endemic island bird

A version of this chapter has been published as:
Sheppard, E.C., Martin, C.A., Armstrong C., González-Quevedo, C., Illera, J.C., Suh, A., Spurgin, L.G. and Richardson D.S. (2024). Molecular Ecology. e17329. doi: 10.1111/mec.17329.

ECS conceived and designed the study with input from DSR and CAM. Fieldwork and molecular laboratory work were carried out by ECS, CAM, CA, CG-Q, JCI, LGS, and DSR. CA contributed restriction site-associated DNA sequencing data. ECS conducted all analyses and wrote the manuscript, receiving input and advice from DSR and AS.

Chapter 4 | Genomic erosion in oceanic island populations of an endemic bird: A spatio-temporal analysis

Sheppard, E.C., Martin, C.A., Illera, J.C., Emerson, B., Suh, A., Irestedt, M., Blom, M.P.K., Scott, M.F. and Richardson D.S.
In preparation.

ECS and DSR conceived and designed the study with input from AS and BE. Field sampling was conducted by JCI. DNA extractions and quality controls were performed by ECS, CAM and MI. ECS carried out bioinformatics and genomic analyses with guidance from MFS and MPKB. ECS wrote the manuscript with input from DSR and MFS.

Acknowledgements

I would first like to thank David Richardson for your continual guidance and vision. You have proved an extraordinarily patient and supportive supervisor, with just the right amount of impatience that has pushed me toward the completion of my thesis. Thank you for your kindness and humour, it has seen me through long days in the field and desk-bound struggles during the pandemic. Alex Suh, as my secondary supervisor, I thank you for always finding time to chat, providing me with valuable contacts and always being enthusiastic about my research, despite transitioning between institutions and a busy schedule. A special thank you to Mike Scott, who kindly took on the role of supervisor to a stressed and time-pressed PhD student. Your willingness to share your expertise, offer bioinformatic guidance, and address methodological questions has been invaluable. I also thank Brent Emerson, my CASE supervisor from IPNA in Tenerife, for his helpful contributions and ideas.

My thanks also go to our lab group, both past and present – Sarah, Tom, Charli, Claudia, Mike, George, Chuen, Alessandro – who provided a supportive space for collaboration and learning. I have enjoyed our coffee breaks, trips and conferences together. A special mention to Claudia Martin, who let me pick her brains on all things pipits, genomics, and life's ups and downs. You have been a source of optimism, support and knowledge for which I owe so much. I am truly grateful that pipits brought us together. Additionally, I appreciate the guidance offered by Maria-Elena Mannarelli in the lab, as her assistance played a crucial role in preparing samples for additional sequencing.

Fieldwork to Madeira and Porto Santo was made possible by a Heredity fieldwork grant from the Genetics Society. I had the privilege of sharing my field trips with dedicated pipit herders and excellent company, including Claudia, Dave, Johnny DeCoriolis, and Tom Brown. Additional field assistance in Lanzarote and La Graciosa was provided by Jordy, Tim, Joe, Elana, and Beth, to whom I am grateful. I appreciate the logistical support and efforts in obtaining the necessary permits by Juan Carlos Illera. I also acknowledge the contributions of past pipit researchers, including Claudia, Claire Armstrong, Catalina González-Quevedo, Lewis Spurgin, and Juan Carlos, who collected the field samples used throughout this thesis.

Claire and Claudia also generated the RAD-seq dataset, and the initial re-sequenced genomes used in this thesis.

I would also like to thank Mozes (Moos) Blom at Museum für Naturkunde Berlin and Martin Irestedt at Stockholm University for their training, guidance and contributions to the fourth chapter of this thesis. Martin performed DNA extractions and preparations for historical genome sequencing, while Moos provided essential bioinformatic training and support. Special thanks to his whole group in Berlin, especially Filip Thorn, for their support, advice, and company during my collaborative visit. This visit was made possible by a Training Grant from the Genetics Society. I also extend my gratitude to numerous museum curators for their generous assistance in sourcing and sampling Berthelot's pipit specimens, in particular to the British Museum of Natural History (Hein Van Grouw and Mark Adams), American Museum of Natural History, New York (Paul Sweet), Naturalis Biodiversity Centre, Leiden (Pepijn Kamminga and Steven van der Mije), and National Museums Liverpool (Tony Parker).

A heartfelt thank you to my family for a lifetime of support and encouragement. I am especially grateful to my mum, whose love and care have been my anchor in the past few months, and to my sister, whose dedication and patience in reading through the entirety of this thesis are a testament to her generosity and support. Finally, to Jordy, who has tolerated the quirks of a semi-migratory partner for several years now – I thank you for making space for my passions and never placing limits on the ever-evolving individuals we are becoming.

Chapter 1

General introduction

Berthelot's pipit in Selvagem Grande

1.1. Evolutionary forces shaping genetic variation, divergence and adaptation

The exploration of genetic divergence and adaptation within and among populations is fundamental for unravelling their genetic dynamics and understanding their evolutionary history. These processes are driven by various microevolutionary forces (Ellegren & Galtier, 2016) and their interactions (Willi *et al.*, 2007; Tigano & Friesen, 2016), which govern allele frequencies within populations through neutral and adaptive mechanisms. Mutation (in its wider sense including recombinational processes) confers genetic novelty by generating new alleles (Lande, 1976; Wright, 2005), which occurs at a greater rate in larger populations where there are more individuals and a greater pool of genetic diversity (Wright, 2005). However, it is the combined actions of genetic drift, gene flow and selection that determine the distribution of these alleles within and among populations from one generation to the next. Genetic drift operates stochastically, reducing genetic variation by subsampling alleles from finite parental populations (Wright, 1931; Lande, 1976; Charlesworth, 2009). This process can lead to population divergence, as the fate of any given allele – whether it becomes fixed or lost from the gene pool – is likely to vary between populations. The magnitude of genetic drift is influenced by population size, with smaller populations more susceptible to losing variants from the gene pool due to fluctuations in allele frequencies between generations (Wright, 1931; Charlesworth, 2009). Conversely, gene flow – the transfer of genetic material as individuals migrate from one population to another and breed therein – has the potential to increase genetic variation by introducing novel alleles or reintroducing alleles that were previously lost (Wright, 1931; Slatkin, 1987). Gene flow can counteract divergence from drift and adaptation and homogenise genetic variation across populations (Slatkin, 1987; Lenormand, 2002), eventually leading to homogenised allele frequencies.

Selection acts locally on individuals based on functional differences between the alleles they carry, resulting in the differential survival and reproductive success of those individuals. Environmental conditions vary spatially, exerting distinct selection pressures, thus determining the adaptive landscape (Wright, 1949). Consequently, populations inhabiting different ecological niches diverge as they adapt to their local environments (Aguirre-Liguori *et al.*, 2019; Wadgymar *et al.*, 2022). Selection can take various forms, including positive

directional selection, which favours beneficial alleles that alter a trait in a given direction, potentially driving an allele towards fixation within the population (reviewed in Biswas & Akey, 2006; Vitti *et al.*, 2013). Conversely, purifying selection disfavours deleterious alleles, removing them from the population (reviewed in Vitti *et al.*, 2013). Finally, balancing selection refers to various selective mechanisms that result in the maintenance of variation at genomic sites, thereby contributing to the overall maintenance of genetic diversity within populations (reviewed in Charlesworth, 2006; Spurgin & Richardson, 2010). Among populations, balancing selection may reduce population differentiation at these sites, or if different populations experience distinct balancing selection regimes, wherein different sets of alleles are favoured in each population, balancing selection can lead to increased population differentiation (Brandt *et al.*, 2018; Dong *et al.*, 2023).

Across the genome, heterogeneity arises from the complex interplay of evolutionary forces, genetic mechanisms, and structural elements acting at any locus; consequently, diversity and divergence are differentially shaped across the genome (Nosil *et al.*, 2009; Nosil & Feder, 2012; Ravinet *et al.*, 2017). Selective pressures act on specific regions of the genome leading to localised patterns of genetic diversity and divergence. Linkage – the tendency of alleles in close physical proximity on the same chromosome to be inherited together – can amplify the impact of selection across the region, influencing allele frequencies at linked sites (Smith & Haigh, 1974). On the other hand, recombination describes the mechanism by which genetic material is exchanged between chromosomes during meiosis, thus disrupting linkage between alleles (Felsenstein, 1974; Barton & Charlesworth, 1998). Both linkage and recombination are therefore fundamental processes in shaping allele co-inheritance across loci, contributing to the formation of linkage disequilibrium (the non-random association between alleles at different loci) influencing the spread of advantageous or deleterious alleles. Furthermore, regions proximal to structural elements (e.g., centromeres and telomeres), often have distinct patterns of variation due to the differences in recombination rates and chromatin structure (Stapley *et al.*, 2017). All together, these evolutionary processes shape the genomic landscape, further shaping the genetic variation that reflects the complex history of adaptation, divergence, and genetic exchange within and between populations.

The nearly neutral theory of molecular evolution suggests that the majority of genetic variation within a population is selectively neutral, meaning it does not affect fitness (Kimura, 1968, 1991; King & Jukes, 1969). Within this framework, genetic drift predominantly governs molecular evolution for alleles at most loci across the genome, while selection acts at a few specific loci within any time frame. This theory therefore serves as a foundational concept for interpreting patterns of genetic diversity within and across populations, aiding in the distinction between neutral and selective forces in shaping genetic variation. Thus, population-level genome-wide data can offer insights into population demography (e.g., Nadachowska-Brzyska *et al.*, 2021), gene flow and genetic structure (e.g., Kersten *et al.*, 2021), allowing for the identification of regions subject to selection pressures and the disentanglement of adaptation from neutral divergence (e.g., Fabian *et al.*, 2012; Qu *et al.*, 2015; Carreras *et al.*, 2017; Clucas *et al.*, 2019; Talla *et al.*, 2019).

In small and isolated populations, the interplay of evolutionary forces on genetic dynamics becomes notably pronounced. These populations are particularly vulnerable to the effects of genetic drift, which reduces genetic diversity (Ellstrand & Elam, 1993; Frankham, 1996; Spielman *et al.*, 2004). This vulnerability is exacerbated in populations that have experienced bottlenecks – significant reductions in size (Nei *et al.*, 1975). Similarly, populations that newly colonise an area may experience similar effects, as they typically originate from a small number of founding individuals (i.e., founder effects) (Lande, 1976; Ramstad *et al.*, 2004; Hawley *et al.*, 2008), with more sequential founder events having a larger impact (Clegg *et al.*, 2002). The combined effects of limited gene flow and persistently small effective population sizes further intensify the impacts of genetic drift, leading to genomic erosion, reduced genetic diversity, and thus reductions in genetic health (reviewed in Bosse & van Loon, 2022). This decline in genetic diversity not only affects neutral variation, but potentially compromises the adaptive potential of these populations (Ouborg *et al.*, 1991; Bijlsma & Loeschcke, 2012). Inbreeding also tends to increase in small populations with reduced diversity, as the limited number of potential mates increases the likelihood of mating between closely related individuals (Lynch *et al.*, 1995). Increased homozygosity resulting from inbreeding exposes recessive deleterious alleles previously masked in heterozygous individuals and also results in the loss of any heterozygote advantage, leading

to detrimental effects on individual and population fitness (termed inbreeding depression) (Keller & Waller, 2002; Willi & Hoffmann, 2009; Åkesson *et al.*, 2016; Stoffel *et al.*, 2021).

In time, populations could recover demographically, and new mutations and gene flow from other populations may introduce new genetic variation (McEachern *et al*., 2011; Frankham, 2015; Jangjoo *et al*., 2016), but the complete reversal of allele frequency changes induced by genetic drift may theoretically require many generations (Nei *et al*., 1975). Indeed, drift debt – reflecting the disparity between the expected and observed genetic diversity in bottlenecked populations (Gilroy *et al*., 2017; Jackson *et al*., 2022; Pinto *et al*., 2023) – may persist until populations reach a mutation-drift equilibrium.

1.2. Pathogens as selective pressures

Pathogens – encompassing a diverse array of disease-causing agents such as viruses, bacteria, fungi, protists, and macroparasites – exert fitness costs on hosts by reducing reproductive output, either through host mortality or morbidity (Anderson & May, 1979; Daszak *et al.*, 2000). The perpetual interaction of hosts and pathogens drives a dynamic process of reciprocal adaptations, resulting in a co-evolutionary arms race (Dawkins & Krebs, 1979; Paterson *et al.*, 2010). Pathogens exert selective pressures favouring host genotypes that confer resistance (i.e., the ability to impede the pathogen's lifecycle and prevent infection) and/or tolerance (i.e., the ability to cope with infection with limited loss of fitness). Conversely, pathogens are driven to evolve strategies to evade host immune defences. The complex interplay of the different mechanisms involved in pathogen-mediated selection (Spurgin & Richardson, 2010) underscores the need for diverse studies investigating the evolutionary theories and basis of host-parasite interactions. While directional selection may favour alleles conferring the highest resistance/tolerance to prevalent pathogens, many immune loci exhibit higher polymorphism than expected given the genomic average (Hughes & Hughes, 1995; Meyer & Thomson, 2001; Lazarus *et al.*, 2002), suggesting the involvement of balancing selection in immune gene evolution. Generally, any selective mechanism that acts to maintain genetic variability in a population can be considered part of balancing selection, but three principal, non-exclusive, pathogen-mediated mechanisms have been proposed: heterozygote advantage (Doherty &

Zinkernagel, 1975), rare-allele advantage (Slade & McCallum, 1992), and fluctuating selection (Hill *et al.*, 1991). Briefly, heterozygote advantage allows individuals with diverse genotypes to better combat either individual pathogens and/or the diversity of pathogens they are exposed to (Doherty & Zinkernagel, 1975). Rare-allele advantage favours less common alleles that pathogens have not yet adapted to (Slade & McCallum, 1992). Finally, fluctuating selection, driven by spatial and temporal variation in selective pressures and trade-offs in the physiological mechanism hosts use to confer resistance/tolerance, maintains polymorphism across populations (Hill *et al.*, 1991).

A wide range of genes may contribute to an organism's ability to resist/tolerate pathogens. Classic immune genes like the vertebrate major histocompatibility complex (MHC) and toll-like receptors (TLRs) have been extensively studied in the context of pathogen-mediated selection (e.g., Tschirren *et al.*, 2011; Collin *et al.*, 2013; Grueber *et al.*, 2014; Quéméré *et al.*, 2015; Biedrzycka & Kloch, 2016; Kloch *et al.*, 2018). However, the advent of the genomics era (see below) has facilitated the identification of additional targets of such selection (e.g., Fumagalli *et al.*, 2011), broadening our understanding of the genetic basis and eco-evolutionary dynamics of host-pathogen interactions and resistance evolution, particularly crucial for non-model organisms where functional genetic information has been scarce (for examples, see Bourgeois *et al.*, 2017; Cornetti & Tschirren, 2020).

1.3. Population genomics in evolution and conservation

Amid the ongoing global decline in biodiversity, largely attributed to human-induced factors such as land use and climate change (Ceballos & Ehrlich, 2002; Butchart *et al.*, 2010; Barnosky *et al.*, 2011; Ceballos *et al.*, 2017), it is crucial to understand the level and distribution of genetic diversity within and between natural populations, and the role of different evolutionary forces in generating those patterns. This is particularly relevant in the context of small and/or fragmented populations, as genetic variation underpins adaptive ability, and thus the survival of individuals and the persistence of populations in the face of future challenges (Reed & Frankham, 2003; Frankham, 2005; Willi *et al.*, 2006; Charlesworth & Willis, 2009; Bozzuto *et al.*, 2019). Thus, population genetics/genomics plays an

increasingly relevant role in evolutionary biology, ecology and conservation (Rokas & Abbot, 2009; Frankham, 2010; Shafer *et al.*, 2015; Supple & Shapiro, 2018; Theissinger *et al.*, 2023).

Traditionally, population genetic studies widely relied on a limited set of markers – allozymes (enzyme polymorphisms), mitochondrial DNA, microsatellites (simple sequence repeats), or amplified fragment length polymorphisms – to investigate, for example, patterns of connectivity, population structure and differentiation (e.g., Hoelzel & Dover, 1991; Fabiani *et al.*, 2003; Boessenkool *et al.*, 2007), colonisation and demographic histories (e.g., Bryan *et al.*, 2005; Funk *et al.*, 2010; Spurgin, Wright, *et al.*, 2014) and heterozygosity and inbreeding (e.g., Eldridge *et al.*, 1999, 2005; Richardson *et al.*, 2004; Brouwer *et al.*, 2007). However, due to their differing mutational processes and evolutionary rates, some of these markers are less suitable for comparing neutral and adaptive regions (Brown *et al.*, 1979; Ellegren, 2000). Intermediate strategies, such as those using a small number of SNPs, have also been employed (e.g., Goossens *et al.*, 2016). However, as discussed above, the genomic landscape can be far from homogeneous, with different regions of the genome exhibiting varying levels and patterns of genetic diversity (Nosil & Feder, 2012; Seehausen *et al.*, 2014). Thus, while all of these markers offer valuable insights, their limited size and number relative to genome size result in low genomic resolution, which could compromise the reliability of genome-wide estimates and inferences (Allendorf *et al.*, 2010; Galla *et al.*, 2020). Moreover, in adaptive evolution studies, the use of a limited number of markers, as in candidate gene studies (e.g., Poelstra *et al.*, 2013), likely overlooks other relevant genomic regions.

In recent decades, progressive advancements in DNA sequencing technologies from traditional Sanger sequencing to high-throughput methods (next-generation sequencing, NGS) have been revolutionary. This shift has allowed rapid and parallel sequencing of millions of DNA fragments, significantly improving the efficiency and scale of the sequencing process (Mardis, 2008; Giani *et al.*, 2020). As a result, it is now feasible to survey patterns of variation (generally SNPs) at genome-wide resolution for tens to hundreds, or even thousands, of individuals. This could take the form of whole-genome resequencing or sequencing a subset of the genome with thousands of markers (reduced representation methods) (Mamanova *et al.*, 2010; Narum *et al.*, 2013; Andrews *et al.*, 2016). Generating whole-genome resequencing data for many individuals across populations can be

impractical, particularly in studies involving non-model organisms with large and complex genomes that do not have an available reference genome. In such cases, extensive sequencing efforts are required, leading to a preference for more cost-effective reduced representation methods like restriction site-associated DNA sequencing (RAD-seq) (Andrews *et al.*, 2016), despite their limitations (theoretically representing < 10% of the genome in most cases) (Lowry *et al.*, 2017). Nevertheless, technological advancements have continued to reduce sequencing costs (Hayden, 2014; van Dijk *et al.*, 2014; Satam *et al.*, 2023), making whole-genome resequencing and the generation of highly complete and contiguous reference genome assemblies more attainable, even for non-model species (Ekblom & Galindo, 2010; Ekblom & Wolf, 2014; Ellegren, 2014; Whibley *et al.*, 2021).

With a significant increase in the number of sites available for analysis, genome-wide datasets enhance the accuracy, statistical power, and inference for numerous metrics central to molecular conservation and evolution studies (Luikart *et al.*, 2003; Allendorf *et al.*, 2010; Kardos *et al.*, 2014; Miller & Coltman, 2014; Hoffmann *et al.*, 2015; Shafer *et al.*, 2015). Moreover, they enable the exploration of more complex ideas relating to demographic history and fitness. For example, identifying runs of homozygosity (ROH) – contiguous stretches of homozygous genotypes within an individual's genome – can provide insights into recent demographic history and inbreeding (Li & Durbin, 2011; Curik *et al.*, 2014). These genome-wide datasets can elucidate the consequences of reduced effective population size on genetic diversity and population fitness (e.g., Robinson *et al.*, 2019; Khan *et al.*, 2021), and provide valuable insights into the influences of different evolutionary forces on population dynamics and structure (e.g., Carreras *et al.*, 2017; Talla *et al.*, 2019). Perhaps one of the most significant advantages of genome-wide data is the ability to identify regions or loci subject to natural selection, providing valuable insights into adaptive evolution and the genetic basis of adaptation.

Numerous approaches have been developed to use genomic data to detect selection, identify genes involved in adaptation and reveal the causal agents of selection (reviewed in Stinchcombe & Hoekstra, 2008; Nadeau & Jiggins, 2010; Stapley *et al.*, 2010; Pardo-Diaz *et al.*, 2015). Forward genetic approaches (including genome-wide association studies (GWAS); reviewed in Korte & Farlow, 2013; Tam *et al.*, 2019), aim to identify loci associated with

variation in a target phenotype, (often already shown to have adaptive significance). Prior knowledge of the factors influencing phenotypic variation is therefore essential for these approaches. Leveraging genetic variation across the genome, GWAS statistically assess which genetic variants exhibit non-random associations with phenotypes. This enables exploration of the genetic architecture underlying complex traits across diverse wild populations (e.g., Santure *et al.*, 2013; Wenzel *et al.*, 2015; Hansson *et al.*, 2018).

Alternatively, reverse genetic approaches (i.e., genome scans), detect outlier loci across the genome (reviewed in Jensen *et al.*, 2016). Outlier loci refer to genetic variants that exhibit unusual patterns of variation compared to the background genome-wide distribution, suggesting they are under selective pressure. From these outliers, the relevant environmental variables and processes responsible for adaptation may be deduced. These methods are based on the underlying assumption that drift and gene flow cause genome-wide effects, whereas selection and recombination have more localised effects (Lewontin & Krakauer, 1973). Statistical tests for excess population differentiation (e.g., fixation index F_{ST}) or signatures of selective sweeps are employed to detect these outlier loci (Oleksyk *et al.*, 2010; Vitti *et al.*, 2013; Ma *et al.*, 2015). Such signatures can include: (i) localised changes in genetic diversity, (ii) shifts in the allele frequency spectrum (AFS; the distribution of polymorphisms), and (iii) patterns of linkage disequilibrium (LD; the non-random association between nearby variants) (Oleksyk *et al.*, 2010; Vitti *et al.*, 2013; Ma *et al.*, 2015). An inherent advantage of this agnostic approach is its lack of an *a priori* hypothesis-driven bias, thus allowing for the identification of unbiased targets of selection. Genome scans can therefore serve as powerful tools for identifying genomic regions undergoing selection within and among ecologically diverse populations (e.g., Fabian *et al.*, 2012; Qu *et al.*, 2015; Clucas *et al.*, 2019).

Genotype-environment association analyses (GEAs), also referred to as environmental association analyses or environmental GWAS, is a specific approach that screens for correlations between allele frequencies and environmental predictor variables presumed to drive selection (Manel *et al.*, 2010; Schoville *et al.*, 2012; Rellstab *et al.*, 2015). While both reverse genetic approaches and GEAs identify divergent loci, they operate on different principles and methodologies. GEAs can be considered a subtype of forward genetics,

resembling GWAS in concept, but focusing on environmental gradients rather than phenotypic traits. The advantage of GEAs lies in their ability to not only identify adaptive alleles, but also pinpoint the relevant selective pressure driving changes (e.g., Dudaniec *et al.*, 2018; Fraik *et al.*, 2020; Yadav *et al.*, 2021).

Ultimately, each approach discussed represents just one tool that we can apply to investigate adaptive variation and divergence. To accurately distinguish the genomic effects of adaptation from other evolutionary processes and mitigate false positives, a combination of approaches and experimental validation is often necessary (Ravinet *et al.*, 2017).

1.4. Temporal genomics: leveraging historical DNA

Population genomics has enormous potential to improve our understanding of evolution and evolutionary processes in wild populations. However, contemporary patterns of genomic variation reflect the outcomes of dynamic processes occurring over extended timescales, shaped by ecological changes, demographic events, and adaptive histories. Thus, considerably more power to discern and quantify such processes can come from leveraging samples from additional time points, known as temporal genomics (Habel *et al.*, 2014; Díez-del-Molino *et al.*, 2018; Jensen & Leigh, 2022; Snead & Clark, 2022; Clark *et al.*, 2023).

Obtaining reference samples for temporal studies of wild organisms presents challenges. Recurrent sampling of the same natural populations across different time points is logistically feasible for only a limited number of species (e.g., Shultz *et al.*, 2016; Davies *et al.*, 2021; Stoffel *et al.*, 2021), due to the immense investment of time and resources required to span multiple generations, especially in longer-lived species. Alternatively, temporal sampling from natural populations can be achieved with samples sourced from natural history and herbarium collections (Holmes *et al.*, 2016; Bieker & Martin, 2018), as well as archaeological and paleontological specimens (Hofman *et al.*, 2015; Orlando *et al.*, 2021). These diverse sources have long provided representative samples of historical populations for various applications in evolutionary biology (e.g., Primack *et al.*, 2004; Moritz *et al.*, 2008; Schroeder *et al.*, 2009), including genetic investigations employing a limited number of markers to identify changes in population size and genetic diversity (e.g.,

Groombridge *et al.*, 2000; Athrey *et al.*, 2012; Spurgin, Wright, *et al.*, 2014). While technically feasible for several decades (Paabo, 1989), retrieving genomic data from these sources was hindered by associated costs and DNA degradation issues, rendering it impractical. Recent advancements have transformed the field (reviewed in Burrell *et al.*, 2015; Billerman & Walsh, 2019; Raxworthy & Smith, 2021), enabling the recovery and sequencing of genomic DNA from these sources. NGS has greatly facilitated this transformation (Bi *et al.*, 2013), offering capabilities similar to traditional population genomics studies at single time points. Short-read NGS technology is well-suited for degraded DNA fragments (< 100–300 bp), efficiently reading through such fragments without significant loss of sequencing efficiency or accuracy (Mardis, 2013; Goodwin *et al.*, 2016). This has made it feasible to incorporate historical DNA (hDNA) from specimens preserved and stored in repositories such as museums (typically < 200 years old) or ancient DNA (aDNA) from much older naturally preserved samples into population genomic studies (Raxworthy & Smith, 2021), a field known as museomics. Commonly used sequencing methods for hDNA include sequence capture, RAD-seq, and hybridization restriction site-associated DNA sequencing (hyRADseq) (Burrell *et al.*, 2015). These cost-effective approaches enable targeted subsampling of the genome, especially valuable in the absence of a reference genome. However, there is a growing trend toward resequencing entire historical genomes, aligning with advancements in contemporary genomics (Mikheyev *et al.*, 2015; Van Der Valk *et al.*, 2019; Parejo *et al.*, 2020; Miranda *et al.*, 2021; Pinsky *et al.*, 2021; Sánchez-Barreiro *et al.*, 2021; Wu *et al.*, 2022; Wood *et al.*, 2023).

As a nascent field, best practices for generating, processing, and analysing hDNA are continuously evolving and being refined (Burrell *et al.*, 2015; Billerman & Walsh, 2019; Raxworthy & Smith, 2021; Irestedt *et al.*, 2022). Historical and ancient specimens typically have lower endogenous DNA content (due to contamination) and are susceptible to post-mortem damage and fragmentation (Lindahl, 1993; Allentoft *et al.*, 2012; Dabney *et al.*, 2013). These properties can be exacerbated by preservation methods (Zimmermann *et al.*, 2008), and often correlate with environmental conditions and specimen age (Allentoft *et al.*, 2012; Sawyer *et al.*, 2012; Weiß *et al.*, 2016). Advances in extraction methods (e.g., Damgaard *et al.*, 2015; Pinhasi *et al.*, 2015; Tsai *et al.*, 2020) and library preparation (e.g., Gansauge & Meyer, 2013; Carøe *et al.*, 2018), have improved success across various types of

specimens. Concurrently, protocols for minimising cross-contamination, including dedicated 'clean' laboratories for all pre-amplification work, have been essential to mitigate risk (Wandeler *et al.*, 2007; Irestedt *et al.*, 2022). The susceptibility of historical samples to post-mortem damage – like the deamination of cytosine to thymine or uracil and guanine to adenine over time (Sawyer *et al.*, 2012; Weiß *et al.*, 2016) – can lead to the introduction of artifact alleles. As part of additional quality control measures, enzymatic repair before sequencing can be employed (Briggs *et al.*, 2010; Bi *et al.*, 2013; McGaughran, 2020).

Importantly the low quantity and quality of hDNA has the potential to hinder downstream analyses if not appropriately addressed and accounted for. For example, post-mortem damage, short fragments, and sequencing errors can result in low coverage data, inaccurate base calls, and misalignments of reads to the reference genome. These issues introduce mapping biases and errors that can potentially inflate estimates of diversity and affect population genetic estimators (e.g., Axelsson *et al.*, 2008; Sánchez-Barreiro *et al.*, 2021). Furthermore, the difficulty of read alignment increases as sequence divergence between the study species and reference genome rises, exacerbating mapping issues (Prüfer *et al.*, 2010; Taron *et al.*, 2018). To address these challenges effectively, it is essential to implement robust strategies for data processing and filtering. Doing so is crucial to ensure the reliability of downstream analyses, particularly when comparing hDNA with contemporary DNA. Basic filtering techniques, such as coverage and missing data thresholds (Sánchez-Barreiro *et al.*, 2021), trimming read ends where base misincorporations are more likely, or filtering out base misincorporations entirely (i.e., excluding transitions), can be used (Schubert *et al.*, 2012; Jónsson *et al.*, 2013). Incorporating measures, such as rescaling base quality scores at positions likely affected by deamination (Jónsson *et al.*, 2013) and using likelihood approaches appropriate for low coverage data (Korneliussen *et al.*, 2014), can also contribute to refining the precision of museomics.

Despite the challenges inherent in using historical DNA (hDNA), the unique advantage of it lies in directly representing populations at specific time points, allowing for the detection and quantification of changes in diversity, adaptation and population structure (Habel *et al.*, 2014; Díez-del-Molino *et al.*, 2018; Jensen & Leigh, 2022; Clark *et al.*, 2023). The incorporation of genomic hDNA into population genomics enables the assessment of

demographic and adaptive processes over extended evolutionary timeframes, spanning decades to centuries, and tens to hundreds of generations. This temporal scope often greatly surpasses the limitations of field-based sampling in natural populations, capturing substantial environmental and anthropogenic changes, and population declines throughout the last century or two (Ceballos & Ehrlich, 2002; Venter *et al.*, 2016; Ceballos *et al.*, 2017; Otto, 2018). Thus, it can provide unique insights into population responses and resilience, and the evolutionary and ecological drivers of genetic variation, genomic erosion and adaptation. For example, recent studies have used museomic methods to uncover how historical population declines contribute to genomic erosion in small populations (Sánchez-Barreiro *et al.*, 2021), how habitat loss and fragmentation has led to genetic diversity loss (Gauthier *et al.*, 2020), and how some species maintain genomic stability despite exploitation (Pinsky *et al.*, 2021). While contemporary population genomics offer valuable insight into historical changes in effective population sizes and divergence times (methods reviewed in Beichman *et al.*, 2018), limitations in accuracy may arise due to uncertainties in mutation rates, generation times, small sample sizes and reliance on the coalescent process, and wide confidence estimates, particularly towards the more recent history (Ryman *et al.*, 2019; Nadachowska-Brzyska *et al.*, 2021). As a result, hDNA can provide the higher resolution needed to detect very recent changes and more direct estimates (Benham & Bowie, 2023). While in its early stages, most population genomic studies incorporating hDNA focus on evolutionary relationships and changes in neutral genetic diversity (Clark *et al.*, 2023). However, it is also possible to detect selection with increased power, following allele trajectories across time and space (e.g., Mikheyev *et al.*, 2015; Bi *et al.*, 2019; Kreiner *et al.*, 2022). In its simplest form, qualitative statements about allele frequency changes at genes known to underlie important ecological traits become possible. Such studies have shown honey bee (*Apis mellifera*) populations can rebound from introduced diseases through the rapid evolution of tolerance (Mikheyev *et al.*, 2015), genetic adaptation could not prevent the range collapse of an endemic species (*Tamias alpinus*; Bi *et al.*, 2019), and intensive agricultural environments have driven rapid adaptation in an agricultural weed, common waterhemp (*Amaranthus tuberculatus*; Kreiner *et al.*, 2022). In evolutionary and conservation population genomics studies, the integration of hDNA thus serves as a powerful tool for unravelling the intricacies of genetic variation and understanding the driving forces shaping populations over time.

1.5. Berthelot's pipit (*Anthus berthelotii*): Disentangling evolutionary complexities in small and isolated populations

Oceanic island systems – often referred to as 'natural laboratories' – may exhibit distinct evolutionary dynamics from continental settings, making them invaluable for investigating fundamental questions regarding evolutionary processes and divergence (reviewed in Emerson, 2002; Losos & Ricklefs, 2009; Warren *et al.*, 2015; Patiño *et al.*, 2017; Whittaker *et al.*, 2017). Aligned with island biogeography theory (MacArthur & Wilson, 1967), species diversity within these systems is largely constrained by island size and distance from continental land (Valente *et al.*, 2017). Initial colonisation events, occurring at varying timescales within the system, prompt divergence from mainland populations and populations among islands (Warren *et al.*, 2015). The geographically distinct and often isolated nature of island populations, facilitated by natural barriers like oceans, may constrain gene flow, promoting genetic divergence among island populations. Founder effects and subsequent genetic drift may play a pivotal role (Yamada *et al.*, 2012; Sendell-Price *et al.*, 2021), but adaptation to diverse niches can also contribute to population divergence (Funk *et al.*, 2016). Island systems are often characterised by multiple populations inhabiting heterogenous environments, shaped by geological variation and differing ecological communities. This variation gives rise to inter- and intra-island pressures, further driving evolutionary processes. Divergence processes within and among island populations can potentially lead to new species over time, and as such, islands often have high rates of endemism (Myers *et al.*, 2000).

Berthelot's pipit (*Anthus berthelotii)* (Figure 1.1)*,* a small, sedentary passerine, is endemic to three archipelagos of Macaronesia in the North Atlantic (Figure 1.2) – Madeira, Selvagens and the Canary Islands, where it inhabits all main islands and the largest islets. It originated from the tawny pipit (*Anthus campestris*) on mainland Africa (Illera *et al.*, 2007), with colonisation of the Canary Islands and divergence approximately two million years ago (Voelker, 1999; Martin *et al.*, 2023). Two subsequent independent northward expansions established populations in Madeira and the Selvagens *ca.* 50,000 and 8,500 years ago, respectively (Spurgin, Illera, *et al.*, 2014; Martin *et al.*, 2023). Populating islands of varying sizes, from less than a square kilometre (large islets) to more than 2,000 square kilometres

(Tenerife) (Florencio *et al.*, 2021), the pipit inhabits diverse open habitats. The species therefore has a restricted overall range but is relatively locally common where habitat is available. These factors closely align with contemporary population estimates, suggesting over 100,000 breeding pairs across all the Canary Islands, and 2,500–10,000 and *ca.* 300 breeding individuals in the archipelagos of Madeira and Selvagens respectively (Spurgin, Illera, *et al.*, 2014; Illera, 2020). Information on population trends and 'recent' historical population size estimates is very limited. The islands also vary in isolation, with distances from the closest neighbouring island ranging from less than 1 km (large islets) to more than 60 km (El Hierro) (Florencio *et al.*, 2021). Low genetic variation and significant differentiation in the most recently established archipelagos, Madeira and Selvagens, are consistent with population bottlenecks associated with the colonisation of these island groups (Spurgin, Illera, *et al.*, 2014; Armstrong *et al.*, 2018; Martin *et al.*, 2021, 2023) and a subsequent lack of gene flow thereafter (Illera *et al.*, 2007; Spurgin, Illera, *et al.*, 2014). Recent studies highlight the role genetic drift may play in shaping genetic and morphological variation across the pipit's range (Spurgin, Illera, *et al.*, 2014; González-Quevedo, Spurgin, *et al.*, 2015; Armstrong *et al.*, 2018).

Despite the loss of genetic variation associated with genetic bottlenecks, certain adaptive immune genes in Berthelot's pipit have demonstrated the retention (González-Quevedo, Phillips, *et al.*, 2015; González-Quevedo, Spurgin, *et al.*, 2015) or even rapid regeneration of variation (Spurgin *et al.*, 2011). Notably, there is substantial and consistent variation in pathogens among Berthelot's pipit populations (Spurgin *et al.*, 2012) and also at the landscape scale within some populations (e.g., Tenerife; González-Quevedo *et al.*, 2014). Furthermore, the overall distribution of pathogen species richness and prevalence largely aligns with biogeographical expectations (MacArthur & Wilson, 1967), with smaller and more isolated islands (from continental Africa) exhibiting lower pathogen diversity (Spurgin *et al.*, 2012). Hence it is likely that the patterns of pathogen-mediated selection are consistent over longer evolutionary time frames, and shape among population patterns of immune-related genetic variation in this species.

Figure 1.1. Berthelot's pipit study system. (**A**) An adult Berthelot's pipit ringed to allow individual identification and prevent re-sampling. (**B**) Retrieval of an individual pipit from a clap trap baited with mealworm (*Tenebrio molitor*) larvae. (**C**) An infected adult Berthelot's pipit with a visible avian pox lesion on the bill. (**D–E**) Varied climates/habitats across the Berthelot's pipit distribution, ranging from dry arid scrub to more humid, vegetated alpine zones.

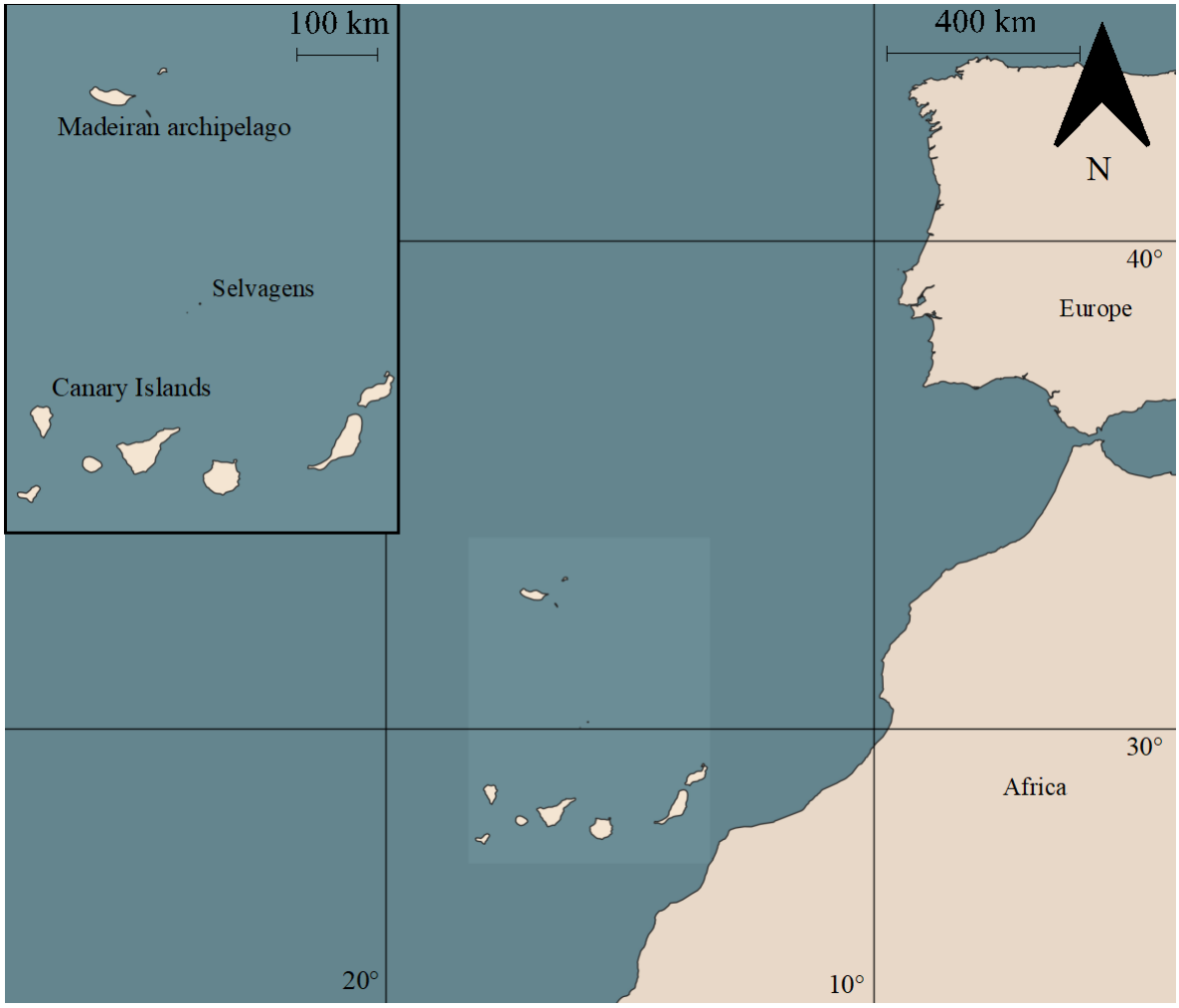

Figure 1.2. The geographic distribution of Berthelot's pipit across three archipelagos of Macaronesia in the north Atlantic. These archipelagos include the Madeiran archipelago, the Selvagens, and the Canary Islands (see inset).

Berthelot's pipit serves as a host for haemosporidian parasites causing avian malaria and malaria-like diseases (Illera *et al.*, 2008; Spurgin *et al.*, 2012), including strains of *Plasmodium* and *Leucocytozoon* , transmitted by dipteran vectors (Martínez-De La Puente *et al.* 2011; Njabo *et al.* 2011). Throughout its range, four strains of *Plasmodium* and four strains of *Leucocytozoon* have been identified in pipits (Spurgin *et al.*, 2012; Armstrong *et al.*, 2019). Additionally, pipits are infected by an endemic strain of avian pox (See Figure 1.1C), belonging to the genus *Avipoxvirus* (Illera *et al.*, 2008; Spurgin *et al.*, 2012), with transmission occurring through direct contact, dipteran vectors, or aerosol transmission (van Riper *et al.*, 2002). Infection by these pathogens may impose fitness costs, allowing only individuals of the highest quality to survive such infection, as evidenced by individuals infected by pox/malaria having significantly greater body mass than those without (Spurgin

et al., 2012). Individual-based associations between malaria infection and genetic variation (using RAD-seq and candidate genes) linked to immune genes have been identified within populations, and across populations of the Canary Islands (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2018, 2019).

Genome scans (using RAD-seq) have revealed divergent allele frequencies among archipelagos in genes related to craniofacial development, immune response, eye development, and metabolism, indicating responses to selection across Berthelot's pipit populations and archipelagos (Martin *et al.*, 2021). While some of these patterns may be linked to pathogens, the exact drivers of selection remain unclear. Ecological variations between islands also include habitat, climate, altitude and anthropogenic influences (Florencio *et al.*, 2021). Islands within archipelagos exhibit climatological diversity due to their oceanic location, varied topography, and altitudinal ranges (Florencio *et al.*, 2021). Eastern islands tend to be arid, rocky, low-lying and vegetated with xerophytic scrubs, whereas western islands are typically more humid, mountainous, and vegetated, featuring extensive laurel forests. Consequently, the pipit's preference for open habitat means it utilises rocky plains, grasslands, coastal scrubs, and even high alpine zones exceeding 2,000 m asl (Clarke, 2006).

The current human population distribution across the pipit's range varies significantly, ranging from nearly uninhabited (Selvagens, Desertas, and other islets) and seldom visited islands, to densely populated islands with more than 500 inhabitants per square kilometre (i.e., Gran Canaria) (Direção regional de estatística da Madeira, 2022; Instituto Canario de Estadística, 2023). The European settlement of Macaronesia was first documented in the early 15th century, although the Canary Islands were already settled by populations from Africa (Rando, 1999; Förster *et al.*, 2009; Fernández-Palacios *et al.*, 2016). These colonisations led to radical changes, including land transformation, intensified agriculture and the introduction of invasive species, resulting in significant alterations in native ecosystems (Morales *et al.*, 2009; del Arco Aguilar *et al.*, 2010; Illera *et al.*, 2012; Fernández-Palacios *et al.*, 2016; de Nascimento *et al.*, 2020). Furthermore, substantial human population growth has occurred throughout many of the islands during the early 20th century (Norder *et al.*, 2020). The subsequent rise in mass tourism, urban expansion, and

agricultural intensification during the 20th century likely contributed to extensive alterations in land use, particularly along coastal areas (Otto *et al.*, 2007). These anthropogenic changes over time and space may have impacted pipit population numbers and altered the pathogen fauna/intensity present. For instance, the level of malaria infection in Berthelot's pipit is influenced by anthropogenic factors, such as distance to artificial water reservoirs and poultry farms (González-Quevedo *et al.*, 2014). Furthermore, proximity to poultry farms was found to influence the presence of certain MHC alleles, some of which were significantly associated with malaria risk, suggesting a potential impact of human activities on fine-scale adaptive evolution in the species (González-Quevedo *et al.*, 2016).

Berthelot's pipit presents an intriguing system where populations exist at various stages of divergence, with discernible population differences despite relatively recent divergence. This offers a valuable opportunity to study the eco-evolutionary processes that act at an early stage in population divergence. The unique evolutionary landscape, coupled with potential human and pathogen-mediated impacts, makes it a valuable model to unravel the complexities of adaptation and divergence within the context of small and isolated populations. Previous research has explored the co-variation of malaria and genetic variation across spatial scales (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2018, 2019), relying on individual-based associations that hinge on accurate infection status assignment – a challenging task in wild populations due to factors like differential exposure of individuals to pathogens. In contrast, the study of avian pox as a selective agent in Berthelot's pipit has received limited attention, and its spatial variation underscores the importance of exploring its evolutionary significance. Incorporating a temporal dimension into analyses of neutral and adaptive variation among oceanic island populations is rare, leaving a gap in understanding recent demographic and environmental changes affecting current island biodiversity.

1.6. Thesis aims

Given the above, Berthelot's pipit – with its demographic history, insular habitat and distribution across varying ecological conditions – provides an excellent system in which to investigate the impact of geographical isolation, local adaptation, and recent historical

processes on the genomic makeup at the population level. My research therefore explores how neutral and adaptive genetic variation undergoes change over space and time within naturally small and isolated wild populations. Together, the empirical chapters employ a multifaceted approach, utilising a range of techniques from candidate gene to genomic methods, and integrate both contemporary and historical samples, providing a nuanced perspective that spans spatio-temporal dimensions and captures the complexity of population dynamics in this island-endemic species. The investigation aims to address key questions in the following chapters:

Chapter 2: This chapter aims to test for genetic variation associated with avian pox at both an individual-level (infection status) and population-level (prevalence) using candidate gene approaches and a genotype-environment association approach in conjunction with RAD-seq data. The goal was to identify candidate genes involved in Berthelot's pipit's response to pox.

Chapter 3: To disentangle the effects of both avian pox and malaria on genetic variation in Berthelot's pipits, this chapter explores the population-level relationship between the two pathogens. Using similar methods to Chapter 2, the aim was to then identify specific genomic loci linked to divergent and spatially variable selection across the species range in response to avian malaria.

Chapter 4: This chapter investigates genomic dynamics within Berthelot's pipit populations using whole-genome resequencing data obtained from historical and contemporary samples. The aim is to unravel spatio-temporal patterns in population structure, differentiation, and genetic diversity. These metrics provide insights into how the pipit's genome has evolved in response to potential demographic, environmental, and anthropogenic factors over the past century. The findings contribute to a deeper understanding of genomic processes in response to environmental challenges and population declines.

Chapter 5: This final chapter provides a brief general discussion of the new insights gained throughout the thesis. It addresses the broad implications of the findings and how they link

across chapters, their relevance to the current understanding of population genomics, and suggests potential directions for future research.

1.7. References

Aguirre-Liguori, J. A., Gaut, B. S., Jaramillo-Correa, J. P., Tenaillon, M. I., Montes-Hernández, S., García-Oliva, F., Hearne, S. J., *et al.* (2019). Divergence with gene flow is driven by local adaptation to temperature and soil phosphorus concentration in teosinte subspecies (*Zea mays parviglumis* and *Zea mays mexicana*). *Molecular Ecology*, *28*, 2814–2830.

Åkesson, M., Liberg, O., Sand, H., Wabakken, P., Bensch, S., & Flagstad, Ø. (2016). Genetic rescue in a severely inbred wolf population. *Molecular Ecology*, *25*, 4745–4756.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*, 697–709.

Allentoft, M. E., Collins, M., Harker, D., Haile, J., Oskam, C. L., Hale, M. L., Campos, P. F., *et al.* (2012). The half-life of DNA in bone: Measuring decay kinetics in 158 dated fossils. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 4724–4733.

Anderson, R. M., & May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, *280*, 361–367.

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.

Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution*, *9*, 12482–12502.

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

Athrey, G., Barr, K. R., Lance, R. F., & Leberg, P. L. (2012). Birds in space and time: genetic changes accompanying anthropogenic habitat fragmentation in the endangered black-capped vireo (*Vireo atricapilla*). *Evolutionary Applications, 5*, 540–552.

Axelsson, E., Willerslev, E., Gilbert, M. T. P., & Nielsen, R. (2008). The effect of ancient DNA damage on inferences of demographic histories. *Molecular Biology and Evolution*, *25*, 2181–2187.

Barnosky, A. D., Matzke, N., Tomiya, S., Wogan, G. O. U., Swartz, B., Quental, T. B., Marshall, C., *et al.* (2011). Has the Earth's sixth mass extinction already arrived? *Nature*, *471*, 51–57.

Barton, N. H., & Charlesworth, B. (1998). Why sex and recombination? *Science*, *281*, 1986–1990.

Beichman, A., Huerta-Sanchez, E., & Lohmueller, K. (2018). Using genomic data to infer historic population dynamics of non-model organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 433–456.

Benham, P. M., & Bowie, R. C. K. (2023). Natural history collections as a resource for conservation genomics: Understanding the past to preserve the future. *Journal of Heredity*, *114*, 367–384.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., *et al.* (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics*, *15*, e1008119.

Bi, K., Linderoth, T., Vanderpool, D., Good, J. M., Nielsen, R., & Moritz, C. (2013). Unlocking the vault: Next-generation museum population genomics. *Molecular Ecology*, *22*, 6018–6032.

Biedrzycka, A., & Kloch, A. (2016). Development of novel associations between MHC alleles and susceptibility to parasitic infections in an isolated population of an endangered mammal. *Infection, Genetics and Evolution*, *44*, 210–217.

Bieker, V. C., & Martin, M. D. (2018). Implications and future prospects for evolutionary analyses of DNA in historical herbarium collections. *Botany Letters*, *165*, 409–418.

Bijlsma, R., & Loeschcke, V. (2012). Genetic erosion impedes adaptive responses to stressful environments. *Evolutionary Applications*, *5*, 117–129.

Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: A review of methods, challenges, applications, and future directions. *Molecular Ecology Resources*, *19*, 1115–1130.

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, *22*, 437–446.

Boessenkool, S., Taylor, S. S., Tepolt, C. K., Komdeur, J., & Jamieson, I. G. (2007). Large mainland populations of South Island robins retain greater genetic diversity than offshore island refuges. *Conservation Genetics*, *8*, 705–714.

Bosse, M., & van Loon, S. (2022). Challenges in quantifying genome erosion for conservation. *Frontiers in Genetics*, *13*, 960958.

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.

Bozzuto, C., Biebach, I., Muff, S., Ives, A. R., & Keller, L. F. (2019). Inbreeding reduces long-term growth of Alpine ibex populations. *Nature Ecology and Evolution*, *3*, 1359–1364.

Brandt, D. Y. C., César, J., Goudet, J., & Meyer, D. (2018). The effect of balancing selection on population differentiation: A study with HLA genes. *G3: Genes, Genomes, Genetics*, *8*, 2805–2815.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, *38*, e87.

Brouwer, L., Komdeur, J., & Richardson, D. S. (2007). Heterozygosity–fitness correlations in a bottlenecked island species: a case study on the Seychelles warbler. *Molecular Ecology*, *16*, 3134–3144.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *76*, 1967–1971.

Bryan, M. B., Zalinski, D., Filcek, K. B., Libants, S., Li, W., & Scribner, K. T. (2005). Patterns of invasion and colonization of the sea lamprey (*Petromyzon marinus*) in North America as revealed by microsatellite genotypes. *Molecular Ecology*, *14*, 3757–3773.

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44.

Butchart, S. H. M., Walpole, M., Collen, B., Van Strien, A., Scharlemann, J. P. W., Almond, R. E. A., Baillie, J. E. M., *et al.* (2010). Global biodiversity: Indicators of recent declines. *Science*, *328*, 1164–1168.

Carøe, C., Gopalakrishnan, S., Vinner, L., Mak, S. S. T., Sinding, M. H. S., Samaniego, J. A., Wales, N., *et al.* (2018). Single-tube library preparation for degraded DNA. *Methods in Ecology and Evolution*, *9*, 410–419.

Carreras, C., Ordóñez, V., Zane, L., Kruschel, C., Nasto, I., MacPherson, E., & Pascual, M. (2017). Population genomics of an endemic Mediterranean fish: differentiation by fine scale dispersal and adaptation. *Scientific Reports, 7*, 43417.

Ceballos, G., & Ehrlich, P. R. (2002). Mammal population losses and the extinction crisis. *Science*, *296*, 904–907.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E6089–E6096.

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

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*, e64.

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

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., Sawkins, A., *et al.* (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources* (in press). Available at: https://doi.org/10.1111/1755-0998.13789.

Clarke, T. (2006). *A Field Guide to the Birds of the Atlantic Islands: Canary Islands, Madeira, Azores, Cape Verde*. London: Bloomsbury Publishing Plc.

Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 8127–8132.

Clucas, G. V., Lou, R. N., Therkildsen, N. O., & Kovach, A. I. (2019). Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evolutionary Applications*, *12*, 1971–1987.

Collin, H., Burri, R., Comtesse, F., & Fumagalli, L. (2013). Combining molecular evolution and environmental genomics to unravel adaptive processes of MHC class IIB diversity in European minnows (*Phoxinus phoxinus*). *Ecology and Evolution*, *3*, 2568–2585.

Cornetti, L., & Tschirren, B. (2020). Combining genome-wide association study and FST-based approaches to identify targets of *Borrelia*-mediated selection in natural rodent hosts. *Molecular Ecology*, *29*, 1386–1397.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*, 26–34.

Dabney, J., Meyer, M., & Pääbo, S. (2013). Ancient DNA damage. *Cold Spring Harbor Perspectives in Biology*, *5*, a012567.

Damgaard, P. B., Margaryan, A., Schroeder, H., Orlando, L., Willerslev, E., & Allentoft, M. E. (2015). Improving access to endogenous DNA in ancient bones and teeth. *Scientific Reports*, *5*, 11184.

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, *287*, 443–449.

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

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

de Nascimento, L., Nogué, S., Naranjo-Cigala, A., Criado, C., McGlone, M., Fernández-Palacios, E., & Fernández-Palacios, J. M. (2020). Human impact and ecological changes during prehistoric settlement on the Canary Islands. *Quaternary Science Reviews*, *239*, 106332.

del Arco Aguilar, M. J., González-González, R., Garzón-Machado, V., & Pizarro-Hernández, B. (2010). Actual and potential natural vegetation on the Canary Islands and its conservation status. *Biodiversity and Conservation*, *19*, 3089–3140.

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying temporal genomic erosion in endangered species. *Trends in Ecology & Evolution*, *33*, 176–185.

Direção regional de estatística da Madeira. (2022). *Estatísticas Demográficas da Região Autónoma da Madeira*.

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

Dong, S., Zhang, B., Huang, K., Ying, M., Yan, J., Niu, F., Hu, H., *et al.* (2023). Balancing selection shapes population differentiation of major histocompatibility complex genes in wild golden snub-nosed monkeys. *Current Zoology* (in press), Available at: https://doi.org/10.1093/cz/zoad043.

Dudaniec, R. Y., Yong, C. J., Lancaster, L. T., Svensson, E. I., & Hansson, B. (2018). Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*). *Molecular Ecology*, *27*, 2576–2593.

Ekblom, R., & Galindo, J. (2010). Applications of next generation sequencing in molecular ecology of non-model organisms. *Heredity*, *107*, 1–15.

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*, 1026–1042.

Eldridge, M. D. B., King, J. M., Loupis, A. K., Spencer, P. B. S., Taylor, A. C., Pope, L. C., & Hall, G. P. (1999). Unprecedented low levels of genetic variation and inbreeding depression in an island population of the black-footed rock-wallaby. *Conservation Biology*, *13*, 531–541.

Eldridge, M. D. B., Kinnear, J. E., Zenger, K. R., McKenzie, L. M., & Spencer, P. B. S. (2005). Genetic diversity in remnant and 'pristine' island populations of three endemic Australian macropods (Marsupalia): *Macropus eugenii*, *Lagorchestes hirsutus* and *Petrogale lateralis*. *Conservation Genetics*, *5*, 325–338.

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

Ellegren, H. (2014). Genome sequencing and population genomics in non-model organisms. *Trends in Ecology and Evolution*, *29*, 51–63.

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

Ellstrand, N. C., & Elam, D. R. (1993). Population genetic consequences of small population size: implications for plant conservation. *Annual Review of Ecology and Systematics*, *24*, 217–242.

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

Fabian, D. K., Kapun, M., Nolte, V., Kofler, R., Schmidt, P. S., Schlötterer, C., & Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology*, *21*, 4748–4769.

Fabiani, A., Hoelzel, A. R., Galimberti, F., & Muelbert, M. M. C. (2003). Long-range paternal gene flow in the southern elephant seal. *Science*, *299*, 676.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics*, *78*, 737–756.

Fernández-Palacios, J. M., Nogué, S., Criado, C., Connor, S., Góis-Marques, C., Sequeira, M., & de Nascimento, L. (2016). Climate change and human impact in Macaronesia. *Past Global Change Magazine*, *24*, 68–69.

Florencio, M., Patiño, J., Nogué, S., Traveset, A., Borges, P. A. V., Schaefer, H., Amorim, I. R., *et al.* (2021). Macaronesia as a fruitful arena for ecology, evolution, and conservation biology. *Frontiers in Ecology and Evolution*, *9*, 718169.

Förster, D. W., GündÜz, I., Nunes, A. C., Gabriel, S., Ramalhinho, M. G., Mathias, M. L., Britton-Davidian, J., *et al.* (2009). Molecular insights into the colonization and

chromosomal diversification of Madeiran house mice. *Molecular Ecology*, *18*, 4477–4494.

Fraik, A. K., Margres, M. J., Epstein, B., Barbosa, S., Jones, M., Hendricks, S., Schönfeld, B., *et al.* (2020). Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *Evolution*, *74*, 1392–1408.

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

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*, 131–140.

Frankham, R. (2010). Challenges and opportunities of genetic approaches to biological conservation. *Biological Conservation*, *143*, 1919–1927.

Frankham, R. (2015). Genetic rescue of small inbred populations: meta-analysis reveals large and consistent benefits of gene flow. *Molecular Ecology*, *24*, 2610–2618.

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.

Funk, W. C., Forsman, E. D., Johnson, M., Mullins, T. D., & Haig, S. M. (2010). Evidence for recent population bottlenecks in northern spotted owls (*Strix occidentalis caurina*). *Conservation Genetics*, *11*, 1013–1021.

Funk, W. C., Lovich, R. E., Hohenlohe, P. A., Hofman, C. A., Morrison, S. A., Sillett, T. S., Ghalambor, C. K., *et al.* (2016). Adaptive divergence despite strong genetic drift: genomic analysis of the evolutionary mechanisms causing genetic differentiation in the island fox (*Urocyon littoralis*). *Molecular Ecology*, *25*, 2176–2194.

Galla, S. J., Moraga, R., Brown, L., Cleland, S., Hoeppner, M. P., Maloney, R. F., Richardson, A., *et al.* (2020). A comparison of pedigree, genetic and genomic estimates of relatedness for informing pairing decisions in two critically endangered birds: implications for conservation breeding programmes worldwide. *Evolutionary Applications*, *13*, 991–1008.

Gansauge, M. T., & Meyer, M. (2013). Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. *Nature Protocols*, *8*, 737–748.

Gauthier, J., Pajkovic, M., Neuenschwander, S., Kaila, L., Schmid, S., Orlando, L., & Alvarez, N. (2020). Museomics identifies genetic erosion in two butterfly species across the 20th century in Finland. *Molecular Ecology Resources*, *20*, 1191–1205.

Giani, A. M., Gallo, G. R., Gianfranceschi, L., & Formenti, G. (2020). Long walk to genomics: History and current approaches to genome sequencing and assembly. *Computational and Structural Biotechnology Journal*, *18*, 9–19.

Gilroy, D. L., Phillips, K. P., Richardson, D. S., & van Oosterhout, C. (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., 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., 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. (2015). 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. (2015). Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology*, *24*, 5852–5863.

Goodwin, S., McPherson, J. D., & McCombie, W. R. (2016). Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics, 17*, 333–351.

Goossens, B., Sharma, R., Othman, N., Kun-Rodrigues, C., Sakon, R., Ancrenaz, M., Ambu, L. N., *et al.* (2016). Habitat fragmentation and genetic diversity in natural populations of the Bornean elephant: Implications for conservation. *Biological Conservation*, *196*, 80–92.

Groombridge, J. J., Jones, C. G., Bruford, M. W., & Nichols, R. A. (2000). 'Ghost' alleles of the Mauritius kestrel. *Nature*, *403*, 616.

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.

Habel, J. C., Husemann, M., Finger, A., Danley, P. D., & Zachos, F. E. (2014). The relevance of time series in molecular ecology and conservation biology. *Biological Reviews*, *89*, 484–492.

Hansson, B., Sigeman, H., Stervander, M., Tarka, M., Ponnikas, S., Strandh, M., Westerdahl, H., *et al.* (2018). Contrasting results from GWAS and QTL mapping on wing length in great reed warblers. *Molecular Ecology Resources*, *18*, 867–876.

Hawley, D. M., Briggs, J., Dhondt, A. A., & Lovette, I. J. (2008). Reconciling molecular signatures across markers: mitochondrial DNA confirms founder effect in invasive North American house finches (*Carpodacus mexicanus*). *Conservation Genetics*, *9*, 637–643.

Hayden, E. C. (2014). The $ 1,000 genome. *Nature*, *507*, 294–295.

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

Hoelzel, A. R., & Dover, G. A. (1991). Genetic differentiation between sympatric killer whale populations. *Heredity*, *66*, 191–195.

Hoffmann, A., Griffin, P., Dillon, S., Catullo, R., Rane, R., Byrne, M., Jordan, R., *et al.* (2015). A framework for incorporating evolutionary genomics into biodiversity conservation and management. *Climate Change Responses*, *2*, 1.

Hofman, C. A., Rick, T. C., Fleischer, R. C., & Maldonado, J. E. (2015). Conservation archaeogenomics: ancient DNA and biodiversity in the Anthropocene. *Trends in Ecology & Evolution*, *30*, 540–549.

Holmes, M. W., Hammond, T. T., Wogan, G. O. U., Walsh, R. E., Labarbera, K., Wommack, E. A., Martins, F. M., *et al.* (2016). Natural history collections as windows on evolutionary processes. *Molecular Ecology*, *25*, 864–881.

Hughes, A. L., & Hughes, M. K. (1995). Natural selection on the peptide-binding regions of major histocompatibility complex molecules. *Immunogenetics*, *42*, 233–243.

Illera, J. C. (2020). '*Anthus berthelotii*, Berthelot's pipit', in Keller, V., Herrando, S., Vorisek, P. *et al.* (eds) *European Breeding Birds Atlas 2: distribution, abundance and change*. Barcelona: European Bird Census Council & Lynx Edicions, 797.

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., Rando, J. C., Richardson, D. S., & Emerson, B. C. (2012). Age, origins and extinctions of the avifauna of Macaronesia: A synthesis of phylogenetic and fossil information. *Quaternary Science Reviews*, *50*, 14–22.

Instituto Canario de Estadística. (2023). Población. Municipios por islas de Canarias y años.

Irestedt, M., Thörn, F., Müller, I. A., Jønsson, K. A., Ericson, P. G. P., & Blom, M. P. K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources*, *22*, 2672–2684.

Jackson, H. A., Percival-Alwyn, L., Ryan, C., Albeshr, M. F., Venturi, L., Morales, H. E., Mathers, T. C., *et al.* (2022). Genomic erosion in a demographically recovered bird species during conservation rescue. *Conservation Biology*, *36*, e13918.

Jangjoo, M., Matter, S. F., Roland, J., & Keyghobadi, N. (2016). Connectivity rescues genetic diversity after a demographic bottleneck in a butterfly population network. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, 10914–10919.

Jensen, E. L., & Leigh, D. M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. *Ecology and Evolution*, *12*, e9340.

Jensen, J. D., Foll, M., & Bernatchez, L. (2016). The past, present and future of genomic scans for selection. *Molecular Ecology*, *25*, 1–4.

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*, 1682–1684.

Kardos, M., Allendorf, F. W., & Luikart, G. (2014). Evaluating the role of inbreeding depression in heterozygosity-fitness correlations: how useful are tests for identity disequilibrium? *Molecular Ecology Resources*, *14*, 519–530.

Keller, L. F., & Waller, D. M. (2002). Inbreeding effects in wild populations. *Trends in Ecology & Evolution*, *17*, 230–241.

Kersten, O., Star, B., Leigh, D. M., Anker-Nilssen, T., Strøm, H., Danielsen, J., Descamps, S., *et al.* (2021). Complex population structure of the Atlantic puffin revealed by whole genome analyses. *Communications Biology, 4*, 922.

Khan, A., Patel, K., Shukla, H., Viswanathan, A., van der Valk, T., Borthakur, U., Nigam, P., *et al.* (2021). Genomic evidence for inbreeding depression and purging of deleterious genetic variation in Indian tigers. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2023018118.

Kimura, M. (1968). Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. *Genetics Research*, *11*, 247–270.

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

King, J. L., & Jukes, T. H. (1969). Non-Darwinian evolution. *Science*, *164*, 788–798.

Kloch, A., Wenzel, M. A., Laetsch, D. R., Michalski, O., Welc-Falęciak, R., & Piertney, S. B. (2018). Signatures of balancing selection in toll-like receptor (TLRs) genes - Novel insights from a free-living rodent. *Scientific Reports*, *8*, 8361.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, *15*, 356.

Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, *9*, 29.

Kreiner, J. M., Latorre, S. M., Burbano, H. A., Stinchcombe, J. R., Otto, S. P., Weigel, D., & Wright, S. I. (2022). Rapid weed adaptation and range expansion in response to agriculture over the past two centuries. *Science*, *378*, 1079–1085.

Lande, R. (1976). Natural selection and random genetic drift in phenotypic evolution. *Evolution*, *30*, 314–334.

Lazarus, R., Vercelli, D., Palmer, L. J., Klimecki, W. J., Silverman, E. K., Richter, B., Riva, A., *et al.* (2002). Single nucleotide polymorphisms in innate immunity genes: abundant variation and potential role in complex human disease. *Immunological Reviews*, *190*, 9–25.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology and Evolution*, *17*, 183–189.

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.

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*, 493–496.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. *Nature*, *362*, 709–715.

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature*, *457*, 830–836.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (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.

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.

Lynch, M., Conery, J., & Burger, R. (1995). Mutation accumulation and the extinction of small populations. *American Naturalist*, *146*, 489–518.

Ma, Y., Ding, X., Qanbari, S., Weigend, S., Zhang, Q., & Simianer, H. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, *115*, 426–436.

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

Mamanova, L., Coffey, A. J., Scott, C. E., Kozarewa, I., Turner, E. H., Kumar, A., Howard, E., *et al.* (2010). Target-enrichment strategies for next-generation sequencing. *Nature Methods*, *7*, 111–118.

Manel, S., Joost, S., Epperson, B. K., Holderegger, R., Storfer, A., Rosenberg, M. S., Scribner, K. T., *et al.* (2010). Perspectives on the use of landscape genetics to detect genetic adaptive variation in the field. *Molecular Ecology*, *19*, 3760–3772.

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*, *24*, 133–141.

Mardis, E. R. (2013). Next-generation sequencing platforms. *Annual Review of Analytical Chemistry*, *6*, 287–303.

Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and within-archipelago adaptation between island bird populations. *Royal Society Open Science*, *8*, 201146.

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging island populations of a bird. *Molecular Ecology*, *32*, 1972–1989.

Martínez-De La Puente, J., Martínez, J., Rivero-De Aguilar, J., Herrero, J., & Merino, S. (2011). On the specificity of avian blood parasites: Revealing specific and generalist relationships between haemosporidians and biting midges. *Molecular Ecology*, *20*, 3275–3287.

McEachern, M. B., Van Vuren, D. H., Floyd, C. H., May, B., & Eadie, J. M. (2011). Bottlenecks and rescue effects in a fluctuating population of golden-mantled ground squirrels (*Spermophilus lateralis*). *Conservation Genetics*, *12*, 285–296.

McGaughran, A. (2020). Effects of sample age on data quality from targeted sequencing of museum specimens: What are we capturing in time? *BMC Genomics*, *21*, 188.

Meyer, D., & Thomson, G. (2001). How selection shapes variation of the human major

histocompatibility complex: A review. *Annals of Human Genetics*, *65*, 1–26.

Mikheyev, A. S., Tin, M. M. Y., Arora, J., & Seeley, T. D. (2015). Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nature Communications*, *6*, 7991.

Miller, J. M., & Coltman, D. W. (2014). Assessment of identity disequilibrium and its relation to empirical heterozygosity fitness correlations: a meta-analysis. *Molecular Ecology*, *23*, 1899–1909.

Miranda, I., Giska, I., Farelo, L., Pimenta, J., Zimova, M., Bryk, J., Dalén, L., *et al.* (2021). Museomics dissects the genetic basis for adaptive seasonal coloration in the Least weasel. *Molecular Biology and Evolution*, *38*, 4388–4402.

Morales, J., Rodríguez, A., Alberto, V., MacHado, C., & Criado, C. (2009). The impact of human activities on the natural environment of the Canary Islands (Spain) during the pre-Hispanic stage (3rd-2nd Century BC to 15th Century AD): An overview. *Environmental Archaeology*, *14*, 27–36.

Moritz, C., Patton, J. L., Conroy, C. J., Parra, J. L., White, G. C., & Beissinger, S. R. (2008). Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. *Science*, *322*, 261–264.

Myers, N., Mittermeler, R. A., Mittermeler, C. G., Da Fonseca, G. A. B., & Kent, J. (2000). Biodiversity hotspots for conservation priorities. *Nature*, *403*, 853–858.

Nadachowska-Brzyska, K., Dutoit, L., Smeds, L., Kardos, M., Gustafsson, L., & Ellegren, H. (2021). Genomic inference of contemporary effective population size in a large island population of collared flycatchers (*Ficedula albicollis*). *Molecular Ecology*, *30*, 3965–3973.

Nadeau, N. J., & Jiggins, C. D. (2010). A golden age for evolutionary genetics? Genomic studies of adaptation in natural populations. *Trends in Genetics*, *26*, 484–492.

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.

Nei, M., Maruyama, T., & Chakraborty, R. (1975). The bottleneck effect and genetic variability in populations. *Evolution*, *29*, 1–10.

Njabo, K. Y., Cornel, A. J., Bonneaud, C., Toffelmier, E., Sehgal, R. N. M., Valkiũnas, G., Russell, A. F., *et al*. (2011). Nonspecific patterns of vector, host and avian malaria parasite associations in a central African rainforest. *Molecular Ecology*, *20*, 1049–1061.

Norder, S. J., de Lima, R. F., de Nascimento, L., Lim, J. Y., Fernández-Palacios, J. M., Romeiras, M. M., Elias, R. B., *et al.* (2020). Global change in microcosms: Environmental and societal predictors of land cover change on the Atlantic Ocean Islands. *Anthropocene*, *30*, 100242.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 332–342.

Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402.

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. (2010). Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 185–205.

Orlando, L., Allaby, R., Skoglund, P., Der Sarkissian, C., Stockhammer, P. W., Ávila-Arcos, M. C., Fu, Q., *et al.* (2021). Ancient DNA analysis. *Nature Reviews Methods Primers*, *1*, 14.

Otto, R., Krüsi, B. O., & Kienast, F. (2007). Degradation of an arid coastal landscape in relation to land use changes in Southern Tenerife (Canary Islands). *Journal of Arid Environments*, *70*, 527–539.

Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proceedings of the Royal Society B: Biological Sciences*, *285,* 20182047.

Ouborg, N. J., van Treuren, R., & van Damme, J. M. M. (1991). The significance of genetic erosion in the process of extinction. *Oecologia*, *86*, 359–367.

Paabo, S. (1989). Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. *Proceedings of the National Academy of Sciences of the United States of America*, *86*, 1939–1943.

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*, 445–464.

Parejo, M., Wragg, D., Henriques, D., Charrière, J.-D., & Estonba, A. (2020). Digging into the genomic past of Swiss honey bees by whole-genome sequencing museum specimens. *Genome Biology and Evolution*, *12*, 2535–2551.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., Quail, M., *et al.* (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*, 275–278.

Patiño, J., Whittaker, R. J., Borges, P. A. V., Fernández-Palacios, J. M., Ah-Peng, C., Araújo, M. B., Ávila, S. P., *et al.* (2017). A roadmap for island biology: 50 fundamental questions after 50 years of The Theory of Island Biogeography. *Journal of Biogeography*, *44*, 963–983.

Pinhasi, R., Fernandes, D., Sirak, K., Novak, M., Connell, S., Alpaslan-Roodenberg, S., Gerritsen, F., *et al.* (2015). Optimal ancient DNA yields from the inner ear part of the human petrous bone. *PLoS ONE*, *10*, e0129102.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., *et al.* (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2025453118.

Pinto, A. V., Hansson, B., Patramanis, I., Morales, H. E., & van Oosterhout, C. (2023). The impact of habitat loss and population fragmentation on genomic erosion. *Conservation Genetics*, *25*, 49–57.

Poelstra, J. W., Ellegren, H., & Wolf, J. B. W. (2013). An extensive candidate gene approach to speciation: Diversity, divergence and linkage disequilibrium in candidate pigmentation genes across the European crow hybrid zone. *Heredity*, *111*, 467–473.

Primack, D., Imbres, C., Primack, R. B., Miller-Rushing, A. J., & Del Tredici, P. (2004). Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. *American Journal of Botany*, *91*, 1260–1264.

Prüfer, K., Stenzel, U., Hofreiter, M., Pääbo, S., Kelso, J., & Green, R. E. (2010). Computational challenges in the analysis of ancient DNA. *Genome Biology*, *11*, R47.

Qu, Y., Tian, S., Han, N., Zhao, H., Gao, B., Fu, J., Cheng, Y., *et al.* (2015). Genetic responses to seasonal variation in altitudinal stress: whole-genome resequencing of great tit in eastern Himalayas. *Scientific Reports*, *5*, 14256.

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

Ramstad, K. M., Woody, C. A., Sage, G. K., & Allendorf, F. W. (2004). Founding events influence genetic population structure of sockeye salmon (*Oncorhynchus nerka*) in Lake Clark, Alaska. *Molecular Ecology*, *13*, 277–290.

Rando, J. C., Cabrera, V. M., Larruga, J. M., Hernández, M., González, A. M., Pinto, F., & Bandelt. H.-J. (1999). Phylogeographic patterns of mtDNA reflecting the colonization of the Canary Islands. *Annals of Human Genetics*, *63*, 413–428.

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. A. F., *et al.* (2017). Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*, 1450–1477.

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*, 1049–1060.

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

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*, 4348–4370.

Richardson, D. S., Komdeur, J., & Burke, T. (2004). Inbreeding in the Seychelles warbler: environment-dependent maternal effects. *Evolution*, *58*, 2037–2048.

Robinson, J. A., Räikkönen, J., Vucetich, L. M., Vucetich, J. A., Peterson, R. O., Lohmueller, K. E., & Wayne1, R. K. (2019). Genomic signatures of extensive inbreeding in Isle Royale wolves, a population on the threshold of extinction. *Science Advances*, *5*, 757–786.

Rokas, A., & Abbot, P. (2009). Harnessing genomics for evolutionary insights. *Trends in Ecology and Evolution*, *24*, 192–200.

Ryman, N., Laikre, L., & Hössjer, O. (2019). Do estimates of contemporary effective population size tell us what we want to know? *Molecular Ecology*, *28*, 1904–1918.

Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrigal, J., Westbury, M. V., de Manuel, M., Margaryan, A., Ciucani, M. M., *et al.* (2021). Historical population declines

prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*). *Molecular Ecology*, *30*, 6355–6369.

Santure, A. W., De Cauwer, I., Robinson, M. R., Poissant, J., Sheldon, B. C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (*Parus major*) population. *Molecular Ecology*, *22*, 3949–3962.

Satam, H., Joshi, K., Mangrolia, U., Waghoo, S., Zaidi, G., Rawool, S., Thakare, R., *et al*. (2023). Next-generation sequencing technology: Current trends and advancements. *Biology*, *12*, 997.

Sawyer, S., Krause, J., Guschanski, K., Savolainen, V., & Pääbo, S. (2012). Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. *PLoS ONE*, *7*, e34131.

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 23–43.

Schroeder, J., Lourenço, P. M., Hooijmeijer, J. C. E. W., Both, C., & Piersma, T. (2009). A possible case of contemporary selection leading to a decrease in sexual plumage dimorphism in a grassland-breeding shorebird. *Behavioral Ecology*, *20*, 797–807.

Schubert, M., Ginolhac, A., Lindgreen, S., Thompson, J. F., AL-Rasheid, K. A. S., Willerslev, E., Krogh, A., *et al.* (2012). Improving ancient DNA read mapping against modern reference genomes. *BMC Genomics*, *13*, 178.

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., *et al.* (2014). Genomics and the origin of species. *Nature Reviews Genetics*, *15*, 176–192.

Sendell-Price, A. T., Ruegg, K. C., Robertson, B. C., & Clegg, S. M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*, *30*, 2495–2510.

Shafer, A. B. A., Wolf, J. B. W., Alves, P. C., Bergström, L., Bruford, M. W., Brännström, I., Colling, G., *et al.* (2015). Genomics and the challenging translation into conservation practice. *Trends in Ecology and Evolution*, *30*, 78–87.

Shultz, A. J., Baker, A. J., Hill, G. E., Nolan, P. M., & Edwards, S. V. (2016). SNPs across time and space: population genomic signatures of founder events and epizootics in the House Finch (*Haemorhous mexicanus*). *Ecology and Evolution*, *6*, 7475–7489.

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.

Smith, J. M., & Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetics Research*, *23*, 23–35.

Snead, A. A., & Clark, R. D. (2022). The biological hierarchy, time, and temporal 'omics in evolutionary biology: a perspective. *Integrative and Comparative Biology*, *62*, 1872–1886.

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

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

Spurgin, L. G., Wright, D. J., van der Velde, M., Collar, N. J., Komdeur, J., Burke, T., & Richardson, D. S. (2014). Museum DNA reveals the demographic history of the endangered Seychelles warbler. *Evolutionary Applications*, *7*, 1134–1143.

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20170279.

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

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

Stoffel, M. A., Johnston, S. E., Pilkington, J. G., & Pemberton, J. M. (2021). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nature Communications*, *12*, 2972.

Supple, M. A., & Shapiro, B. (2018). Conservation of biodiversity in the genomics era. *Genome Biology*, *19*, 1–12.

Talla, V., Johansson, A., Dincă, V., Vila, R., Friberg, M., Wiklund, C., & Backström, N. (2019). Lack of gene flow: Narrow and dispersed differentiation islands in a triplet of *Leptidea* butterfly species. *Molecular Ecology*, *28*, 3756–3770.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., & Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nature Reviews Genetics*, *20*, 467–484.

Taron, U. H., Lell, M., Barlow, A., & Paijmans, J. L. A. (2018). Testing of alignment parameters for ancient samples: Evaluating and optimizing mapping parameters for ancient samples using the TAPAS tool. *Genes*, *9*, 157.

Theissinger, K., Fernandes, C., Formenti, G., Bista, I., Berg, P. R., Bleidorn, C., Bombarely, A., *et al.* (2023). How genomics can help biodiversity conservation. *Trends in Genetics*, *39*, 545–559.

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, *25*, 2144–2164.

Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2020). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, *20*, 1220–1227.

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

Valente, L., Illera, J. C., Havenstein, K., Pallien, T., Etienne, R. S., & Tiedemann, R. (2017). Equilibrium bird species diversity in Atlantic Islands. *Current Biology*, *27*, 1660–1666.

Van Der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology*, *29*, 165–170.

van Dijk, E., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in Genetics*, *30*, 418–426.

van Riper, C., van Riper, S. G., & Hansen, W. R. (2002). Epizootiology and effect of avian pox on Hawaiian forest birds. *The Auk*, *119*, 929–942.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., *et al.* (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, *7*, 12558.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *The Annual Review of Genetics*, *47*, 97–120.

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

Wadgymar, S. M., Demarche, M. L., Josephs, E. B., Sheth, S. N., & Anderson, J. T. (2022). Local adaptation: Causal agents of selection and adaptive trait divergence. *Annual Review of Ecology, Evolution, and Systematics*, *53*, 87–111.

Wandeler, P., Hoeck, P. E. A., & Keller, L. F. (2007, December). Back to the future: museum specimens in population genetics. *Trends in Ecology and Evolution*, *22*, 634–642.

Warren, B. H., Simberloff, D., Ricklefs, R. E., Aguilée, R., Condamine, F. L., Gravel, D., Morlon, H., *et al.* (2015). Islands as model systems in ecology and evolution: Prospects fifty years after MacArthur-Wilson. *Ecology Letters*, *18*, 200–217.

Weiß, C. L., Schuenemann, V. J., Devos, J., Shirsekar, G., Reiter, E., Gould, B. A., Stinchcombe, J. R., *et al.* (2016). Temporal patterns of damage and decay kinetics of DNA retrieved from plant herbarium specimens. *Royal Society Open Science, 3*, 160239.

Wenzel, M. A., James, M. C., Douglas, A., & Piertney, S. B. (2015). Genome-wide association and genome partitioning reveal novel genomic regions underlying variation in gastrointestinal nematode burden in a wild bird. *Molecular Ecology*, *24*, 4175–4192.

Whibley, A., Kelley, J. L., & Narum, S. (2021). The changing face of genome assemblies: Guidance on achieving high-quality reference genomes. *Molecular Ecology Resources*, *21*, 641–652.

Whittaker, R. J., Fernández-Palacios, J. M., Matthews, T. J., Borregaard, M. K., & Triantis, K. A. (2017). Island biogeography: Taking the long view of natures laboratories. *Science*, *357*, aam8326.

Willi, Y., & Hoffmann, A. A. (2009). Demographic factors and genetic variation influence population persistence under environmental change. *Journal of Evolutionary Biology*, *22*, 124–133.

Willi, Y., Van Buskirk, J., & Hoffmann, A. (2006). Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 433–458.

Willi, Y., Van Buskirk, J., Schmid, B., & Fischer, M. (2007). Genetic isolation of fragmented populations is exacerbated by drift and selection. *Journal of Evolutionary Biology*, *20*, 534–542.

Wood, A. W., Szpiech, Z. A., Lovette, I. J., Smith, B. T., & Toews, D. P. L. (2023). Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers. *Current Biology*, *33*, 2823–2829.

Wright, A. F. (2005). 'Genetic variation: Polymorphisms and mutations', in *Encyclopedia of Life Sciences*. Chichester: John Wiley and Sons, 959–968.

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

Wright, S. (1949). The genetical structure of populations. *Annals of Eugenics*, *15*, 323–354.

Wu, M. Y., Lau, C. J., Xin Ng, E. Y., Baveja, P., Gwee, C. Y., Sadanandan, K., Ferasyi, T. R., *et al.* (2022). Genomes from historic DNA unveil massive hidden extinction and terminal endangerment in a tropical Asian songbird radiation. *Molecular Biology and Evolution*, *39*, msac189.

Yadav, S., J. Stow, A., & Dudaniec, R. Y. (2021). Microgeographical adaptation corresponds to elevational distributions of congeneric montane grasshoppers. *Molecular Ecology*, *30*, 481–498.

Yamada, T., Biogeography, M. M.-J. of, & 2012, undefined. (2012). Impact of geographical isolation on genetic differentiation in insular and mainland populations of *Weigela coraeensis* (Caprifoliaceae) on Honshu and the Izu Islands. *Journal of Biogeography*, *39*, 901–917.

Zimmermann, J., Hajibabaei, M., Blackburn, D. C., Hanken, J., Cantin, E., Posfai, J., & Evans, T. C. (2008). DNA damage in preserved specimens and tissue samples: A molecular assessment. *Frontiers in Zoology*, *5*, 18.

Chapter 2

Genomic associations with poxvirus across divergent island populations in Berthelot's pipit

A pox-infected Berthelot's pipit with characteristic lesion on the bill

2.1. Abstract

Understanding the mechanisms and genes that enable animal populations to adapt to pathogens is important from an evolutionary, health and conservation perspective. Berthelot's pipit (*Anthus berthelotii*) experiences extensive and consistent spatial heterogeneity in avian pox infection pressure across its range of island populations, thus providing an excellent system with which to examine how pathogen-mediated selection drives spatial variation in immunogenetic diversity. Here we test for evidence of genetic variation associated with avian pox at both an individual and population-level. At the individual level, we find no evidence that variation in MHC class I and TLR4 (both known to be important in recognising viral infection) was associated with pox infection within two separate populations. However, using genotype-environment association (Bayenv) in conjunction with genome-wide (ddRAD-seq) data, we detected strong associations between population-level avian pox prevalence and allele frequencies of single nucleotide polymorphisms (SNPs) at a number of sites across the genome. These sites were located within genes involved in cellular stress signalling and immune responses, many of which have previously been associated with responses to viral infection in humans and other animals. Consequently, our analyses indicates that pathogen-mediated selection may play a role in shaping genomic variation among relatively recently colonised island bird populations and highlights the utility of genotype-environment associations for identifying candidate genes potentially involved in host-pathogen interactions.

2.2. Introduction

Infection with pathogens can have considerable impact on individual fitness by reducing reproductive success, via increased mortality or morbidity (Anderson & May, 1979; Daszak *et al.*, 2000). Consequently, pathogen-mediated selection has the potential to affect the population dynamics, adaptation, and genetic variation of hosts (O'Brien & Evermann, 1988; Hudson *et al.*, 1998; Ortego *et al.*, 2007; Spurgin & Richardson, 2010). Selection on host immunity-related genes – that is, those involved in pathogen recognition and elimination – may decrease within-population genetic diversity and increase between-population divergence as the alleles that provide the most benefit approach fixation (Mukherjee *et al.*, 2009). However, evolutionary dynamics in natural disease systems are often more complex given that interactions between hosts and pathogens promote continual selection for reciprocal adaptations, resulting in an evolutionary arms race (Dawkins & Krebs, 1979; Paterson *et al.*, 2010).

Selection as a result of diverse or rapidly evolving pathogens can promote and maintain genetic diversity in the host population (reviewed in Charlesworth, 2006; Spurgin & Richardson, 2010). Any selective regime that acts to maintain multiple alleles at a locus can be considered a mechanism of such 'balancing' selection, including heterozygote advantage (Doherty & Zinkernagel, 1975), rare allele advantage (Slade & McCallum, 1992), and fluctuating selection (resulting from spatial and/or temporal changes in infection prevalence; Hill *et al.*, 1991). However, these mechanisms are complicated and difficult to untangle in any given system, not least because they are non-exclusive, and may also interact with each other and with other processes (such as sexual selection) to shape genetic diversity (Apanius *et al.*, 1997; Ejsmond *et al.*, 2014). Understanding how pathogens and immune genes covary spatially within and across populations may provide some insight into the mechanisms and genes involved in adaptive evolution.

The search for pathogen associated genes has often involved a candidate gene approach (Bernatchez & Landry, 2003; Netea *et al.*, 2012), however this is limited in its ability to include pre-identified loci. Using genome-wide markers would allow for discovery of further loci but even traditional population-level based approaches may have limited power to

detect small frequency shifts responsible for adaptation in polygenic traits (Pritchard *et al.*, 2010). Alternatively, landscape genomic approaches such as genotype-environment association studies, also called environmental association analyses, test for correlations between population patterns of allele frequencies with environmental factors across a spatial scale (Rellstab *et al.*, 2015; Hoban *et al.*, 2016). This method has been successfully employed, for example, to identify adaptation to the climatic environment in a range-expanding damselfly (Dudaniec *et al.*, 2018), to uncover parallel adaptive responses in congeneric grasshoppers along elevational gradients (Yadav *et al.*, 2021), and to understand abiotic stress tolerance in barley landraces (Lei *et al.*, 2019). Comparatively few genotype-environment associations have attempted to identify genes that confer resistance to infectious diseases (but see Mackinnon *et al.*, 2016; Fraik *et al.*, 2020). We believe that such an approach could also overcome some of the difficulties that other individual-based association methods experience when used for infectious disease data in wild populations. That is, rather than test for associations between individual genotypes and disease status – which could include individuals that have survived infection or are yet to be exposed to the pathogen in the same 'no infection' set – testing for associations between allele frequencies and pathogen prevalence represents the selection generated by the pathogen in the population.

Avian pox is caused by a double-stranded DNA virus (genus *Avipoxvirus*; family Poxviridae) and has been reported in numerous bird species worldwide (Bolte *et al.*, 1999; van Riper & Forrester, 2007; Williams *et al.*, 2021). Transmission occurs by biting insect vectors, skin-to-skin contact, or indirect contact when virions persist in the environment (van Riper *et al.*, 2002; Smits *et al.*, 2005). Infection manifests as proliferative lesions: the most common occur on featherless skin, but internal lesions can form on the diphtheritic membrane of the mouth, respiratory tract and digestive system (Tripathy, 1993). In severe cases, lesions may lead to impaired vision, feeding, flight and breathing, emaciation and death (Davidson *et al.*, 1980; Docherty *et al.*, 1991; Orós *et al.*, 1997). Pox infections have been shown to have implications for host predation (Laiolo *et al.*, 2007), secondary infections (van Riper & Forrester, 2007), male pairing success (Kleindorfer & Dudaniec, 2006), reproductive output (Lachish *et al.*, 2012; Vanderwerf & Young, 2016) and productivity (Carrete *et al.*, 2009). Severe population declines in endemic island species have even been linked to avian pox

outbreaks (van Riper *et al.*, 2002; Kleindorfer & Dudaniec, 2006; Alley *et al.*, 2010). Although minor pox lesions typically heal, birds are often left with deformities, including missing digits and misshapen bills (Vanderwerf, 2001; van Riper *et al.*, 2002), which may affect foraging success. Despite these fitness effects, few studies have investigated how pox infection varies both spatially and temporally (Carrete *et al.*, 2009; Lawson *et al.*, 2012; Spurgin *et al.*, 2012; Zylberberg *et al.*, 2012; Samuel *et al.*, 2018), and none have investigated how it shapes patterns of host genetic diversity. Such information would provide insight into the selective pressure exerted by *Avipoxvirus* and its consequence for host immunogenetic adaptation in wild populations.

Berthelot's pipit (*Anthus berthelotii*) – a small passerine, endemic to three archipelagos of Macaronesia – provides an excellent natural system with which to investigate pathogen-mediated selection and the evolution of immune genes. The pipit colonised the Canary Islands from Africa *ca.* 2.5 Mya and from there colonised the Selvagens and the Madeiran archipelago in two independent events relatively recently *ca.* 8,500–50,000 years ago (Illera *et al.*, 2007; Spurgin *et al.*, 2014; Martin *et al.*, 2021). Population bottlenecks associated with these colonisation events, and a lack of subsequent gene flow, have resulted in low genetic diversity within these populations and high genetic structuring across archipelagos (Illera *et al.*, 2007; Spurgin *et al.*, 2014; Armstrong *et al.*, 2018; Martin *et al.*, 2021). Importantly, pathogen pressures – specifically the prevalence of avian pox and malaria – are temporally consistent within populations but spatially variable between populations (Illera *et al.*, 2008; Spurgin *et al.*, 2012), and at a finer scale within populations (González-Quevedo *et al.*, 2014). Therefore, these pathogens are likely to have shaped host genetic diversity across populations of Berthelot's pipits.

Candidate loci that may be involved in combating avian pox infection have already been identified in Berthelot's pipit. At the population-level, balancing selection appears to have maintained functionally divergent loci at the class I major histocompatibility complex (MHC) in Berthelot's pipit (Spurgin *et al.*, 2011; González-Quevedo, Phillips, *et al.*, 2015). The molecules encoded by these loci are key receptors involved in the acquired immune system presenting antigens in cells infected by intracellular pathogens, such as viruses (Hewitt, 2003). Similarly, while overall patterns of diversity at Toll-like receptors (TLRs; involved in the

innate immune response) are shaped by drift across the pipit's range, polymorphisms have been retained in most populations, and evidence for positive selection at some TLR loci exists (González-Quevedo, Spurgin, *et al.*, 2015). This includes TLR4, a locus that plays a role in virus sensing (Barton, 2007); TLR4-dependent signalling has been shown to be important against influenza (Shinya *et al.*, 2012) and vaccinia virus (the prototypic poxvirus; Hutchens *et al.*, 2008). Evidence also exists for pathogen-mediated selection on candidate immune genes within populations of Berthelot's pipits (González-Quevedo *et al.*, 2014, 2016; Armstrong *et al.*, 2019), however these studies have only focused on malaria as the selective agent. Thus, MHC class I and TLR4 are candidates for investigating immunogenetic adaptation to avian poxvirus. Nevertheless, these genes only represent a fraction of the 144 immune-related genes identified in the avian genome (Ekblom *et al.*, 2010) and it is also important to assess variation at a greater number of sites throughout the genome.

Here, we take advantage of the heterogeneity in pox prevalence within and between populations of Berthelot's pipit to identify loci that may be important in host response to pox infection. First, we test the hypothesis that variation at previously identified candidate immune effectors (TLR4 and MHC class I exon 3) will be associated with pox infection status in individual-based analyses, consistent with pathogen-mediated selection. We further hypothesise that differing pox-mediated selective pressure among populations will shape the distribution of variation across the genome such that allele frequencies at the specific loci involved will show exceptional correlations with local pox prevalence. To test this, we perform a genotype-environment association with genome-wide restriction-site associated DNA sequence (RAD-seq) data from populations across the Berthelot's pipit range, encompassing the entire gradient of pox prevalence in this species. Finally, we identify likely candidate genes in close proximity to the loci identified in the genotype-environment association, to assess the potential role they may play in adaptation to poxvirus.

2.3. Materials and Methods

2.3.1. Field sampling and data collection

Berthelot's pipits were caught during periods between 2005–2020 (Table S2.1) across 12 islands of their range (Figure 2.1). All sampling undertaken prior to 2019 has been described previously (Illera *et al.*, 2007; Spurgin *et al.*, 2012; González-Quevedo *et al.*, 2014; Armstrong *et al.*, 2019). The remaining samples were collected in April 2019 (Lanzarote; *n* = 83) and February–June 2020 (Lanzarote, La Graciosa, La Gomera and Tenerife; *n* = 200). Multiple sampling localities were chosen across each island to achieve a representative sample of the entire island population. In total, *n* = 1,661 individuals were sampled.

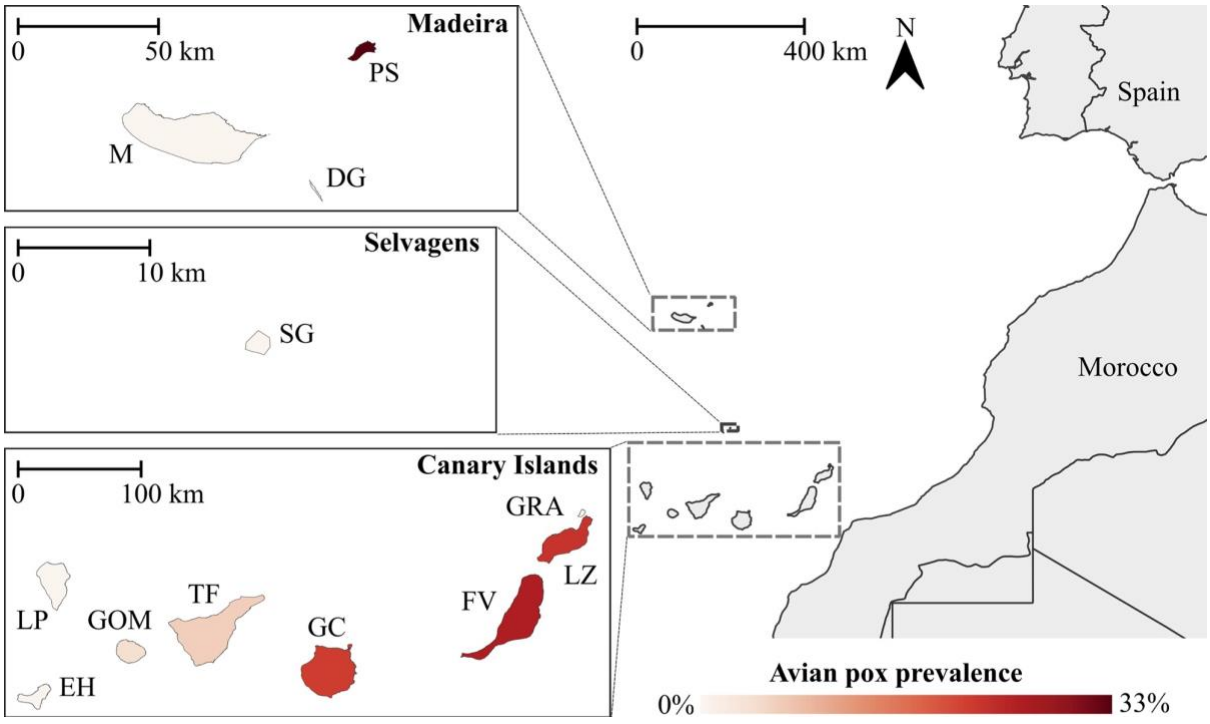

Figure 2.1. Map of the 12 islands sampled for Berthelot's pipits across its Macaronesian range. Populations are coloured according to their overall estimated pox prevalence across the 15-year sampling period. M = Madeira; PS = Porto Santo; DG = Deserta Grande; SG = Selvagem Grande; LP = La Palma; EH = El Hierro; GOM = La Gomera; TF = Tenerife; GC = Gran Canaria; FV = Fuerteventura; GRA = La Graciosa; LZ = Lanzarote.

Birds were captured in spring traps baited with *Tenebrio molitor* larvae. A blood sample (*ca*. 25 µl) was taken by brachial venipuncture and stored in absolute ethanol (800 µl) at room

temperature. Birds were ringed with a uniquely numbered metal band issued by the Spanish or Portuguese authorities as relevant. Age (adult/juvenile) was determined according to feather moult pattern (Cramp, 1985), and mass (± 0.1 g), wing (± 1 mm), tarsus, head length, and bill height, length and width (all ± 0.1 mm) were measured.

To date, there is no reliable molecular/serological technique for diagnosing avian pox infection from blood samples (Smits *et al.*, 2005; Farias *et al.*, 2010; Williams *et al.*, 2014; Baek *et al.*, 2020). In our study, each bird was carefully assessed for evidence of pox infection – based on the presence/absence of lesions around the eyes, beak, feet, legs, or sparsely feathered areas. It is therefore possible that some infected birds may have been asymptomatic and incorrectly classified as uninfected. *Avipoxvirus* DNA has been amplified from samples taken from skin lesions for seven pipits included in this study (six from Porto Santo and one from Lanzarote), and identified as a single strain (see Illera *et al.*, 2008). There is no evidence of other pathogens that result in similar lesions in this system. Previous work has also identified poxvirus infection in Berthelot's pipit, and other species in the Canary Islands (Medina *et al.*, 2004; Smits *et al.*, 2005), including short-toed larks (*Calandrella rufescens*) which inhabit the same shrub steppes. However, the specific strain of avian pox in Berthelot's pipits is not found in any of the other species, and thus appears to be host-specific, potentially indicating long-term coevolution between the strain and the host. Phylogenetic studies place this Berthelot's pipit lineage within the *Canarypox virus* clade, as an outlier to a subclade of lineages formed primarily of passerines (Illera *et al.*, 2008; Gyuranecz *et al.*, 2013).

DNA was extracted from blood following a salt extraction protocol (Richardson *et al.*, 2001) and host sex was determined molecularly (Griffiths *et al.*, 1998). To detect infection with *Haemoproteus* and *Plasmodium* spp. (here termed avian malaria for simplicity), we used a nested polymerase chain reaction (PCR) method described by Waldenström *et al.* (2004) to screen each sample at least twice (three times if the first two attempts gave contrasting results). Individuals were considered to be infected with malaria if the PCR produced a positive result twice, and all positive and negative controls gave expected results. For a subset of 400 infected individuals, the amplicon was sequenced to identify the strains

present (see Illera *et al.*, 2008; Spurgin *et al.*, 2012; González-Quevedo *et al.*, 2014; Armstrong *et al.*, 2019).

2.3.2. Sequencing and genotyping of candidate gene variants

To investigate genetic associations with pox at the individual level, we utilised targeted genotyping of TLR4 and MHC class I loci. The TLR4 dataset had been generated previously for a study of associations with malaria (Armstrong *et al.*, 2019); a complete description of methods and discussion of TLR4 variation, can be found therein. In brief, 780 individuals were screened for TLR4 variation from the islands of Porto Santo (*n* = 190) and Tenerife (*n* = 590). These include individuals from a previous cross-population study by González-Quevedo *et al.* (2015) in which 23–30 individuals were TLR4 genotyped from 13 Berthelot's pipit populations. A section of the extracellular region (660 bp) of the TLR4 gene (leucine-rich repeat domain) – directly involved in pathogen recognition – was amplified using primers PauTLR4F and PauTLR4R (Grueber *et al.*, 2012). The 129 samples from Porto Santo (2016) were genotyped using Sanger Sequencing, while the remaining individuals from Tenerife and Porto Santo (*n* = 577) were genotyped using KASP, a proprietary technology of LGC Genomics, as part of Armstrong *et al.* (2019). Five single nucleotide polymorphisms (SNPs) were reported, but one was later excluded because the SNP had a minor allele frequency (MAF) < 0.05 (Armstrong *et al.*, 2019). Another of these SNPs was tri-allelic, therefore the least frequent alternate allele (T) was treated as missing (Porto Santo *n* = 1; Tenerife *n* = 12). Five nucleotide haplotypes were inferred using DnaSP v6 (Librado & Rozas, 2009), and translated into four protein haplotypes (Armstrong *et al.*, 2019; details regarding the SNPs, their positions, and the haplotypes are provided in Table S2.2). Samples with phase probabilities < 0.90 were excluded from further analyses.

González-Quevedo *et al.* (2015) used 454 sequencing to screen variation at exon 3 of MHC class I (which partially encodes the antigen binding region) in 310 individuals from Tenerife (2011). In brief, two replicate PCR reactions were performed for each sample using different sets of fusion primers. Amplicon products were purified, pooled in equimolar amounts, and sequenced using a GS FLX Titanium system. Stringent variant/artefact identification and validation criteria were applied to identify putative MHC alleles (described in full in

González-Quevedo, Phillips, *et al.*, 2015) and 22 alleles were found within this population sample (GenBank accession numbers: JN799601–JN799604, JN799606, JN799610–JN799612, JN799623, JN799625, JN799636–JN799639, JN799641, and KM593305–KM593311). As recommended in González-Quevedo *et al.* (2015), we excluded two alleles with low amplification efficiencies (ANBE3 and ANBE31) from downstream analyses because their presence/absence cannot always be reliably ascertained for every individual, and one sample was discarded due to poor coverage. Here, the term 'allele' refers to unique sequence variants amplified across a number of duplicated loci (the estimate for exon 3 of MHC class I genes in Berthelot's pipit is six loci; González-Quevedo, Phillips, *et al.*, 2015).

2.3.3. RAD sequencing

To explore population-level genetic associations with pox, we analysed previously published double digest RAD-seq (ddRAD-seq) data from 20 individuals selected from each population (22 from lowland Tenerife; Armstrong *et al.*, 2018). Pipits on the mountain of El Teide (2,000 m above sea level) are considered a separate population from the rest of lowland Tenerife (Armstrong *et al.*, 2018), thus there are 13 populations in this dataset. Individuals screened were selected from the 2006 and 2009 sampling, in even sex ratios where possible, and across a wide geographical coverage within each population to reduce the likelihood of including related individuals.

Library preparation and initial bioinformatics followed the protocol in DaCosta and Sorenson (2014). Genotyping was performed by mapping reads to the zebra finch (*Taeniopygia guttata*) genome sequence (v3.2.4; Warren *et al.*, 2010). Within the 'Berthelot's' dataset (Armstrong *et al.*, 2018) – used here for analyses – the loci that could not be confidently genotyped in a minimum of four samples, and those with missing or ambiguous genotypes for > 10% of samples, were treated as missing data.

2.3.4. Identification of genetic variants associated with individual disease status

Generalised linear models (GLMs), with and without mixed effects, were used to identify predictors of pox infection at the individual level. For all models including genetic predictors,

each island was modelled separately because population-specific associations between genetic variation and pox may have evolved, and different genetic variants were present in the different populations (Armstrong *et al.*, 2019). Predictor variables were assessed for collinearity using variance inflation factors (VIFs) or, in the case of categorical variables with more than two levels, generalised variance inflation factors (GVIFs; Fox & Monette, 1992). To obtain values equivalent to VIFs, GVIFs were transformed by squaring the standardised GVIF ($GVIF^{1/2df}$, where *df* is degrees of freedom; Fox & Weisberg, 2019). VIFs ranged from 1.00 to 1.95 for all predictor variables used in the following models, thus none were excluded (threshold value: VIFs < 3; Zuur *et al.*, 2010).

Prior to testing for associations with key candidate genes within populations, we used a multi-population dataset to model the relationship between non-genetic variables and pox infection status (not infected/infected) to build a base model using the maximum sample size. This dataset consisted of all individuals screened for variation at candidate genes, but two individuals were excluded because the sexing PCR failed (n = 778). This generalised linear mixed model (GLMM) included age class (adult/juvenile), sex (male/female), island (Porto Santo/Tenerife) and malaria infection status (not infected/infected) as fixed factors, with sampling year as a random factor. We also considered interactions between age class and island, and malaria infection status and island. Non-significant predictors and interaction terms (p > 0.05) were removed sequentially leaving a minimal model (Bolker *et al.*, 2009). Such an approach can inflate the probability of type-1 errors (Mundry & Nunn, 2009), thus all removed variables were re-entered into the minimal model one at a time to determine their significance and parameter estimates. Genetic variables were later added to the minimal model to assess their significance in explaining variation in pox infection status.

GLMMs were used to test for associations between TLR4 variation and pox infection status. We focused on associations with protein haplotypes rather than single SNPs because these should better reflect functional differences at the TLR. Nonetheless, different protein haplotypes may have the same functional properties. In future, an *in-silico* approach to assess the specific functional/structural/regulatory effects of these different variants, and ultimately their contribution to disease development, could be undertaken. Different TLR4 variants were represented as fixed factors in the following ways: (i) protein haplotype

(presence/absence), (ii) protein haplotype heterozygosity (homozygote/heterozygote), and (iii) specific protein genotype (to investigate whether there is a synergistic effect when two specific protein haplotypes are present). Rare variants and genotypes (< 0.05 in frequency) were removed from analyses (see Table S2.2) because they lack sufficient power to test effects. The effect of individual haplotypes, heterozygosity and genotypes were investigated separately to avoid problems of collinearity.

Associations between MHC class I variation and pox infection were assessed for birds sampled from Tenerife in 2011 (n = 309). We could not assign alleles to specific loci and resolve haplotypes because MHC alleles were amplified across multiple unidentified loci, thus we were unable to test for associations at the haplotype level. Instead, separate GLMs were performed to evaluate the effects of (i) MHC diversity (number of alleles per individual, 3–10) and optimality (the quadratic of allele number; to investigate whether intermediate MHC heterozygosity has the greatest fitness benefit), and (ii) individual MHC alleles (presence/absence). Rare and almost fixed MHC alleles (< 0.05 and > 0.95 in frequency) were removed from the second model because they provide no resolution. We also removed two alleles (ANBE16 and ANBE49) because they only occurred in individuals that were not infected with pox (12.00 and 7.44 in frequency, respectively) and therefore prevented model convergence. We detected a strong positive association between malaria and pox infection (see Section 2.4.1), and therefore considered the possibility that including malaria as a variable in the minimal model may have masked the genetic effects on pox infection status. Thus, all genetic models above were performed again without malaria infection status as a predictor.

For all models above, we fitted a binomial error structure and used a logit link function. All modelling was performed using R v4.0.2 (R Development Core Team, 2020), with GLMMs constructed using the 'lme4' package (Bates *et al.*, 2015), VIFs calculated using the 'car' package (Fox & Weisberg, 2019), and the explained variance (R^2) calculated according to the delta method (Nakagawa *et al.*, 2017) using the 'r.squaredGLMM' function in the 'MuMIn' package (Barton, 2020).

2.3.5. Identification of SNPs correlated with population-level pox prevalence

We applied a Bayesian approach to identify SNPs strongly associated with differences in pox prevalence across populations. This was implemented in Bayenv v2.0, which first estimates a null model based on covariances of observed allele frequencies between populations that arise due to shared evolutionary history, and then assesses each SNP individually for linear correlations between population allele frequencies and environmental variables (Coop *et al.*, 2010; Günther & Coop, 2013). We used the ddRAD dataset, described above, and performed additional filtering steps in Plink v1.9 (Chang *et al.*, 2015) to generate a set of independent markers: a MAF threshold of 0.05 was applied to remove rare SNPs, and the remaining markers were filtered to remove loci in strong linkage disequilibrium (LD; Plink command: --indep-pairwise 50 5 0.5). PGDspider (v2.1.1.5; Lischer & Excoffier, 2012) was used to convert to a Bayenv input file format. Population covariance matrices were generated using this pruned marker set in ten replicate runs of Bayenv, each of 100,000 iterations and with different seed numbers, to ensure convergence. The last matrices of the ten independent runs were averaged to obtain a final, single covariance matrix, and the equivalent correlation matrix (Table S2.3) was compared to previously published pairwise F_{ST} values derived from both microsatellites and the ddRAD-seq SNPs (Illera *et al.*, 2007; Martin *et al.*, 2021) to ensure population structure was well estimated. In both cases, this matrix was consistent with the previous estimates of structure, indicating no problems in the labelling of populations.

To obtain best estimates of pox prevalence for each population and account for variability across different years due to natural fluctuations or sampling error, we used the complete field dataset (2005–2020). Prevalence was calculated as the total number of individuals caught with visible pox lesions across all sampling years, divided by the total number of individuals caught across all sampling years. One bird from La Palma that was originally identified as having pox (Spurgin *et al.*, 2012) was later revised after more experience diagnosing such infections, thus no confirmed cases of pox have been identified on La Palma. Prior to analyses with Bayenv, the population prevalence estimates were standardised to a mean of zero and variance of one.

Using the mean covariance matrix estimated above as a null model, we ran Bayenv for five independent replicates, on the same pruned set of SNPs. We report estimates for both Bayes factor (BF) values (measure of support for the alternative model in which the genotype shows a linear association with the tested environmental variable compared to the null model) and non-parametric Spearman's rank coefficients (ρ; strength of correlation between allele frequencies and the environmental variable). The latter served to reduce potential false positives as high BF values can also result from single outlying populations (Günther & Coop, 2013). Bayenv can also show high run-to-run variability (Blair *et al.*, 2014). We sought to reduce false positives due to this variance by testing the concordance between BF values and Spearman's ρ across the five replicate runs at different numbers of iterations (100,000, 200,000, and 500,000 iterations) and averaging these estimates across runs as advised by Blair *et al.* (2014). At 500,000 iterations, the correlation observed between runs consistently reached > 0.99 for Spearman's ρ and ranged from 0.08 – 0.53 for BF values. SNPs were therefore considered candidates if their average BF and average absolute value of Spearman's ρ across five replicate runs, each of 500,000 iterations, ranked in the highest 1% (BF $\geq$ 7.4) and 10%, respectively. BF values were interpreted according to a classification scheme adjusted from Jeffreys' scale of evidence: BF > 3, BF > 10, BF > 30, BF > 100, indicative of moderate, strong, very strong and extreme evidence of selection, respectively (Lee & Wagenmakers, 2013; modified from Jeffreys, 1961). To check whether BF values of candidate SNPs were driven by the overall correlation across the pox prevalence gradient rather than a single population, we also visually investigated the allele frequency structure of candidate SNPs across populations: MAFs for all candidate SNPs were calculated within each population using Plink.

We compiled a list of candidate genes located within 10 kbp up or downstream of each candidate SNP in the zebra finch assembly Taeniopygia_guttata-3.2.4 using NCBI Genome Data Viewer v5.1 (www.ncbi.nlm.nih.gov/genome/gdv/browser). Additionally, we used BEDTools v2.29.2 (Quinlan & Hall, 2010) and the genome positions of known immune-related genes within the same zebra finch assembly, given by Ekblom *et al.* (2010), to quantify the proportion of avian immune genes that will have been overlooked for associations with pox prevalence when using our marker set and methods. Throughout,

SNPs are identified by RAD-tag ID, followed by distance in base pairs from the start of the tag.

2.4. Results

2.4.1. Individual predictors of pox infection status

In Berthelot's pipits on Tenerife and Porto Santo, there was no significant association between age or sex and pox infection status (Table 2.1). However, there was a highly significant effect of malaria infection status and island identity (Table 2.1 and Figure 2.2; marginal R^2 = 0.16). Individuals infected with malaria had an increased likelihood of pox infection (i.e., 24.3%, compared to just 4.9% among individuals not infected with malaria) and pox prevalence was higher among individuals on Porto Santo (32.4%) than on Tenerife (6.4%). Thus, malaria infection status was retained as a predictor variable in subsequent genetic models.

There was no association between TLR4 protein haplotype, heterozygosity, or genotype and pox infection status within individuals on either island when controlling for malaria infection status (Table 2.2), nor when malaria was removed from the models (Table S2.5).

The GLM limited to individuals from Tenerife for which we had MHC variation information showed no association between MHC diversity, optimality, or individual alleles and pox infection status after controlling for malaria infection (Table 2.3) and when malaria was removed from the models (Table S2.6).

Table 2.1. Generalised linear mixed models (GLMM) testing individual-level predictors of pox infection status in Berthelot's pipits on Porto Santo and Tenerife (*n* = 778). Estimates and significance levels for each predictor represent the values upon re-entry into the minimal model. Those in bold were retained in the minimal model. Reference categories for each predictor is as follows: malaria infection status = not infected, island identity = Porto Santo, sex = female, and age = adult. Significant terms: *** = $p < 0.001$.

Fixed effects	**Estimate**	**Std. error**	**Z**	***p*-value**
Intercept	-1.678	0.342	-4.905	
Malaria	1.584	0.280	5.653	**< 0.001*** **
Island identity	-1.682	0.362	-4.645	**< 0.001*** **
Age	-0.123	0.406	-0.304	0.761
Sex	0.180	0.296	0.608	0.543
Random effects	**Variance**			
Sampling year	0.127	5 sampling years		

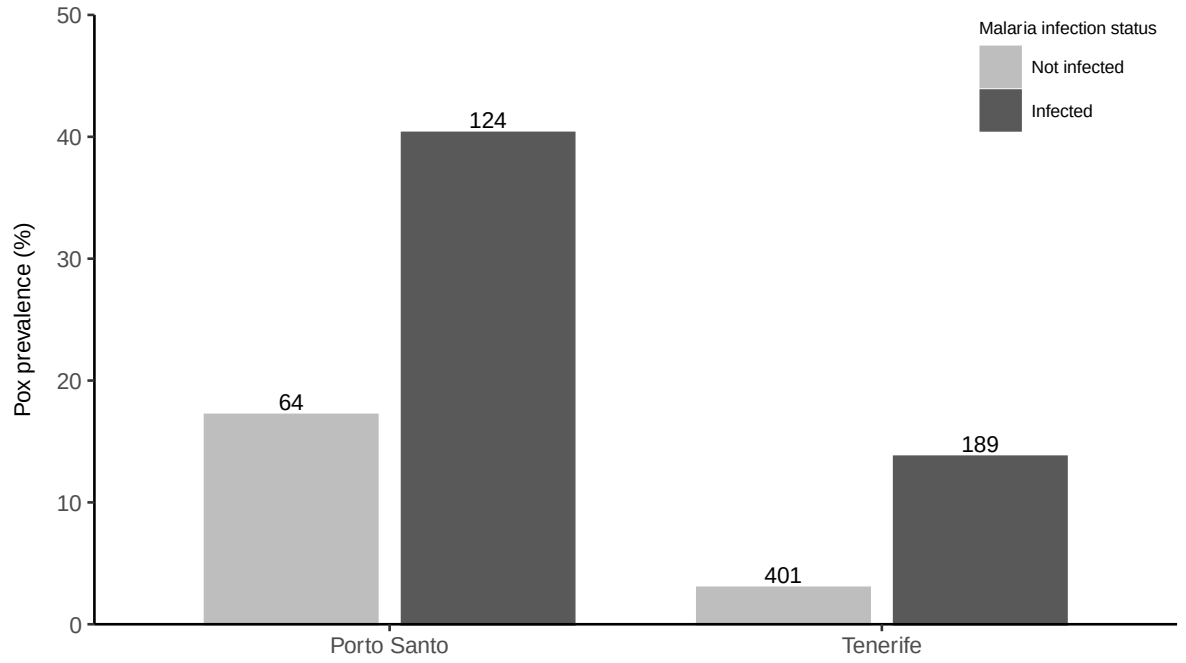

Figure 2.2. Prevalence of pox infection among Berthelot's pipits with and without malaria infection on Porto Santo and Tenerife. Numbers above the bars represent sample sizes.

Table 2.2. TLR4 variation in relation to pox infection status in Berthelot's pipits on Porto Santo and Tenerife. Generalised linear mixed models (GLMMs) were used to test for associations between (A) TLR4 protein haplotype (presence/absence), (B) TLR4 protein haplotype heterozygosity (homozygote/heterozygote), (C) TLR4 protein genotype, and pox infection status. Reference factor levels: malaria infection status = not infected, TLR_P = absence, TLR4_het = homozygote, and TLR4_genotype = 1,1. Significant terms: ** = $p < 0.01$, and *** = $p < 0.001$.

	Porto Santo					***Tenerife***				
Model A	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	Malaria	1.555	0.444	3.506	**< 0.001*** **	*Malaria*	1.660	0.383	4.334	**< 0.001*** **
	TLR4_P1	-0.164	0.506	-0.324	0.746	*TLR4_P1*	0.313	0.368	0.852	0.394
	TLR4_P2	-0.247	0.413	-0.598	0.550	*TLR4_P2*	0.074	0.661	0.112	0.911
	TLR4_P3	-0.141	0.381	-0.369	0.712					
	Random effects	**Variance**				**Random effects**	**Variance**			
	Sampling year	0.151	3 sampling years			*Sampling year*	0.096	4 sampling years		
	***n* = 184**					***n* = 578**				
Model B	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	Malaria	1.545	0.443	3.486	**< 0.001*** **	*Malaria*	1.655	0.383	4.327	**< 0.001*** **
	TLR4_het	0.040	0.348	0.116	0.908	*TLR4_het*	0.279	0.353	0.790	0.429
	Random effects	**Variance**				**Random effects**	**Variance**			
	Sampling year	0.151	3 sampling years			*Sampling year*	0.094	4 sampling years		
	***n* = 184**					***n* = 578**				
Model C	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	Malaria	1.539	0.486	3.165	**0.002****	*Malaria*	1.660	0.383	4.334	**< 0.001 *** **
	TLR4_genotype1,2	-0.187	0.511	-0.366	0.714	*TLR4_genotype1,2*	0.074	0.661	0.112	0.911
	TLR4_ genotype1,3	-0.005	0.444	-0.011	0.991	*TLR4_genotype2,2*	-0.239	0.663	-0.361	0.718
	TLR4_ genotype2,3	-0.054	0.577	-0.094	0.925					
	TLR4_ genotype3,3	-0.165	0.769	-0.215	0.830					
	Random effects	**Variance**				**Random effects**	**Variance**			
	Sampling year	0.188	3 sampling years			*Sampling year*	0.096	4 sampling years		
	***n* = 168**					***n* = 578**				

Table 2.3. Variation at MHC class I exon 3 in relation to pox infection status in Berthelot's pipits on Tenerife (*n* = 309). Generalised linear models (GLMs) were used to test for associations between (A) MHC diversity (number of alleles per individual, 3–10) and optimality (quadratic of MHC allele number), and (B) presence of specific MHC alleles (presence/absence), and pox infection status. Reference factor levels: malaria infection status = not infected, and ANBE = absence. Significant terms: ** = $p < 0.01$.

	Fixed effects	Estimate	Std. error	z	*p*-value
Model A	*Malaria*	1.787	0.570	3.138	**0.002****
	N.alleles	-1.403	1.233	-1.138	0.255
	N.alleles.squared	0.100	0.093	1.082	0.279
Model B	*Malaria*	1.937	0.594	3.260	**0.001****
	ANBE10	0.341	1.214	0.281	0.779
	ANBE8	-0.881	0.794	-1.109	0.267
	ANBE4	-0.059	0.572	-0.104	0.917
	ANBE43	-0.194	0.539	-0.360	0.719
	ANBE1	0.936	0.530	1.766	0.077
	ANBE44	0.548	0.765	0.716	0.474
	ANBE45	1.067	0.967	1.104	0.270
	ANBE9	1.029	0.680	1.513	0.130
	ANBE46	0.846	0.866	0.977	0.329
	ANBE47	-0.624	0.681	-0.917	0.359
	ANBE11	-0.408	0.705	-0.579	0.563
	ANBE6	-0.426	0.917	-0.465	0.642
	ANBE38	0.669	0.892	0.751	0.453

2.4.2. Signatures of pox-driven selection at the population level

The ddRAD library produced by Armstrong *et al.* (2018) contained 9,960 high-quality SNPs. After filtering SNPs with a MAF of less than 0.05, we retained 3,525. This dataset was further reduced by removing SNPs in strong LD to generate a set of independent markers for analysis in Bayenv, resulting in a final dataset of 2,334 SNPs.

Pox prevalence varied greatly among populations of Berthelot's pipit, ranging from 0% to 32.6% (shown in Table S2.4), but was broadly consistent within populations across the different sampling years (Figure S2.1). Population prevalence levels were highly correlated between 2006 and 2009 (2005 and 2009 for Selvagem Grande), when all populations had been sampled (Pearson correlation: $R = 0.72$, $p < 0.01$). We have never observed evidence of

pox infection in the Selvagens, or in three islands of the Canaries (El Hierro, La Palma and La Graciosa) and two islands of the Madeiran archipelago (Deserta Grande and Madeira). Yet, the third island of the Madeiran archipelago, Porto Santo, had the highest pox prevalence of all populations. The remaining Canary Islands generally showed a negative east-west gradient in pox prevalence (Figure 2.1), ranging from 25.6% to 5.4% (prevalence in the mountain population of El Teide was 1.9%).

Analysis with Bayenv identified 14 candidate SNPs (0.6% of total) where allele frequencies were highly associated with population-level pox prevalence, as indicated by their BF value and Spearman's ρ (Figure 2.3) across five independent runs. Raw allele frequency patterns for these candidate SNPs showed variable trends across the gradient of pox prevalence (note that Bayenv accounts for patterns of demography in observed allele frequencies but the raw data is shown in Figure 2.4). Within the Canary Islands, the top candidate SNP (444s109) generally showed higher MAFs as pox prevalence decreased. The minor allele was also absent in Porto Santo only (Madeiran archipelago; highest overall pox prevalence). In contrast, some SNPs (e.g., 3493s67) showed lower MAFs as pox prevalence decreased. For other SNPs, such as 2177s14 and 1796s91, the minor allele was very rare or absent within populations free of pox in the Canary Islands, but generally had a higher prevalence within populations where pox had been recorded.

Of the 14 SNPs identified by Bayenv, eight were located within annotated genes in the zebra finch genome (57.1%) identified using the NCBI genome browser. For three candidate SNPs, we identified genes within 10 kbp up- or downstream, and there were three SNPs that were not close to genes (21.4%; closest genes 27,722 – 201,990 bp from SNP). In total, we identified 12 candidate genes from 11 SNPs (while SNP 3439s47 was located within the gene *MFSD2A*, it was also only approximately 4,000 bp upstream of another gene, *MYCL*, so in this case both genes are identified). Ten of these identified genes have been implicated in cellular stress responses and/or have shown associations with viral infections (Table 2.4). Specific candidate SNPs that showed very strong evidence of selection (BF > 30) include those that were found approximately 6,000 bp from the gene *HSPA8* (heat shock protein 70 family member 8), within the gene *MTHFD1L* (methylenetetrahydrofolate dehydrogenase NADP+ dependent 1 like), and approximately 4,000 bp from the gene *ABLIM3* (actin binding

LIM protein family member 3). Of the immune genes identified by Ekblom *et al.* (2010) for which there was a mapped location, nine were located within 10 kbp of a SNP typed in the present study, and were therefore assessed for an association with pox (*ca*. 7%, Table S2.7).

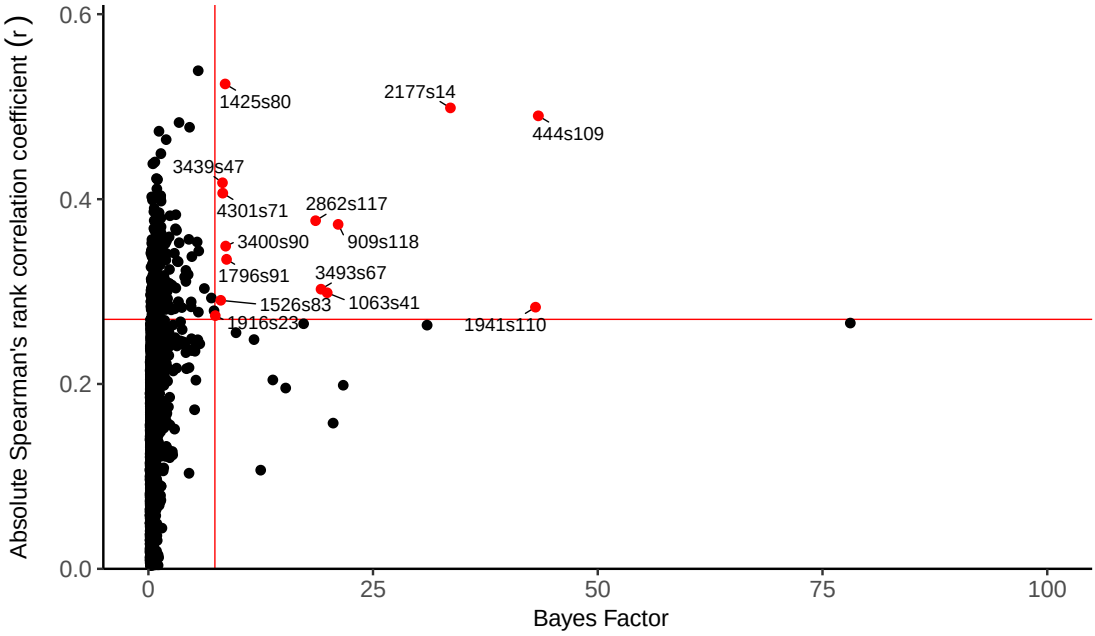

Figure 2.3. Bayes factor values versus absolute Spearman's rank correlation coefficients (ρ), averaged from five replicate runs, for genome-wide ddRAD SNPs among 13 Berthelot's pipit populations. SNPs were considered candidates for adaptation to population-level pox prevalence by Bayenv if they ranked in the highest 1% of Bayes factor values (≥ 7.4, threshold indicated by the vertical red line) and 10% of Spearman's ρ (threshold indicated by the horizontal red line). Fourteen SNPs were identified as candidates (those in red).

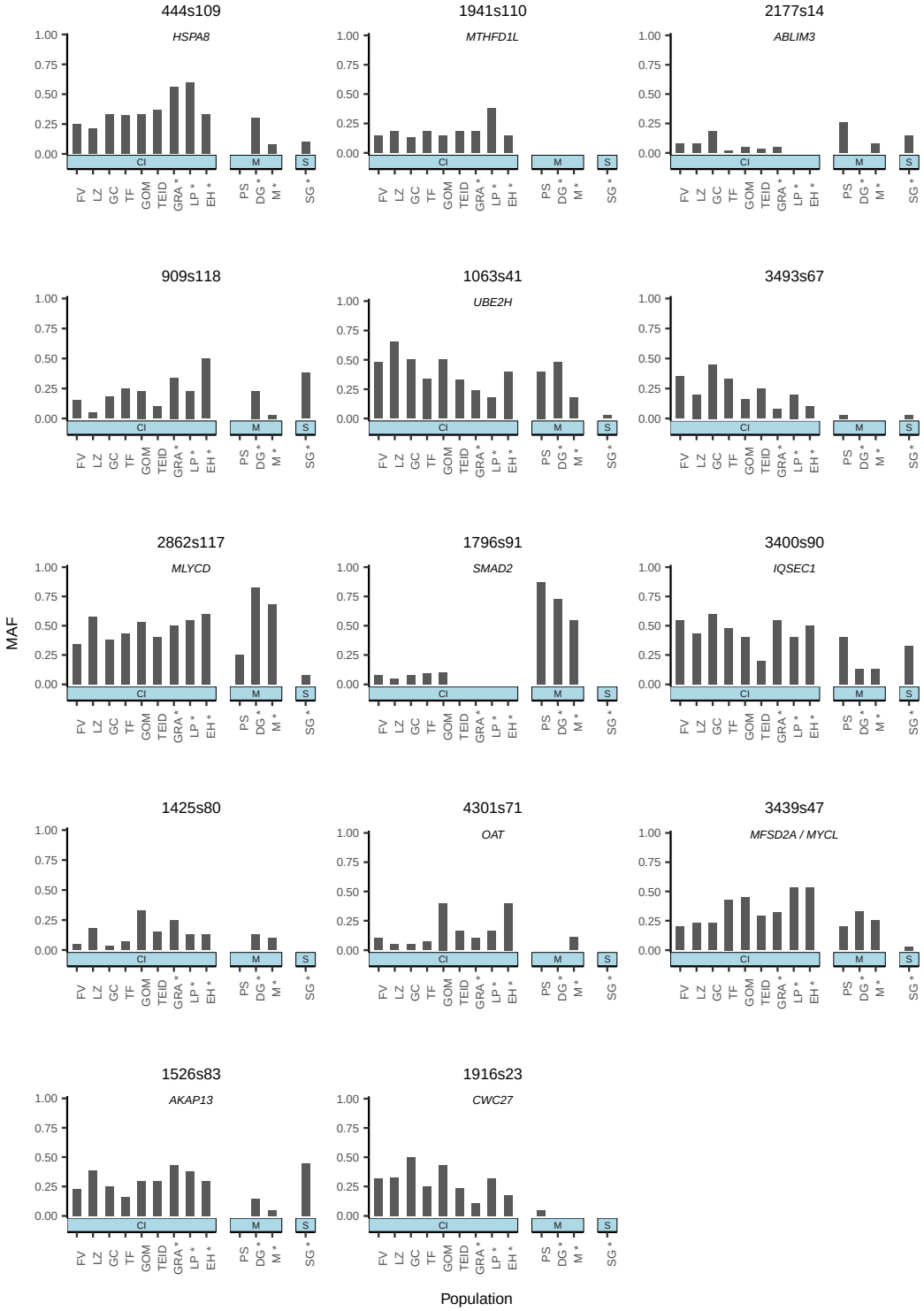

Figure 2.4. Minor allele frequency (MAF) distribution patterns of 14 candidate SNPs associated with population-level pox prevalence identified by Bayenv across populations of Berthelot's pipit. Nearby genes are noted below the SNP names. Populations are first grouped by archipelago (CI = Canary Islands; M = Madeira; S = Selvagens) and then ordered according to population-level pox prevalence (highest–lowest). Pox-free populations are indicated by an asterisk. Acronyms: FV = Fuerteventura; LZ = Lanzarote; GC = Gran Canaria; TF = Tenerife; GOM = La Gomera; TEID = Teide; GRA = La Graciosa; LP = La Palma; EH = El Hierro; PS = Porto Santo; DG = Deserta Grande; M = Madeira; SG = Selvagem Grande.

Table 2.4. SNPs identified as associated with population-level pox prevalence by Bayenv in Berthelot's pipits (those ranked in the highest 1% and 10% of Bayes factor (BF) values and Spearman's ρ respectively). Candidate genes identified within ± 10 kbp windows upstream (US) or downstream (DS) of the focal SNP.

SNP	Genomic location (Chr: bp)	BF	ρ	Candidate gene(s)	Gene name/description	Putative function	Evidence for role in the response to viral infection
444s109	24:3526218	43.39	0.49	*HSPA8* (6,319 bp US)	heat shock protein family A (Hsp70) member 8	ATP-dependent molecular chaperone that plays a role in protein folding processes	Hsp70 isoforms play a role in viral infection (Santoro *et al.*, 2009), upregulated during poxvirus infection (Jindal & Young, 1992; Sedger & Ruby, 1994; Brum *et al.*, 2003; Kowalczyk *et al.*, 2005; Cheng *et al.*, 2018).
1941s110	3:56769033	43.07	0.28	*MTHFD1L* (in gene)	methylenetetrahydrofolate dehydrogenase (NADP+ dependent) 1 like	Catalyses the synthesis of tetrahydrofolate in mitochondria in the folic acid cycle	Linked to avian influenza, may have regulatory role in replication (Zhang *et al.*, 2020).
2177s14	13:1724905	33.60	0.50	*ABLIM3* (4,111 bp US)	actin binding LIM protein family member 3	Interacts with actin filaments and may occur within adherens junctions	Silencing reduced the replication of hepatitis C virus (Blackham *et al.*, 2010).
909s118	21:5744237	21.12	0.37	–	–	–	–
1063s41	1A:217173	19.90	0.30	*UBE2H* (in gene)	ubiquitin conjugating enzyme E2 H	Catalyses the attachment of ubiquitin to other proteins in the ubiquitin/proteasome degradation pathway	Linked to herpes simplex virus (Lutz *et al.*, 2017) and identified in response to infection with Aleutian mink disease virus (Karimi *et al.*, 2021).

Table 2.4. (Continued).

SNP	Genomic location (Chr: bp)	BF	ρ	Candidate gene(s)	Gene name/description	Putative function	Evidence for role in the response to viral infection
3493s67	17:4549983	19.21	0.30	–	–	–	–
2862s117	11:1265660	18.60	0.38	*MLYCD* (3,598 bp DS)	malonyl-CoA decarboxylase	Catalyses the conversion of malonyl-CoA to acetyl-CoA during fatty acid metabolism	Highly expressed in cells infected with influenza virus (Coombs *et al.*, 2010; Dove *et al.*, 2012; Kroeker *et al.*, 2012).
1796s91	Z:421273	8.71	0.33	*SMAD2* (in gene)	SMAD family member 2	Transcriptional modulator and downstream effector of the transforming growth factor (TGF)-β signalling pathway	Linked to West Nile virus (Slonchak *et al.*, 2016), Epstein-Barr virus (Wood *et al.*, 2007), herpesvirus (Xu *et al.*, 2011), and polyomavirus (Sung *et al.*, 2014).
3400s90	12:1391473	8.60	0.35	*IQSEC1* (in gene)	IQ motif and sec7 domain ArfGEF 1	Guanine nucleotide exchange factor, involved in the regulation of ADP-ribosylation factor (ARF) protein signal transduction	–
1425s80	9:25142202	8.55	0.52	–	–	–	–
4301s71	6:32665728	8.28	0.41	*OAT* (in gene)	ornithine aminotransferase	Catalyses ornithine-glutamate interconversion during metabolism of arginine and glutamine	Decreased expression in influenza infected cells (Ding *et al.*, 2016) and herpes infected spat (Jouaux *et al.*, 2013).

Table 2.4. (Continued).

SNP	Genomic location (Chr: bp)	BF	ρ	Candidate gene(s)	Gene name/description	Putative function	Evidence for role in the response to viral infection
3439s47	23:3676275	8.26	0.42	*MFSD2A* (in gene) *MYCL* (4,164 bp US)	major facilitator superfamily domain containing 2A MYCL proto-oncogene, bHLH transcription factor	Sodium-dependent lysophosphatidylcholine transporter Transcription activity, regulating the expression of many pro-proliferative genes	Plays a role in the pathogenesis of Zika virus (Zhou *et al.*, 2019). –
1526s83	10:14176084	8.05	0.29	*AKAP13* (in gene)	A-kinase anchoring protein 13	Scaffold protein involved in assembling signalling complexes	Plays a role in the initiation of human immunodeficiency virus replication (König *et al.*, 2008).
1916s23	Z:50640938	7.45	0.27	*CWC27* (in gene)	CWC27 spliceosome associated cyclophilin	Pre-mRNA splicing factor recruited by the spliceosome	Many cyclophilins play a role in infection by diverse viruses (Frausto *et al.*, 2013), including some poxviruses (Castro *et al.*, 2003; Zhou *et al.*, 2021).

2.5. Discussion

We used both a candidate gene approach and landscape genomic approach to identify variation associated with spatial heterogeneity in avian pox prevalence within and across island populations. Linear modelling of individual-level infection data within two populations of Berthelot's pipit, found no evidence that variation in two previously identified candidate genes (TLR4 and MHC class I) was associated with pox infection. However, we consistently observed a positive association between pox and malaria infection. At the population-level across the species range using ddRAD-seq data, we identified 14 sites across the genome that showed pox-associated clinal patterns in allele frequency after controlling for population genetic structure. At these sites we identified 12 genes, many of which are involved in cellular stress response pathways and have been previously associated with infection by a range of viruses, including poxviruses, and in different animals, including humans (see Table 2.4).

Within-species variation in pox infection has previously been attributed to ecological (van Riper *et al.*, 2002; Samuel *et al.*, 2018), anthropic (Carrete *et al.*, 2009) and physiological factors (Zylberberg *et al.*, 2012). Here, we assessed predictors of pox infection at the individual-level using two populations of Berthelot's pipits so that we could control for these factors in later genetic models. We found no evidence for an effect of age or sex on pox infection status. Previous studies have shown little indication that sex influences pox infection prevalence (Ruiz-Martínez *et al.*, 2016; Samuel *et al.*, 2018). Prevalence is commonly reported to be higher in juvenile birds than in adults (Gortázar *et al.*, 2002; Buenestado *et al.*, 2004; Ruiz-Martínez *et al.*, 2016), which authors have linked to immunological naivety. However, opposite patterns have also been observed (Atkinson *et al.*, 2005; Smits *et al.*, 2005).

Malaria infection was the most significant predictor of pox infection in our system. This result adds to a growing body of evidence that avian pox and malaria infections are not independent among individuals, but instead show positive associations, as previously documented in native Hawaiian birds (Atkinson *et al.*, 2005; Atkinson & Samuel, 2010; Samuel *et al.*, 2018) and Berthelot's pipits with a much smaller dataset (Spurgin *et al.*, 2012).

Whether this association is due to simultaneous vector-borne transmission, reduced immunity following infection by the first pathogen and therefore susceptibility to secondary infections, or differential mortality among individuals with singular and concomitant infections needs further investigation. It is possible the pipit host could contain variants at key genes that confer shared resistance/susceptibility to both pathogens. However, different taxa of pathogens (i.e., virus versus protist) are normally recognised by different receptors, e.g., TLR3 specifically recognises viral DNA (Barton, 2007), though other components of the immune defences may be shared. In the present study, we did not detect any SNPs that had previously been identified as associated with malaria in Berthelot's pipits (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2018, 2019; discussed further below). Alternatively, the association between avian pox and malaria might reflect increased likelihood of exposure to both pathogens. For example, artificial water sources and poultry farms were associated with increased local prevalence of malaria in the pipit system (González-Quevedo *et al.*, 2014), and in multiple birds species across Tenerife (Padilla *et al.*, 2017), probably due to the increased local density of vectors and hosts at these sites. The same could be true for pox prevalence, but as of yet no landscape-scale study has investigated the factors driving local pox prevalence in Berthelot's pipit. Though, similar effects of animal husbandry on poxvirus infection rates have been documented in the short-toed lark in the Canary Islands (Carrete *et al.*, 2009).

We examined whether host genetic variation could explain the observed variation in patterns of infection among individuals. Both TLR4 and MHC class I were considered potential candidates that may interact with poxvirus based on evidence of selection at these loci in Berthelot's pipit (González-Quevedo, Phillips, *et al.*, 2015; González-Quevedo, Spurgin, *et al.*, 2015), and their involvement in the immune pathways associated with poxvirus pathogenesis (Guerin *et al.*, 2002; Hutchens *et al.*, 2008). Some variants included in this study were previously found to be associated with malaria infection and risk among pipits in Porto Santo and Tenerife (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2019). However, after controlling for other predictors mentioned above, we did not find any evidence for an association between pox infections and TLR4 heterozygosity, or individual protein haplotypes or genotypes. Nor did we find evidence for a link with MHC diversity, optimality, or individual alleles as would be expected under different scenarios of heterozygote

advantage, fluctuating selection or rare allele advantage, and optimized heterozygosity. Given we found such a strong association between pox and malaria infection, it is perhaps even more surprising we did not find that any candidate variants were associated with pox infection. However, some difficulty lies with the complexity of classifying individual infection status. For example, individuals with highly susceptible genotypes may be included in the set of non-infected individuals due to them never being exposed to *Avipoxvirus*. Also, most individuals we identified with pox are perhaps just those individuals that have successfully survived pox, rather than those particularly susceptible to pox. Both of these factors could obscure association data. Nevertheless, these loci may not play a direct role in pox infection. Unfortunately, we could not assess these genes for population-level associations with pox prevalence because there were no markers close enough to these genes in our ddRAD dataset. The closest SNP (SNP 3201s69) to the TLR4 gene was approximately 30 kbp away. Passerine MHC genes are difficult to map (He *et al.*, 2021); MHC class I genes have been identified on chromosomes 16 and 22 in zebra finch (Balakrishnan *et al.*, 2010; Ekblom *et al.*, 2011) but exact locations are unknown.

The candidate gene approach is also limited to key genes with an already hypothesised role in the host's response to a particular pathogen, thereby excluding the possibility of identifying new genes, especially those that participate in hitherto unknown mechanisms underlying host-pathogen interactions. That said, the concentrated sampling effort, largely within the same population, focused on a single or few gene(s), tends to offer higher statistical power compared to genome-wide approaches (Amos *et al.*, 2011). Therefore, it is possible to detect population-specific associations between genetic variants and infection with pathogens with such an approach (e.g., Bonneaud *et al.*, 2006). Further, candidate gene approaches allow for thorough investigation of highly polymorphic loci, where we might expect alleles to differ across populations, while the genome-wide population-level Bayenv approach only used SNPs that are consistent across populations. Accordingly, the candidate gene approach is very useful, but may be more relevant once further candidates – potentially those involved in polygenic responses or understudied mechanisms – have been identified using genome scans or GWAS.

We identified novel candidate loci associated with population-level of avian pox prevalence using a ddRAD-seq marker set. The highest-ranking SNP identified (444s109) was located on chromosome 24, *ca.* 6,000 bp from the gene *HSPA8*. This gene encodes a member of the heat shock protein 70 (Hsp70) family – molecular chaperones that assist in protein folding, degradation, and trafficking (Mayer & Bukau, 2005; Kampinga & Craig, 2010). During viral infection, the cellular heat shock response is induced and Hsp70 genes are upregulated (e.g., Brum *et al.*, 2003; Burch & Weller, 2004; Manzoor *et al.*, 2014; Howe *et al.*, 2016), however the role these genes serve remains unclear. Although Hsp70 proteins are essential for protecting cells from stress, stabilising the cell's own proteins, preventing viral replication (Li *et al.*, 2011), and signalling to the innate immune system (Kim *et al.*, 2013), they may also support viral genome replication (Ye *et al.*, 2013; Manzoor *et al.*, 2014) or be exploited as molecular chaperones to process or stabilise viral proteins (Taguwa *et al.*, 2015; Zhang *et al.*, 2015). Indeed, Hsp70 appears to play a role in poxvirus replication (Jindal & Young, 1992; Cheng *et al.*, 2018) and in the suppression of apoptosis, which lengthens the time the poxvirus has for replication (Kowalczyk *et al.*, 2005). Given the evidence above, *HSPA8* should be considered a strong candidate for involvement in interactions between hosts and avian poxvirus, but further work is needed to understand the mechanistic basis for how variation at this locus affects infection.

In addition to *HSPA8*, many of the genes linked to SNPs identified in our study are involved in enzymatic pathways and cell signalling transduction and have been linked to viral infection. Poxviruses, and other large viruses such as herpesvirus, have dedicated much of their genomes to encoding proteins that allow them to subvert anti-viral mechanisms and regulatory controls (reviewed in McFadden & Murphy, 2000; Smith & Kotwal, 2002; Seet *et al.*, 2003; Leão & Fonseca, 2014). These include proteins that mimic extracellular host immune molecules and block the innate immune response (Kotwal & Moss, 1988; Alcami & Smith, 1992; Alcami, 2007; Mann *et al.*, 2008). Other proteins may mask signals between the infected cell and the acquired immune system, (Boshkov *et al.*, 1992; Guerin *et al.*, 2002), or interfere with intracellular pathways such as signalling from cytokines, the ubiquitin pathway (Zhang *et al.*, 2009), and other processes that promote cell death (reviewed in Nichols *et al.*, 2017). Among the avian pox-associated loci in this study, we specifically found genes that are involved in ubiquitin pathways and cytokine signalling: *UBE2H*, which likely catalyses the

modification of proteins for degradation (Stewart *et al.*, 2016), and *SMAD2*, a downstream effector of the transforming growth factor (TGF)-β signalling pathway, which regulates the transcription of target genes including those leading to apoptosis (Derynck *et al.*, 1998). Interestingly, the genome of avian fowl pox virus contains a putative homolog of the eukaryotic TGF-β gene, which is unique to this genera of poxviruses and is likely to have immunomodulatory effects (Afonso *et al.*, 2000). Thus, our findings fit well with the known mechanisms of immune evasion employed by poxviruses and their interactions with host proteins.

Despite the strong association we observed between pox and malaria infection at the individual-level, we could not control for malaria prevalence in the population-level analyses. However, none of the 14 candidate SNPs identified in this study were located near (within 10 kbp) the malaria-associated SNPs previously identified using the same genome-wide dataset for Berthelot's pipit and an EigenGWAS analysis (Armstrong *et al.*, 2018). Clearly further work is needed to explore the interacting effects of avian pox and malaria infection and to identify which specific candidate genes respond to either disease agent. This is an important question if we aim to understand the mechanisms and agents driving genetic variation in these genes.

When identifying candidate SNPs, we applied strict criteria to reduce false positive associations with pox prevalence. We acknowledge that this could also mean that other avian pox-associated loci may have gone undetected. That our Bayenv analyses detected 14 candidate loci that were strongly (BF > 7.4) associated with avian pox prevalence after using the stringent cut-offs suggests that this pathogen may have influence on genetic variation within – and structuring among – populations of Berthelot's pipit. While RAD sequencing is a cost-effective method for rapidly genotyping large numbers of polymorphisms, such reduced representation sequencing approaches are only able to assess a small portion of the genome for genotype-environment associations. Indeed, we were only able to evaluate *ca.* 7% of the previously identified avian immune-related genes (Ekblom *et al.*, 2010) for associations with pox prevalence. However, this is a consideration for any genome scan or genome-wide association study (GWAS) that does not use high resolution whole-genome sequencing. It does not undermine the validity of the associations that we do detect. The aim of this study

was not to identify every possible correlation between allele frequencies and pox prevalence, but rather, to identify some strong candidate SNPs for future study of pathogen-mediated selection in wild bird populations. Generally, the approach applied in this study should be considered just one of many complementary tools that can be used in the search for genes involved in host-pathogen interactions; it is by no means exhaustive.

Further studies in Berthelot's pipits and/or other pox-infected avian populations are needed to validate the candidate loci identified in the current study and to identify other loci under selection from avian pox. In Berthelot's pipit, spatial variation in pox prevalence is independent of neutral population structure across the range, but is shaped by certain biogeographical factors (Spurgin *et al.*, 2012), that is, smaller and more isolated islands are less likely to be infected with pox. This suggests that biogeographical factors, and other environmental factors, largely determine patterns of pox prevalence across populations. Finally, the Bayenv analysis is specifically designed to control for neutral population structure. Therefore, it is unlikely that the associations we observed between avian pox prevalence and specific SNPs were due to the prior distribution of susceptibility alleles caused by demographic history; rather associations are likely to be a consequence of pathogen-mediated selection pressure. Nevertheless, studies assessing the infection status, health and survival of a large number of genotyped individuals within pox-infected populations are necessary to verify the role of the candidate loci in pox resistance. Ultimately, these tests provide evidence of an association, not necessarily a causal link and thus, an empirical demonstration of their function is required, either using gene expression profiling among infected and non-infected individuals or a direct assay (Pardo-Diaz *et al.*, 2015). We acknowledge this is more difficult for a non-model organism and endemic species such as Berthelot's pipit. Testing for signatures of selection at candidate SNPs using resequencing of individual birds would also corroborate whether selection is acting on specific genes. In addition, resequencing candidates across temporal samples either across generations in contemporary populations (e.g., Davies *et al.*, 2021) or over longer time spans using museum samples (e.g., Alves *et al.*, 2019) could enable inference of selection and may help to elucidate the mechanisms of adaptation by testing for allele frequency changes at these loci (Bank *et al.*, 2014; Malaspinas, 2016).

2.5.1. Conclusions

Our study is one of the first to attempt to identify loci involved in avian poxvirus response across wild passerine populations. Using two different approaches, we were able to test for associations between disease and host genetic variation at both an individual and population level. Genotype-environment associations were detected across populations of Berthelot's pipits exposed to different pox prevalence levels after controlling for demographic history/neutral genetic population structure. Thus, the results potentially evidence the evolution of hosts in response to local pathogen pressure, but this needs to be confirmed in the future using extensive within-population level analyses to link variants to the presence/absence of pox and/or individual survival. We identified novel pox-associated genes involved in cellular stress signalling and immune responses, such as the heat shock response, cytokine signalling pathways, and apoptosis. We suggest that these genes represent promising new candidates with which to understand pathogen-mediated selection in wild bird populations.

2.6. References

Afonso, C. L., Tulman, E. R., Lu, Z., Zsak, L., Kutish, G. F., & Rock, D. L. (2000). The genome of fowlpox virus. *Journal of Virology*, *74*, 3815–3831.

Alcami, A. (2007). New insights into the subversion of the chemokine system by poxviruses. *European Journal of Immunology*, *37*, 880–883.

Alcami, A., & Smith, G. L. (1992). A soluble receptor for interleukin-1β encoded by vaccinia virus: A novel mechanism of virus modulation of the host response to infection. *Cell*, *71*, 153–167.

Alley, M. R., Ha, H. J., Howe, L., Hale, K. A., & Cash, W. (2010). Concurrent avian malaria and avipox virus infection in translocated South Island saddlebacks (*Philesturnus carunculatus carunculatus*). *New Zealand Veterinary Journal*, *58*, 218–223.

Alves, J. M., Carneiro, M., Cheng, J. Y., de Matos, A. L., Rahman, M. M., Loog, L., Campos, P. F., *et al.* (2019). Parallel adaptation of rabbit populations to myxoma virus. *Science*, *363*, 1319–1326.

Amos, W., Driscoll, E., & Hoffman, J. I. (2011). Candidate genes versus genome-wide associations: which are better for detecting genetic susceptibility to infectious disease? *Proceedings of the Royal Society B: Biological Sciences*, *278*, 1183–1188.

Anderson, R. M., & May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, *280*, 361–367.

Apanius, V., Penn, D., Slev, P. R., Ruff, L. R., & Potts, W. K. (1997). The nature of selection on the major histocompatibility complex. *Critical Reviews in Immunology*, *17*, 179–224.

Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution*, *9*, 12482–12502.

Armstrong, C., Richardson, D. S., Hipperson, H., Horsburgh, G. J., Küpper, C., Percival-Alwyn, L., Clark, M., *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., Lease, J. K., Dusek, R. J., & Samuel, M. D. (2005). Prevalence of pox-like lesions and malaria in forest bird communities on Leeward Mauna Loa Volcano, Hawaii. *The Condor*, *107*, 537–546.

Atkinson, C. T., & Samuel, M. D. (2010). Avian malaria *Plasmodium relictum* in native Hawaiian forest birds: epizootiology and demographic impacts on àapapane *Himatione sanguinea*. *Journal of Avian Biology*, *41*, 357–366.

Baek, H. E., Bandivadekar, R. R., Pandit, P., Mah, M., Sehgal, R. N. M., & Tell, L. A. (2020). TaqMan quantitative real-time PCR for detecting *Avipoxvirus* DNA in various sample types from hummingbirds. *PLoS ONE*, *15*, e0230701.

Balakrishnan, C. N., Ekblom, R., Völker, M., Westerdahl, H., Godinez, R., Kotkiewicz, H., Burt, D. W., *et al.* (2010). Gene duplication and fragmentation in the zebra finch major histocompatibility complex. *BMC Biology*, *8*, 29.

Bank, C., Ewing, G. B., Ferrer-Admettla, A., Foll, M., & Jensen, J. D. (2014). Thinking too positive? Revisiting current methods of population genetic selection inference. *Trends in Genetics*, *30*, 540–546.

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

Barton, K. (2020). MuMIn: Multi-Model Inference. R package version 1.43.17. Available at: https://cran.r-project.org/package=MuMIn.

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

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

Biswas, S., & Akey, J. M. (2006). Genomic insights into positive selection. *Trends in Genetics*, *22*, 437–446.

Blackham, S., Baillie, A., Al-Hababi, F., Remlinger, K., You, S., Hamatake, R., & McGarvey, M. J. (2010). Gene expression profiling indicates the roles of host oxidative stress, apoptosis, lipid metabolism, and intracellular transport genes in the replication of hepatitis C virus. *Journal of Virology*, *84*, 5404–5414.

Blair, L. M., Granka, J. M., & Feldman, M. W. (2014). On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, *8*, 1.

Bolker, B. M., Brooks, M. E., Clark, C. J., Geange, S. W., Poulsen, J. R., Stevens, M. H. H., &

White, J. S. S. (2009). Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, *24*, 127–135.

Bolte, A. L., Meurer, J., & Kaleta, E. F. (1999). Avian host spectrum of avipoxviruses. *Avian Pathology*, *28*, 415–432.

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

Boshkov, L. K., Macen, J. L., & McFadden, G. (1992). Virus-induced loss of class I MHC antigens from the surface of cells infected with myxoma virus and malignant rabbit fibroma virus. *The Journal of Immunology*, *148*, 881–887.

Brum, L. M., Lopez, M. C., Varela, J. C., Baker, H. V., & Moyer, R. W. (2003). Microarray analysis of A549 cells infected with rabbitpox virus (RPV): A comparison of wild-type RPV and RPV deleted for the host range gene, SPI-1. *Virology*, *315*, 322–334.

Buenestado, F., Gortázar, C., Millán, J., Höfle, U., & Villafuerte, R. (2004). Descriptive study of an avian pox outbreak in wild red-legged partridges (*Alectoris rufa*) in Spain. *Epidemiology and Infection*, *132*, 369–374.

Burch, A. D., & Weller, S. K. (2004). Nuclear sequestration of cellular chaperone and proteasomal machinery during herpes simplex virus type 1 infection. *Journal of Virology*, *78*, 7175–7185.

Carrete, M., Serrano, D., Illera, J. C., López, G., Vögeli, M., Delgado, A., & Tella, J. L. (2009). Goats, birds, and emergent diseases: apparent and hidden effects of exotic species in an island environment. *Ecological Applications*, *19*, 840–853.

Castro, A. P. V., Carvalho, T. M. U., Moussatché, N., & Damaso, C. R. A. (2003). Redistribution of cyclophilin A to viral factories during vaccinia virus infection and its incorporation into mature particles. *Journal of Virology*, *77*, 9052–9068.

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

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*, e64.

Cheng, W., Jia, H., Wang, X., He, X., Jin, Q., Cao, J., & Jing, Z. (2018). Ectromelia virus upregulates the expression of heat shock protein 70 to promote viral replication. *International Journal of Molecular Medicine*, *42*, 1044–1053.

Coombs, K. M., Berard, A., Xu, W., Krokhin, O., Meng, X., Cortens, J. P., Kobasa, D., *et al.* (2010). Quantitative proteomic analyses of influenza virus-infected cultured human lung cells. *Journal of Virology*, *84*, 10888–10906.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

Cramp, S. (1985). *The Birds of the Western Palearctic* (Vol. 5). Oxford: Oxford University Press.

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.

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, *287*, 443–449.

Davidson, W. R., Kellogg, F. E., & Doster, G. L. (1980). An epornitic of avian pox in wild bobwhite quail. *Journal of Wildlife Diseases*, *16*, 293–298.

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

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

Derynck, R., Zhang, Y., & Feng, X.-H. (1998). Transcriptional activators of TGF-β responses: Smads. *Cell*, *95*, 737–740.

Ding, X., Lu, J., Yu, R., Wang, X., Wang, T., Dong, F., Peng, B., *et al.* (2016). Preliminary proteomic analysis of A549 cells infected with avian influenza virus H7N9 and influenza A virus H1N1. *PLoS ONE*, *11*, e0156017.

Docherty, D. E., Long, R. I., Flickinger, E. L., & Locke, L. N. (1991). Isolation of poxvirus from debilitating cutaneous lesions on four immature grackles (*Quiscalus* sp.). *Avian Diseases*, *35*, 244–247.

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

Dove, B. K., Surtees, R., Bean, T. J. H., Munday, D., Wise, H. M., Digard, P., Carroll, M. W., *et al.* (2012). A quantitative proteomic analysis of lung epithelial (A549) cells infected with 2009 pandemic influenza A virus using stable isotope labelling with amino acids in cell culture. *Proteomics*, *12*, 1431–1436.

Dudaniec, R. Y., Yong, C. J., Lancaster, L. T., Svensson, E. I., & Hansson, B. (2018). Signatures of local adaptation along environmental gradients in a range-expanding damselfly (*Ischnura elegans*). *Molecular Ecology*, *27*, 2576–2593.

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

Ekblom, R., Stapley, J., Ball, A. D., Birkhead, T., Burke, T., & Slate, J. (2011). Genetic mapping of the major histocompatibility complex in the zebra finch (*Taeniopygia guttata*). *Immunogenetics*, *63*, 523–530.

Farias, M. E. M., la Pointe, D. A., Atkinson, C. T., Czerwonka, C., Shrestha, R., & Jarvi, S. I. (2010). Taqman real-time PCR detects *Avipoxvirus* DNA in blood of Hawaìi `Amakihi (*Hemignathus virens*). *PLoS ONE*, *5*, e10745.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, *87*, 178–183.

Fox, J., & Weisberg, S. (2019). *An R Companion to Applied Regression* (3rd edn). California:

Sage, Thousand Oaks.

Fraik, A. K., Margres, M. J., Epstein, B., Barbosa, S., Jones, M., Hendricks, S., Schönfeld, B., *et al.* (2020). Disease swamps molecular signatures of genetic-environmental associations to abiotic factors in Tasmanian devil (*Sarcophilus harrisii*) populations. *Evolution*, *74*, 1392–1408.

Frausto, S. D., Lee, E., & Tang, H. (2013). Cyclophilins as modulators of viral replication. *Viruses*, *5*, 1684–1701.

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., 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. (2015). 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. (2015). Drift, not selection, shapes toll-like receptor variation among oceanic island populations. *Molecular Ecology*, *24*, 5852–5863.

Gortázar, C., Millán, J., Höfle, U., Buenestado, F. J., Villafuerte, R., & Kaleta, E. F. (2002). Pathology of avian pox in wild red-legged partridges (*Alectoris rufa*) in Spain. *Annals of the New York Academy of Sciences*, *969*, 354–357.

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., 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.

Guerin, J.-L., Gelfi, J., Boullier, S., Delverdier, M., Bellanger, F.-A., Bertagnoli, S., Drexler, I., *et al.* (2002). Myxoma virus leukemia-associated protein is responsible for major histocompatibility complex class I and Fas-CD95 down-regulation and defines scrapins, a new group of surface cellular receptor abductor proteins. *Journal of Virology*, *76*, 2912–2923.

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*, 205–220.

Gyuranecz, M., Foster, J. T., Dan, A., Ip, H. S., Egstad, K. F., Parker, P. G., Higashiguchi, J. M., *et al.* (2013). Worldwide phylogenetic relationship of avian poxviruses. *Journal of Virology*, *87*, 4938–4951.

He, K., Minias, P., & Dunn, P. O. (2021). Long-read genome assemblies reveal extraordinary variation in the number and structure of MHC loci in birds. *Genome Biology and Evolution*, *13*, evaa270.

Hewitt, E. W. (2003). The MHC class I antigen presentation pathway: strategies for viral

immune evasion. *Immunology*, *110*, 163–169.

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

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., *et al.* (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *American Naturalist*, *188*, 379–397.

Howe, M. K., Speer, B. L., Hughes, P. F., Loiselle, D. R., Vasudevan, S., & Haystead, T. A. J. (2016). An inducible heat shock protein 70 small molecule inhibitor demonstrates anti-dengue virus activity, validating Hsp70 as a host antiviral target. *Antiviral Research*, *130*, 81–92.

Hudson, P. J., Dobson, A. P., & Newborn, D. (1998). Prevention of population cycles by parasite removal. *Science*, *282*, 2256–2258.

Hutchens, M. A., Luker, K. E., Sonstein, J., Núñez, G., Curtis, J. L., & Luker, G. D. (2008). Protective effect of Toll-like receptor 4 in pulmonary vaccinia infection. *PLoS Pathogens*, *4*, 1000153.

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.

Jeffreys, H. (1961). *Theory of Probability* (3rd edn). Oxford: Oxford University Press.

Jindal, S., & Young, R. A. (1992). Vaccinia virus infection induces a stress response that leads to association of Hsp70 with viral proteins. *Journal of Virology*, *66*, 5357–5362.

Jouaux, A., Lafont, M., Blin, J.-L., Houssin, M., Mathieu, M., & Lelong, C. (2013). Physiological change under OsHV-1 contamination in Pacific oyster *Crassostrea gigas* through massive mortality events on fields. *BMC Genomics*, *14*, 590.

Kampinga, H. H., & Craig, E. A. (2010). The HSP70 chaperone machinery: J proteins as drivers of functional specificity. *Nature Reviews Molecular Cell Biology*, *11*, 579–592.

Karimi, K., Farid, A. H., Myles, S., & Miar, Y. (2021). Detection of selection signatures for response to Aleutian mink disease virus infection in American mink. *Scientific Reports*, *11*, 2944.

Kim, M. Y., Shu, Y., Carsillo, T., Zhang, J., Yu, L., Peterson, C., Longhi, S., *et al.* (2013). Hsp70 and a novel axis of type I interferon-dependent antiviral immunity in the measles virus-infected brain. *Journal of Virology*, *87*, 998–1009.

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

König, R., Zhou, Y., Elleder, D., Diamond, T. L., Bonamy, G. M. C., Irelan, J. T., Chiang, C. -Y., *et al.* (2008). Global analysis of host-pathogen interactions that regulate early-stage HIV-1 replication. *Cell*, *135*, 49–60.

Kotwal, G. J., & Moss, B. (1988). Vaccinia virus encodes a secretory polypeptide structurally related to complement control proteins. *Nature*, *335*, 176–178.

Kowalczyk, A., Guzik, K., Slezak, K., Dziedzic, J., & Rokita, H. (2005). Heat shock protein and heat shock factor 1 expression and localization in vaccinia virus infected human monocyte derived macrophages. *Journal of Inflammation*, *2*, 12.

Kroeker, A. L., Ezzati, P., Halayko, A. J., & Coombs, K. M. (2012). Response of primary human airway epithelial cells to influenza infection: A quantitative proteomic study. *Journal of Proteome Research*, *11*, 4132–4146.

Lachish, S., Bonsall, M. B., Lawson, B., Cunningham, A. A., & Sheldon, B. C. (2012). Individual and population-level impacts of an emerging poxvirus disease in a wild population of great tits. *PLoS ONE*, *7*, 48545.

Laiolo, P., Serrano, D., Tella, J. L., Carrete, M., Lopez, G., & Navarro, C. (2007). Distress calls reflect poxvirus infection in lesser short-toed lark *Calandrella rufescens*. *Behavioral Ecology*, *18*, 507–512.

Lawson, B., Lachish, S., Colvile, K. M., Durrant, C., Peck, K. M., Toms, M. P., Sheldon, B. C., *et al.* (2012). Emergence of a novel avian pox disease in British tit species. *PLoS ONE*, *7*, 40176.

Leão, T. L., & Fonseca, F. G. da. (2014). Subversion of cellular stress responses by poxviruses. *World Journal of Clinical Infectious Diseases*, *4*, 27–40.

Lee, M. D., & Wagenmakers, E.-J. (2013). *Bayesian Cognitive Modeling: A Practical Course*. Cambridge: Cambridge University Press.

Lei, L., Poets, A. M., Liu, C., Wyant, S. R., Hoffman, P. J., Carter, C. K., Shaw, B. G., *et al.* (2019). Environmental association identifies candidates for tolerance to low temperature and drought. *G3: Genes, Genomes, Genetics*, *9*, 3423–3438.

Li, G., Zhang, J., Tong, X., Liu, W., & Ye, X. (2011). Heat shock protein 70 inhibits the activity of influenza a virus ribonucleoprotein and blocks the replication of virus In Vitro and In Vivo. *PLoS ONE*, *6*, e16546.

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

Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, *28*, 298–299.

Lutz, G., Jurak, I., Kim, E. T., Kim, J. Y., Hackenberg, M., Leader, A., Stoller, M. L., *et al.* (2017). Viral ubiquitin ligase stimulates selective host microRNA expression by targeting ZEB transcriptional repressors. *Viruses*, *9*, 210.

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

Malaspinas, A. S. (2016). Methods to characterize selective sweeps using time serial samples: An ancient DNA perspective. *Molecular Ecology*, *25*, 24–41.

Mann, B. A., Huang, J. H., Li, P., Chang, H. C., Slee, R. B., O'Sullivan, A., Mathur, A., *et al.* (2008). Vaccinia virus blocks Stat1-dependent and Stat1-independent gene expression

induced by type I and type II interferons. *Journal of Interferon and Cytokine Research*, *28*, 367–379.

Manzoor, R., Kuroda, K., Yoshida, R., Tsuda, Y., Fujikura, D., Miyamoto, H., Kajihara, M., *et al.* (2014). Heat shock protein 70 modulates influenza A virus polymerase activity. *Journal of Biological Chemistry*, *289*, 7599–7614.

Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and within-archipelago adaptation between island bird populations. *Royal Society Open Science*, *8*, 201146.

Mayer, M. P., & Bukau, B. (2005). Hsp70 chaperones: Cellular functions and molecular mechanism. *Cellular and Molecular Life Sciences*, *62*, 670–684.

McFadden, G., & Murphy, P. M. (2000). Host-related immunomodulators encoded by poxviruses and herpesviruses. *Current Opinion in Microbiology*, *3*, 371–378.

Medina, F. M., Ramírez, G. A., & Hernández, A. (2004). Avian pox in white-tailed laurel-pigeons from the Canary Islands. *Journal of Wildlife Diseases*, *40*, 351–355.

Mundry, R., & Nunn, C. L. (2009). Stepwise model fitting and statistical inference: turning noise into signal pollution. *The American Naturalist*, *173*, 119–123.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, *14*, 20170213.

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

Nichols, D. B., De Martini, W., & Cottrell, J. (2017). Poxviruses utilize multiple strategies to inhibit apoptosis. *Viruses*, *9*, 215.

O'Brien, S. J., & Evermann, J. F. (1988). Interactive influence of infectious disease and genetic diversity in natural populations. *Trends in Ecology & Evolution*, *3*, 254–259.

Orós, J., Rodríguez, F., Rodríguez, J. L., Bravo, C., & Fernández, A. (1997). Debilitating cutaneous poxvirus infection in a Hodgson's grandala (*Grandala coelicolor*). *Avian Diseases*, *41*, 481–483.

Ortego, J., Aparicio, J. M., Calabuig, G., & Cordero, P. J. (2007). Risk of ectoparasitism and genetic diversity in a wild lesser kestrel population. *Molecular Ecology*, *16*, 3712–3720.

Padilla, D. P., Illera, J. C., Gonzalez-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.

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*, 445–464.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., Quail, M., *et al.* (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*, 275–278.

Pritchard, J., Pickrell, J., & Coop, G. (2010). The genetics of human adaptation: Hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*, *20*, 208–215.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing

genomic features. *Bioinformatics*, *26*, 841–842.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*, 4348–4370.

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.

Ruiz-Martínez, J., Ferraguti, M., Figuerola, J., Martínez-De La Puente, J., Williams, R. A. J., Herrera-Dueñas, A., Aguirre, J. I., *et al.* (2016). Prevalence and genetic diversity of avipoxvirus in house sparrows in Spain. *PLoS ONE*, *11*, e0168690.

Samuel, M. D., Woodworth, B. L., Atkinson, C. T., Hart, P. J., & LaPointe, D. A. (2018). The epidemiology of avian pox and interaction with avian malaria in Hawaiian forest birds. *Ecological Monographs*, *88*, 621–637.

Santoro, M. G., Amici, C., & Rossi, A. (2009). 'Role of heat shock proteins in viral infection', in Pockley, A. G., Calderwood, S. K., & Santoro, M. G. (eds) *Prokaryotic and Eukaryotic Heat Shock Proteins in Infectious Disease*. Netherlands: Springer, 51–84.

Sedger, L., & Ruby, J. (1994). Heat shock response to vaccinia virus infection. *Journal of Virology*, *68*, 4685–4689.

Seet, B. T., Johnston, J. B., Brunetti, C. R., Barrett, J. W., Everett, H., Cameron, C., Sypula, J., *et al.* (2003). Poxviruses and immune evasion. *Annual Review of Immunology*, *21*, 377–423.

Shinya, K., Ito, M., Makino, A., Tanaka, M., Miyake, K., Eisfeld, A. J., & Kawaoka, Y. (2012). The TLR4-TRIF pathway protects against H5N1 influenza virus infection. *Journal of Virology*, *86*, 19–24.

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

Slonchak, A., Shannon, R. P., Pali, G., & Khromykh, A. A. (2016). Human microRNA miR-532-5p exhibits antiviral activity against West Nile virus via suppression of host genes SESTD1 and TAB3 required for virus replication. *Journal of Virology*, *90*, 2388–2402.

Smith, S. A., & Kotwal, G. J. (2002). Immune response to poxvirus infections in various animals. *Critical Reviews in Microbiology*, *28*, 149–185.

Smits, J. E., Tella, J. L., Carrete, M., Serrano, D., & Lopez, G. (2005). An epizootic of avian pox in endemic short-toed larks (*Calandrella rufescens*) and Berthelot's pipits (*Anthus berthelotti*) in the Canary Islands, Spain. *Veterinary Pathology*, *42*, 59–65.

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

Stewart, M. D., Ritterhoff, T., Klevit, R. E., & Brzovic, P. S. (2016). E2 enzymes: more than just middle men. *Cell Research*, *26*, 423–440.

Sung, C. K., Yim, H., Andrews, E., & Benjamin, T. L. (2014). A mouse polyomavirus-encoded microRNA targets the cellular apoptosis pathway through Smad2 inhibition. *Virology*, *468–470*, 57–62.

Taguwa, S., Maringer, K., Li, X., Bernal-Rubio, D., Rauch, J. N., Gestwicki, J. E., Andino, R., *et al.* (2015). Defining Hsp70 subnetworks in dengue virus replication reveals key vulnerability in flavivirus infection. *Cell*, *163*, 1108–1123.

Tripathy, D. N. (1993). 'Avipox Viruses', in *Virus Infections of Birds*. New York: Elsevier, 5–15.

van Riper, C., & Forrester, D. J. (2007). 'Avian Pox', in *Infectious Diseases of Wild Bird*. Oxford: Wiley Blackwell Publishing, 131–176.

van Riper, C., van Riper, S. G., & Hansen, W. R. (2002). Epizootiology and effect of avian pox on Hawaiian forest birds. *The Auk*, *119*, 929–942.

Vanderwerf, E. A. (2001). Distribution and potential impacts of avian poxlike lesions in ’ elepaio at hakalau forest national wildlife refuge. *Studies in Avian Biology*, *22*, 247–253.

Vanderwerf, E. A., & Young, L. C. (2016). Juvenile survival, recruitment, population size, and effects of avian pox virus in Laysan Albatross (*Phoebastria immutabilis*) on Oahu, Hawaii, USA. *Condor*, *118*, 804–814.

Vitti, J. J., Grossman, S. R., & Sabeti, P. C. (2013). Detecting natural selection in genomic data. *The Annual Review of Genetics*, *47*, 97–120.

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.

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

Williams, R. A. J., Escudero Duch, C., Pérez-Tris, J., & Benítez, L. (2014). Polymerase chain reaction detection of avipox and avian papillomavirus in naturally infected wild birds: comparisons of blood, swab and tissue samples. *Avian Pathology*, *43*, 130–134.

Williams, R. A. J., Truchado, D. A., & Benitez, L. (2021). A review on the prevalence of poxvirus disease in free-living and captive wild birds. *Microbiology Research*, *12*, 403–418.

Wood, V. H. J., O’Neil, J. D., Wei, W., Stewart, S. E., Dawson, C. W., & Young, L. S. (2007). Epstein-Barr virus-encoded EBNA1 regulates cellular gene transcription and modulates the STAT1 and TGFβ signaling pathways. *Oncogene*, *26*, 4135–4147.

Xu, S., Xue, C., Li, J., Bi, Y., & Cao, Y. (2011). Marek’s disease virus type 1 microRNA miR-M3

suppresses cisplatin-induced apoptosis by targeting SMAD2 of the transforming growth factor beta signal pathway. *Journal of Virology*, *85*, 276–285.

Yadav, S., J. Stow, A., & Dudaniec, R. Y. (2021). Microgeographical adaptation corresponds to elevational distributions of congeneric montane grasshoppers. *Molecular Ecology*, *30*, 481–498.

Ye, J., Chen, Z., Zhang, B., Miao, H., Zohaib, A., Xu, Q., Chen, H., *et al.* (2013). Heat shock protein 70 is associated with replicase complex of Japanese encephalitis virus and positively regulates viral genome replication. *PLoS ONE*, *8*, e75188.

Zhang, C., Kang, K., Ning, P., Peng, Y., Lin, Z., Cui, H., Cao, Z., *et al.* (2015). Heat shock protein 70 is associated with CSFV NS5A protein and enhances viral RNA replication. *Virology*, *482*, 9–18.

Zhang, L., Villa, N. Y., & McFadden, G. (2009). Interplay between poxviruses and the cellular ubiquitin/ubiquitin-like pathways. *FEBS Letters*, *583*, 607–614.

Zhang, X., Lin, X., Qin, C., Huang, K., Sun, X., Zhao, L., & Jin, M. (2020). Avian chaperonin containing TCP1 subunit 5 supports influenza A virus replication by interacting with viral nucleoprotein, PB1, and PB2 proteins. *Frontiers in Microbiology*, *11*, 538355.

Zhou, J., Chi, X., Cheng, M., Huang, X., Liu, X., Fan, J., Xu, H., *et al.* (2019). Zika virus degrades the ω-3 fatty acid transporter Mfsd2a in brain microvascular endothelial cells and impairs lipid homeostasis. *Science Advances*, *5*, eaax7142.

Zhou, Y., Gao, F., Lv, L., Wang, S., He, W., Lan, Y., Li, Z., *et al.* (2021). Host factor cyclophilin B affects Orf virus replication by interacting with viral ORF058 protein. *Veterinary Microbiology*, *258*, 109099.

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.

Zylberberg, M., Lee, K. A., Klasing, K. C., & Wikelski, M. (2012). Increasing avian pox prevalence varies by species, and with immune function, in Galápagos finches. *Biological Conservation*, *153*, 72–79.

2.7. Supplements

Table S2.1. Sampling periods for the 13 populations of Berthelot's pipit across Macaronesia and their use in different datasets. Where the number of individuals included in each dataset is less than the total sample size for that sampling period (*n*), the number of individuals is denoted in brackets.

Archipelago	Population	Year	Month(s)	*n*	Dataset(s)			
					Pox prevalence	TLR4 variation	MHC I variation	ddRAD-seq
Madeira	Deserta Grande	2006	Sept–Oct	31	X			X (18)
		2009	Apr, Sept–Oct	4	X			X (2)
	Madeira	2005	Apr	1	X			
		2006	Sept	32	X			
		2009	Mar–Apr, Sept–Oct	29	X			X (20)
		2016	Apr–May	29	X			
	Porto Santo	2006	Sept	31	X	X		X (3)
		2009	Mar	30	X	X		X (17)
		2016	May–June	129	X	X		
Selvagens	Selvagem Grande	2005	Apr	52	X			
		2009	Mar	42	X			X (20)
Canary Islands	La Graciosa	2006	Mar	24	X			X (2)
		2009	Jan	26	X			X (18)
		2020	Mar	44	X			
	Lanzarote	2006	Mar	13	X			
		2008	Nov	1	X			
		2009	Jan	30	X			X (20)
		2019	Apr	83	X			
		2020	Feb–Apr	113	X			
	Fuerteventura	2006	Feb–Mar	12	X			X (2)
		2008	Nov	1	X			
		2009	Jan	30	X			X (18)
	Gran Canaria	2006	Jan–Feb	31	X			
		2009	Jan–Feb	33	X			X (20)
	El Teide	2006	Mar–Apr	30	X	X (29)		
		2008	Oct	3	X			
		2009	Apr–May, Aug	22	X	X		X (20)
		2011	Mar–Apr	30	X	X	X (26)	
		2020	May–June	19	X			

Table S2.1. (Continued).

Archipelago	Population	Year	Month(s)	*n*	Dataset(s)			
					Pox prevalence	TLR4 variation	MHC I variation	ddRAD-seq
Canary Islands	Tenerife	2005	May	7	X			
		2006	Feb–Mar	25	X	X (22)		X (8)
		2009	Jan–Aug	35	X	X (34)		X (14)
		2010	Apr–May	96	X	X		
		2011	Jan–May	358	X	X (357)	X (284)	
		2020	May–June	18	X			
	La Gomera	2006	Feb	30	X			X (7)
		2009	Jan, Apr	20	X			X (13)
		2020	June	6	X			
	La Palma	2006	Jan–Feb	28	X			X (3)
		2009	Feb, July	22	X			X (17)
	El Hierro	2006	Jan–Feb	31	X			
		2009	Feb	30	X			X (20)

Table S2.2. Details of TLR4 variation among Berthelot's pipits on Porto Santo (PS) and Tenerife (TF).

Nucleotide haplotype	Nucleotide sequence				Protein haplotype	Amino acid sequence	Protein haplotype frequency *	
	SNP1 (905 bp)	SNP2 (970 bp)	SNP3 (990 bp)	SNP4 † (992 bp)			PS	TF
1	A	G	T	A	***TLR4_P1***	DGPK	0.53	0.29
2	A	G	C	C	***TLR4_P2***	DGPT	0.17	0.71
3	A	G	T	C				
4	A	A	C	C	***TLR4_P3***	DDPT	0.26	0.00
5	G	G	C	C	***TLR4_P4***	GGPT	0.04	0.00

Positions based on the TLR4 protein coding region in the zebra finch genome.
A fifth SNP (1,010 bp, nonsynonymous) was excluded due to low minor allele frequency.
* Low frequency variants (< 0.05) were excluded from analyses.
† This tri-allelic SNP was treated as biallelic by excluding the least frequent allele (T).

Table S2.3. Equivalent correlation matrix to the final population covariance matrix (averaged from 10 independent runs) which was used as a null model in Bayenv analyses. Acronyms: PS = Porto Santo; FV = Fuerteventura; LZ = Lanzarote; GC = Gran Canaria; TF = Tenerife; GOM = La Gomera; TEID = Teide; DG = Deserta Grande; EH = El Hierro; GRA = La Graciosa; LP = La Palma; M = Madeira; SG = Selvagem Grande.

	GRA	**LZ**	**FV**	**GC**	**TF**	**TEID**	**GOM**	**LP**	**EH**	**SG**	**M**	**PS**	**DG**
GRA													
LZ	0.955												
FV	0.955	0.958											
GC	0.954	0.957	0.959										
TF	0.953	0.958	0.959	0.959									
TEID	0.953	0.958	0.959	0.959	0.960								
GOM	0.950	0.954	0.956	0.956	0.956	0.956							
LP	0.952	0.956	0.958	0.958	0.957	0.957	0.956						
EH	0.946	0.950	0.952	0.952	0.952	0.953	0.953	0.952					
SG	0.742	0.745	0.747	0.747	0.745	0.751	0.743	0.741	0.748				
M	0.661	0.680	0.677	0.684	0.682	0.686	0.681	0.674	0.668	0.497			
PS	0.650	0.670	0.666	0.673	0.671	0.675	0.670	0.662	0.657	0.482	0.975		
DG	0.655	0.674	0.671	0.679	0.677	0.681	0.677	0.67	0.663	0.493	0.961	0.962	

Table S2.4. Prevalence of avian pox in 13 populations of Berthelot's pipit across Macaronesia. Prevalence was estimated from field data collected from 2005–2020 (the number of individuals caught with characteristic pox lesions divided by the total number of individuals caught throughout the entire period). The prevalence in each island was standardised by subtracting the mean and dividing by the standard deviation of prevalence across populations. Standardised prevalence was then used as an input variable for genotype-environment association analyses.

Archipelago	**Population**	***n***	**Pox prevalence %**	**Standardised prevalence**
Madeira	Deserta Grande	35	0.00	-0.75
	Madeira	91	0.00	-0.75
	Porto Santo	190	32.63	1.95
Selvagens	Selvagem Grande	94	0.00	-0.75
Canary Islands	La Graciosa	94	0.00	-0.75
	Lanzarote	240	23.33	1.18
	Fuerteventura	43	25.58	1.36
	Gran Canaria	64	21.88	1.06
	El Teide	104	1.92	-0.59
	Tenerife	539	7.24	-0.15
	La Gomera	56	5.36	-0.31
	La Palma	50	0.00	-0.75
	El Hierro	61	0.00	-0.75

Table S2.5. Results of generalised linear mixed models used to investigate associations between TLR4 variation and pox infection status when the variable malaria infection status was removed: (A) TLR4 protein haplotype (presence/absence), (B) TLR4 protein haplotype heterozygosity homozygote/heterozygote), (C) TLR4 protein genotype. Reference factor levels: TLR4_P = absence, TLR4_het = homozygote, and TLR4_genotype = 1,1.

	Porto Santo					***Tenerife***				
Model A	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	TLR4_P1	-0.145	0.476	-0.305	0.761	*TLR4_P1*	0.222	0.360	0.616	0.394
	TLR4_P2	-0.203	0.392	-0.517	0.605	*TLR4_P2*	0.085	0.647	0.131	0.911
	TLR4_P3	-0.093	0.363	-0.256	0.798					
	Random effects	Variance				**Random effects**	**Variance**			
	Sampling year	$4.676e^{-15}$	3 sampling years			*Sampling year*	$4.000e^{-14}$	4 sampling years		
	***n* = 184**					***n* = 578**				
Model B	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	TLR4_het	0.097	0.332	0.292	0.770	*TLR4_het*	0.201	0.345	0.583	0.429
	Random effects	**Variance**				**Random effects**	**Variance**			
	Sampling year	0	3 sampling years			*Sampling year*	$4.000e^{-14}$	4 sampling years		
	***n* = 184**					***n* = 578**				
Model C	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**	**Fixed effects**	**Estimate**	**Std. error**	**z**	***p*-value**
	TLR4_genotype1,2	-0.152	0.484	-0.313	0.754	*TLR4_genotype1,2*	0.085	0.648	0.131	0.911
	TLR4_ genotype1,3	0.059	0.421	0.140	0.889	*TLR4_genotype2,2*	-0.137	0.650	-0.210	0.718
	TLR4_ genotype2,3	0.031	0.554	0.056	0.956					
	TLR4_ genotype3,3	-0.257	0.742	-0.346	0.729					
	Random effects	**Variance**				**Random effects**	**Variance**			
	Sampling year	**0**	**3 sampling years**			***Sampling year***	**0**	**4 sampling years**		
	***n* = 168**					***n* = 578**				

Table S2.6. Results of generalised linear models used to investigate associations between MHC class I variation and pox infection status when the variable malaria infection status was removed: (A) MHC diversity (number of alleles per individual, 3–10) and optimality (quadratic of MHC allele number), and (B) presence of specific MHC alleles (presence/absence). Reference factor levels: ANBE = absence.

	Fixed effects	**Estimate**	**Std. error**	**z**	***p*-value**
Model A	*N.alleles*	-1.530	1.191	-1.284	0.199
	N.alleles.squared	0.111	0.090	1.234	0.217
Model B	*ANBE10*	0.518	1.160	0.447	0.655
	ANBE8	-0.819	0.731	-1.120	0.263
	ANBE4	-0.082	0.543	-0.152	0.879
	ANBE43	-0.217	0.525	-0.413	0.680
	ANBE1	0.798	0.509	1.568	0.117
	ANBE44	0.382	0.702	0.545	0.586
	ANBE45	1.150	0.900	1.278	0.201
	ANBE9	0.874	0.653	1.340	0.180
	ANBE46	0.616	0.833	0.739	0.460
	ANBE47	-0.646	0.667	-0.968	0.333
	ANBE11	-0.186	0.682	-0.272	0.785
	ANBE6	-0.035	0.851	-0.041	0.967
	ANBE38	0.634	0.837	0.757	0.449

Table S2.7. Detectable immune genes and their overlapping SNPs in the ddRAD-seq dataset for analyses with Bayenv. Genes and their genomic locations taken from Ekblom *et al.* (2010).

Avian immune gene	Chromosome	Gene start position	Gene end position	SNP	SNP position
CD247	1	102,472,923	102,479,033	2003s115	102,472,695
CD200	1	114,120,531	114,124,170	3088s53	114,125,011
				3088s65	114,125,023
TNFRSF11B	2	142,879,487	142,894,631	2752s29	142,880,794
CTSB	3	110,883,194	110,889,580	2361s40	110,887,301
Gal 8	3	110,809,584	110,810,285	397s64	110,810,106
				397s76	110,810,118
CD81	5	13,447,673	13,455,198	2998s79	13,440,274
CD28	7	50,658	60,186	2150s95	51,599
				2150s39	51,655
				2150s37	51,657
				2150s15	51,679
TNFSF15	17	2,727,000	2,741,576	2140s105	2,727,606
SOCS3	18	1,664,631	1,665,260	3441s90	1,660,387

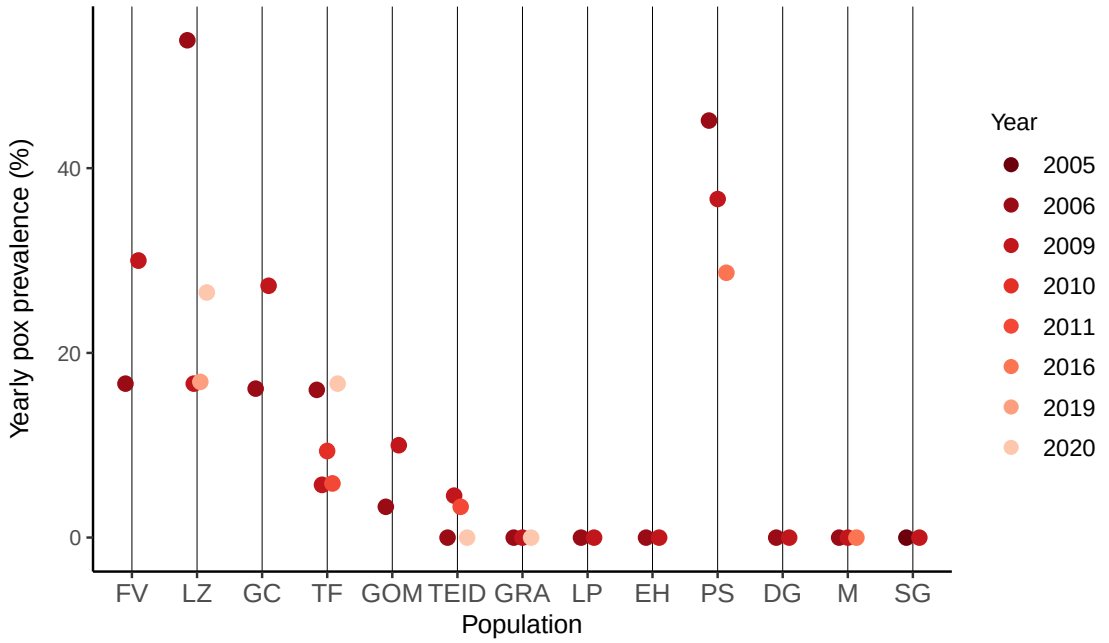

Figure S2.1. Dot plot showing multiple pox prevalence values from different years within populations. Patterns of prevalence are similar over time. All populations have been sampled at least twice. Prevalence values only included where $n > 10$. PS = Porto Santo; FV = Fuerteventura; LZ = Lanzarote; GC = Gran Canaria; TF = Tenerife; GOM = La Gomera; TEID = Teide; DG = Deserta Grande; EH = El Hierro; GRA = La Graciosa; LP = La Palma; M = Madeira; SG = Selvagem Grande.

Chapter 3

Genotype-environment associations reveal genes potentially linked to avian malaria infection in populations of an endemic island bird

Herding pipits on La Graciosa, Canary Islands

3.1. Abstract

Patterns of pathogen prevalence are, at least partially, the result of coevolutionary host-pathogen interactions. Thus, exploring the distribution of host genetic variation in relation to infection by a pathogen within and across populations can provide important insights into mechanisms of host defence and adaptation. Here we use a landscape genomics approach (Bayenv) in conjunction with genome-wide data (ddRAD-seq) to test for associations between avian malaria (*Plasmodium*) prevalence and host genetic variation across 13 populations of the island endemic Berthelot's pipit (*Anthus berthelotii)*. Considerable and consistent spatial heterogeneity in malaria prevalence was observed among populations over a period of 15 years. The prevalence of malaria infection was also strongly positively correlated with pox (*Avipoxvirus*) prevalence. Multiple host loci showed significant associations with malaria prevalence after controlling for genome-wide neutral genetic structure. These sites were located near to or within genes linked to metabolism, stress response, transcriptional regulation, complement activity and the inflammatory response, many previously implicated in vertebrate responses to malarial infection. Our findings identify diverse genes – not just limited to the immune system – that may be involved in host protection against malaria and suggests that spatially variable pathogen pressure may be an important evolutionary driver of genetic divergence among wild animal populations, such as Berthelot's pipit. Further, our data indicate that spatio-temporal variation in multiple different pathogens (e.g., malaria and pox in this case) may have to be studied together to develop a more holistic understanding of host pathogen-mediated evolution.

3.2. Introduction

Pathogens are – by definition – harmful to hosts, influencing both reproduction and survival (Anderson & May, 1979; Daszak *et al.*, 2000), thereby acting as a strong selective pressure on genes encoding for host defence (Sommer, 2005). Reciprocal selection pressure from hosts on infectivity-related genes also drives counter-adaptation in pathogens, resulting in strong coevolution (Dawkins & Krebs, 1979; Paterson *et al.*, 2010). However, pathogen-mediated selection does not always lead to the most beneficial host variants becoming fixed in the host population; multiple variants at a locus can be maintained via balancing selection (e.g., Ferrer-Admetlla *et al.*, 2008; Loiseau *et al.*, 2009; Hawley & Fleischer, 2012). Different mechanisms of balancing selection (reviewed in Charlesworth, 2006; Spurgin & Richardson, 2010) include heterozygote advantage (Doherty & Zinkernagel, 1975), rare allele advantage (Slade & McCallum, 1992), and spatio-temporally fluctuating selection (Hill *et al.*, 1991). Importantly, these mechanisms and divergent host-pathogen coevolutionary cycles across a host's geographic range may lead to spatial differences in the variants conferring defence against a given pathogen at any given time. Investigating the distribution of genetic variation within and among host populations facing a given pathogen can help to identify the loci and mechanisms that underlie the host's evolutionary response to that pathogen.

It can, however, be challenging to identify the underlying host loci involved in modulating host-pathogen interactions in natural systems. Analyses of vertebrate pathogen defences tend to have focused on well-known resistance related genes (i.e., those that prevent or limit infection) (Aguilar *et al.*, 2004; Brouwer *et al.*, 2010; Tschirren *et al.*, 2011; Hawley & Fleischer, 2012; Quéméré *et al.*, 2015; Davies *et al.*, 2021). By comparison, the genes involved in tolerance mechanisms which act to mitigate fitness costs by minimising direct or immune-mediated damage have been largely overlooked, despite evidence that tolerance is an important host defence strategy (Råberg *et al.*, 2007, 2009). The candidate gene approach, as well as other approaches such as genome-wide association studies (GWAS), also requires reliable individual-based phenotypic datasets to establish robust links between genetic variants and pathogen infection. However, limited knowledge of individual-level exposure to, or recovery from, infectious diseases in wild host systems can confound or conceal associations with genotypes. In such complex scenarios, bottom-up genome

scanning methods therefore offer a valuable alternative to uncover potentially adaptive loci and identify candidate genes for testing in future in-depth studies. These methods involve systematically screening thousands of variants across the genome, operating without the constraints of pre-existing knowledge about candidate genes or specific individual host phenotypes (Pardo-Diaz *et al.*, 2015).

Conducting genome scans to investigate genomic variation linked to divergent and spatially variable selection entails two primary methods: genetic differentiation-based approaches and genotype-environment association (GEA) analyses (Schoville *et al.*, 2012; Hoban *et al.*, 2016). While differentiation-based methods (such as those that involve F_{ST} outlier tests or alternative statistics) identify loci that exhibit extreme values of differentiation among populations compared to the genome-wide average (Lewontin & Krakauer, 1973; Beaumont, 2005), GEA analyses identify loci that show a strong association between the geographic pattern of allele frequencies and spatial variation in an environmental variable (Rellstab *et al.*, 2015). GEA methods can therefore help to elucidate the environmental selective pressure driving the allele frequency patterns, yet they do not explicitly incorporate individual phenotypic data. Instead, they operate on the assumption that population-level variation in an environmental variable (e.g., in pathogen prevalence) is representative of the strength of selective pressure acting on phenotypes in that population. Importantly, a number of model-based GEA tools exist that also account for the effects of neutral genetic structure across populations (e.g., latent factor mixed models, Frichot *et al.*, 2013; Bayenv, Günther & Coop, 2013; and BayPass, Gautier, 2015). Thus the application of GEA in host-pathogen systems can help to identify the genetic variants that may have accumulated in different populations of the same host species due to spatially variable selective pressure from a specific pathogen (e.g., Fumagalli *et al.*, 2011; Garroway *et al.*, 2013; Zueva *et al.*, 2014, 2018; Mackinnon *et al.*, 2016).

Avian malaria (here defined widely as infection by *Plasmodium* and *Haemoproteus* spp.) is a well-studied disease of wild birds due to its near-global distribution, wide host range and relevance for host health and conservation (Valkiunas, 2005). Infected birds typically undergo an acute phase, whereby high parasite loads are briefly experienced soon after infection, followed by a chronic phase with low parasitaemia (that may be undetectable in

the blood) which can last throughout an individual's lifetime with potential relapses (Lapointe *et al.*, 2012). Evidence from experimentally infected birds has shown reductions in condition and survival as a result of high parasitaemia (Atkinson *et al.*, 1995; Yorinks *et al.*, 2000; Palinauskas *et al.*, 2008). Wild-caught birds are more likely to have chronic levels of malaria rather than acute infections because the probability of capturing an individual depends on its survival and mobility (Mukhin *et al.*, 2016), however numerous studies have also shown negative effects on fitness including survival, lifespan (Marzal *et al.*, 2008; Lachish *et al.*, 2011; Asghar *et al.*, 2015), and reproductive success (Asghar *et al.*, 2011; Pigeault *et al.*, 2020). Collectively, evidence suggests that malaria has the potential to drive host variation in both immune genes and other parts of the genome related to health and homeostasis.

Berthelot's pipit (*Anthus berthelotii*) is endemic across three archipelagos in the North Atlantic – the Canary Islands, the Madeira archipelago and the Selvagens. It has only relatively recently dispersed across this island system (*ca*. 8,500–50,000 years ago) (Illera *et al.*, 2007; Spurgin *et al.*, 2014; Martin *et al.*, 2023) since its origin in the Canary Islands *ca.* 2.1–2.5 Mya (Voelker, 1999; Martin *et al.*, 2023). Previous genomic work has shown an absence of subsequent gene flow between populations (Martin *et al.*, 2021), thus potentially permitting population differentiation within and among archipelagos through local adaptation. The pipit therefore provides a useful system for studying the initial evolutionary responses to differential selective pressures in a wild vertebrate, including pathogen-mediated selection (e.g., González-Quevedo *et al.*, 2015, 2016; Armstrong *et al.*, 2018, 2019). Patterns of avian blood pathogen pressure are due to the combined effect of climatic, biogeographic and anthropogenic factors (Spurgin *et al.*, 2012; González-Quevedo *et al.*, 2014; Sehgal, 2015; Illera *et al.*, 2017; Padilla *et al.*, 2017) which are highly heterogenous within and among these island archipelagos. Importantly, prevalence of pathogen infections, specifically avian malaria and pox (*Avipoxvirus* species), differs markedly between Berthelot's pipit populations, aligning closely with biogeographical expectations based on island size and isolation (Spurgin *et al.*, 2012). Simultaneously, patterns of prevalence demonstrate relative consistency, as evidenced by studies conducted over a time frame ranging up to 15 years (Illera *et al.*, 2008; Spurgin *et al.*, 2012; Sheppard *et al.* 2022). Different populations are therefore likely to have been exposed to consistently different levels of pathogen-mediated

selection leading to variation in disease resistance/tolerance evolving across the Berthelot's pipit geographic range.

Analytical approaches based on individual-level assessments of infection, including candidate gene and GWAS methods, have previously been employed to identify pathogen related loci in Berthelot's pipit (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2018, 2019). Polymorphisms in key genes such as Toll-like receptors (TLRs) and the major histocompatibility complex (MHC), as well as a variant near to interleukin-16 (an inflammatory regulator), were found to be associated with malaria (González-Quevedo *et al.*, 2016; Armstrong *et al.*, 2018, 2019). However, it remains uncertain whether the variants underlying individual trait variation exhibit patterns of local adaptation across varying spatial scales, given the potential influence of distinct coevolutionary processes. Observations within populations or archipelagos may not necessarily extrapolate to broader spatial contexts (Bonneaud *et al.*, 2006; Kloch *et al.*, 2010; Armstrong *et al.*, 2018). Genome scans using measures of population differentiation (EigenGWAS; Chen *et al.*, 2016) have also revealed other loci affecting immunity, homeostasis and metabolism that are highly divergent between and/or within archipelagos (Armstrong *et al.*, 2018; Martin *et al.*, 2021). However, such an approach does not incorporate environmental or disease data, thus revealing little about how these drivers of spatially variable selection affect specific sites in the genome. Furthermore, differentiation-based methods primarily identify the most divergent SNPs, whereas GEA methods are more likely to detect loci with gradual allele frequency changes in response to spatially varying environments (Hancock *et al.*, 2010). There is also evidence that GEA tests possess higher power than differentiation outlier methods in detecting adaptive loci within island models (Lotterhos & Whitlock, 2015). A broad, multi-population, multi-gene assessment of associations between pathogen prevalence and genetic variation would therefore be beneficial within this species to gain further insights into the genetic basis of disease.

We recently utilised a Bayenv method with restriction site-associated DNA sequencing (RAD-seq) data across the Berthelot's pipit range to uncover candidate genes involved in the host's response to avian poxvirus (Sheppard *et al.*, 2022). Here, we use RAD-seq data and population-level measures of avian malaria prevalence from Berthelot's pipit, in conjunction

with the Bayenv method, to identify loci that may have evolved in response to avian malaria infection. We then explore the overlap between previously identified avian pox-associated single nucleotide polymorphisms (SNPs) and malaria-associated SNPs in an attempt to disentangle the effects of both diseases. We hypothesise that there could be a limited degree of overlap in SNPs linked to genes governing resistance, primarily due to the specificity of pathogen-targeting mechanisms. In contrast, SNPs linked to genes influencing tolerance are expected to exhibit a higher degree of overlap, given the shared mechanisms aimed at mitigating harm during infections.

3.3. Materials and Methods

3.3.1. Blood sampling

We used 1,574 Berthelot's pipits blood samples obtained between 2005–2020 as part of other studies (Illera *et al.*, 2007; Spurgin *et al.*, 2012; González-Quevedo *et al.*, 2014; Armstrong *et al.*, 2019; Sheppard *et al.*, 2022) (details in Table S3.1). Sampling included the 12 islands of the species' range across Macaronesia (Figure 3.1). Individuals inhabiting the plateau of Mount Teide (> 2,000 m above sea level) were considered a thirteenth population, separated from the lowland Tenerife population by the dense forest vegetation at moderate elevations which the pipit does not inhabit. Individuals were caught using spring nets baited with *Tenebrio molitor* larvae, and blood samples (*ca*. 25 µl), collected from the brachial vein, were stored in absolute ethanol (800 µl) at room temperature. Every bird was ringed with a numbered metal band from the relevant Spanish or Portuguese authorities to prevent resampling.

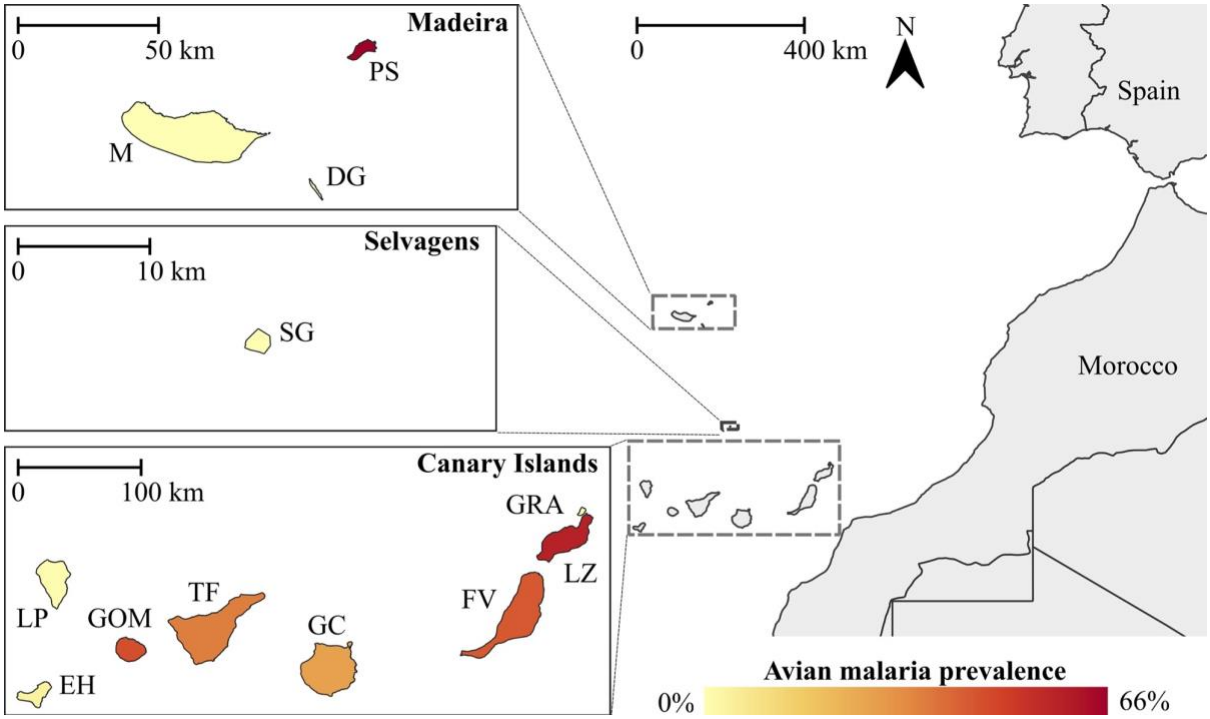

Figure 3.1. The distribution of Berthelot's pipit populations and the mean malaria (*Plasmodium*) prevalence (percentage infected) in those populations across a 15-year sampling period. Canary Island populations: FV = Fuerteventura; LZ = Lanzarote; GC = Gran Canaria; TF = Tenerife; GOM = La Gomera; TEID = El Teide; GRA = La Graciosa; LP = La Palma; EH = El Hierro. Madeiran populations: PS = Porto Santo; DG = Deserta Grande; M = Madeira. Selvagens population: SG = Selvagem Grande.

3.3.2. DNA extraction and molecular detection of malaria infection

Genomic DNA was extracted from blood using a salt extraction protocol (Richardson *et al.*, 2001). Samples were screened for *Haemoproteus* and *Plasmodium* spp. based on mitochondrial genome sequences using either the nested PCR protocols described in Waldenström *et al.* (2004) (samples collected prior to 2019) or the multiplex PCR assay described in Ciloglu *et al.* (2019) (samples collected in 2019 and onwards). Recent research into the comparative effectiveness of nested and multiplex PCR methods in diagnosing malaria suggests that they exhibit comparable sensitivity for *Plasmodium* and *Haemoproteus* (Ciloglu *et al.*, 2019; Musa *et al.*, 2022). However, the multiplex PCR assay outperforms the nested PCR in detecting co-infections, involving more than one species or strain of pathogen. Negative and positive controls were included in each PCR run. Each sample underwent two separate successful PCR screening events (three if the first two tests were inconsistent) and any samples that were positive twice were assigned as infected. All positive samples

collected prior to 2019 (*n* = 400) had been previously Sanger sequenced to identify the strain of haemosporidian (see Illera *et al.*, 2008; Spurgin *et al.*, 2012; González-Quevedo *et al.*, 2014; Armstrong *et al.*, 2019). No evidence of *Haemoproteus* spp. infection was detected, but four strains of *Plasmodium* (LK6, LK5, KYS9 and PS1530) were detected with no evidence of mixed-strain infections. The LK6 strain accounted for > 95% of infections, therefore testing for differences in strain-specific genetic associations was not possible. Instead, birds were classified by their *Plasmodium* infection status (not infected/infected) only.

3.3.3. RAD sequencing

RAD-seq markers have previously been used in GWAS analyses (based on individual assessments of disease) to identify genetic variants associated with LK6 infection status in Berthelot's pipits (Armstrong *et al.*, 2018), and in a GEA to identify genetic variants associated with poxvirus prevalence (population-level) across their range (Sheppard *et al.*, 2022). Here we used the 'Berthelot's' double digest RAD-seq (ddRAD-seq) library generated in Armstrong *et al.* (2018) and applied the same filtering steps as in Sheppard *et al.* (2022). The initial ddRAD-seq library was prepared and processed according to DaCosta & Sorenson (2014) with modifications detailed in Armstrong *et al.* (2018), and consisted of 20–22 individuals per population (putatively unrelated individuals and equal numbers of both sexes) sampled in 2006 and 2009. Reads were mapped to the zebra finch (*Taeniopygia guttata*) genome v3.2.4 (Warren *et al.*, 2010). Loci with ambiguous genotypes in ≤ 3 samples or those that showed an excess of missing/ambiguous genotypes (> 10% of samples) were treated as missing data. Multiple SNPs were outputted from each RAD-tag and treated as separate loci (SNPs are denoted by the RAD-tag followed by their bp distance from the beginning of the tag), resulting in 9,960 SNPs. We then performed additional filtering steps using Plink v1.9 (Chang *et al.*, 2015), removing loci with a minor allele frequency (MAF) of < 0.05 and one variant from each pair showing strong linkage disequilibrium (LD; $r^2 > 0.5$) with a sliding window of 50 kbp (step size of five).

3.3.4. Pathogen covariables for association analyses

Population malaria prevalence (Table S3.2) was calculated using the *Plasmodium* infection status of all birds sampled between 2005–2020 to account for year-to-year variability arising from natural fluctuations in infection rates and differences in sample size (cases of infected individuals / total number of individuals x 100). We considered utilising a binomial logistic regression to model the relationship between malaria and pox prevalence (previously estimated in Sheppard *et al.*, 2022) across populations because prevalence is a proportion derived from counts of infected and uninfected individuals. However, model diagnostics suggested that a beta distribution provided a more suitable fit for our data. Consequently, we represent prevalence as the proportion of infected birds and employed a zero-inflated beta regression using the glmmTMB package (v1.1.8; Brooks *et al.*, 2017) in R v4.3.1 (R Core Team, 2023). A beta regression model is well-suited for variables that are doubly bounded and often exhibit non-normal distributions and heteroscedasticity (Douma & Weedon, 2019). As our results indicated a strong positive correlation between malaria prevalence and pox prevalence across populations (see Section 3.4), we used both the "raw" malaria prevalence values and (in a separate model) residuals from malaria prevalence regressed on pox prevalence as inputs for the Bayenv models. This approach allowed us to examine the role of malaria independently of the influence of pox virus, while the "raw" values encompass more environmental variability and therefore provides greater power. All values were standardised to a mean of zero and a standard deviation of one prior to input into GEA analyses.

3.3.5. Genotype-environment association analyses

GEAs were conducted using Bayenv v2.0 (Günther & Coop, 2013) to identify associations between the allele frequencies of each SNP across populations and the gradient of malaria prevalence (or residual malaria) observed across populations. Subsequent analyses and visualisation of the output were performed using R v4.3.1 (R Core Team, 2023).

Bayenv employs a univariate Bayesian framework and explicitly accounts for shared evolutionary history and population structure by incorporating a covariance matrix of

population allele frequencies as the null model (Coop *et al.*, 2010; Günther & Coop, 2013). The matrix was generated previously in Sheppard *et al.* (2022) by running Bayenv with the filtered ddRAD-seq data and averaging the last covariance matrices of 10 replicate runs, each of 100,000 iterations.

For each SNP in the dataset, the analysis assigns a Bayes factor (BF) value, which describes the strength of evidence provided by the data for the alternative model – that is, a linear relationship between allele frequency and the pathogen covariable across populations – relative to the null model. Rather than relying solely on the parametric model, we also calculated Spearman's rank correlation coefficients (ρ) and ensured candidates indicated by BFs are not confounded by outlying populations (Günther & Coop, 2013). To further confirm this, we plotted population-level malaria prevalence against population allele frequencies for any candidate SNPs, using Plink v1.9 (Chang *et al.*, 2015) to estimate minor allele frequencies (MAFs). Following the recommendation of Blair *et al.* (2014), we averaged estimates across five independent runs, each of 1,500,000 iterations and with different random seeds, because Markov Chain Monte Carlo (MCMC) algorithms can generate high run-to-run variability (mean pairwise correlation between the five runs for the main analysis: for BF values $R = 0.690 \pm 0.222$, for ρ $R = 0.997 \pm 0.0001$). SNPs ranked above both the 99th percentile of averaged BF and the 90th percentile of averaged absolute value of ρ were then considered as robust candidates, in line with the manual (Gunther & Coop, 2018).

We used the zebra finch genome Taeniopygia_guttata-3.2.4 (Warren *et al.*, 2010) in the NCBI Genome Data Viewer v5.1 (www.ncbi.nlm.nih.gov/genome/gdv/browser) to identify candidate genes. These candidates were defined as genes located within a 10 kbp proximity either upstream or downstream of a malaria-associated SNP. Given the observed extent of LD within populations of Berthelot's pipit (Armstrong *et al.*, 2018; Martin *et al.*, 2021), we consider candidate loci located at this distance to be physically linked to the respective gene.

3.4. Results

Across the Berthelot's pipit geographic range there was considerable inter-population variation in malaria (*Plasmodium*) prevalence (Table S3.2). Two populations were completely

free of infection (Deserta Grande and Selvagem Grande) and prevalence in infected populations ranged between 1.1% in Madeira to 66.3% in Porto Santo – neighbouring populations in the Madeiran archipelago. Malaria was detected in every population within the Canary Islands, where the general pattern of prevalence is characterised by a negative east-west gradient (Figure 3.1). At the population level, the proportion of individuals infected with malaria was positively associated with previous estimates (Sheppard *et al.*, 2022) of pox infection within this species (beta regression estimate = 9.16 $\pm$ 1.99, $p < 0.01$; Figure 3.2). Estimates of malaria prevalence within populations were largely consistent between 2006 and 2009 (years in which all populations were sampled: Pearson correlation, $R = 0.64$, $p = 0.02$; Figure S3.1).

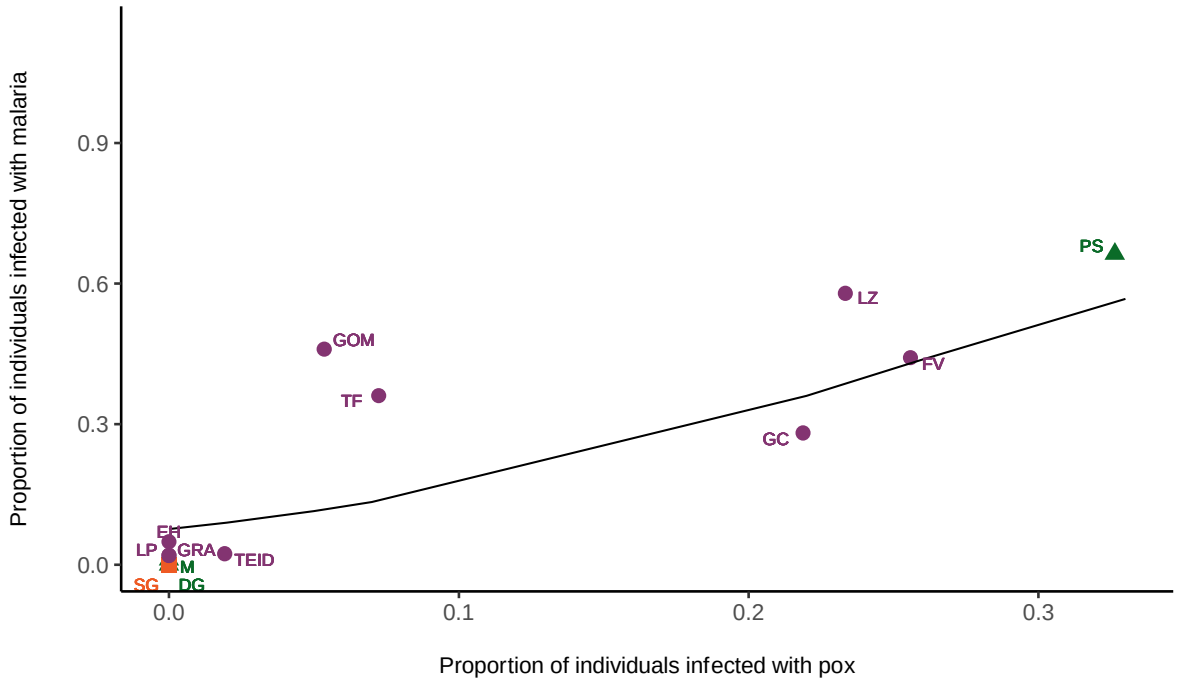

Figure 3.2. Spatial patterns of avian malaria (*Plasmodium*) and avian pox (*Avipoxvirus*) prevalence (proportion of individuals infected) across populations of Berthelot's pipit. Values of pox prevalence previously estimated in Sheppard *et al.* (2022). The fit line is a zero-inflated beta regression between the x- and y-axes. Grey shading indicates 95% confidence intervals. Points depict raw data. Canary Island populations (purple): FV = Fuerteventura; LZ = Lanzarote; GC = Gran Canaria; TF = Tenerife; GOM = La Gomera; TEID = El Teide; GRA = La Graciosa; LP = La Palma; EH = El Hierro. Madeiran populations (green): PS = Porto Santo; DG = Deserta Grande; M = Madeira. Selvagens population (orange): SG = Selvagem Grande.

A total of 2,334 SNPs with minor allele frequency > 0.05 were retained for analysis with Bayenv following LD filtering (r^2 > 0.5) of the ddRAD-seq dataset (Armstrong *et al.*, 2018). Twenty-three SNPs (*ca*. 1% of the total SNPs) passed the stringent thresholds (Figure 3.3) and were identified as having an association between allele frequencies and malaria prevalence at the population level following five runs of Bayenv. These associations did not seem to be driven by single outlying populations (Figure 3.4). The SNPs found in both the top 1% of the averaged BFs (> 6.68) and 10% of the averaged absolute values of Spearman's ρ (> 0.25) were located in proximity to (within 10 kbp upstream or downstream) or within 20 annotated candidate genes (Table 3.1). Specifically, eight SNPs were located within genes (34.8%), seven were located 10 kbp up- or downstream of a gene(s) (30.4%), one was located within 10 kbp of a pseudogene (4.3%), and seven were not near genes (30.4%).

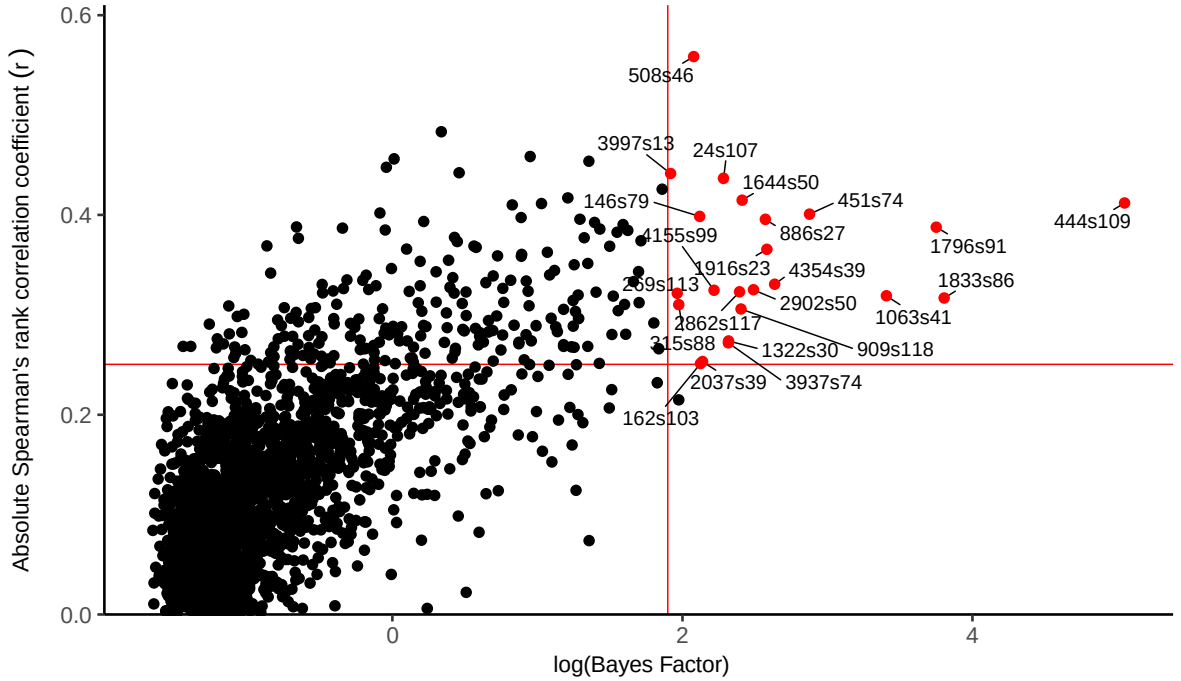

Figure 3.3. The distribution of Bayes factor values and absolute Spearman's rank correlation coefficients (ρ), generated and averaged from five replicate runs of Bayenv, for genome-wide ddRAD SNPs among 13 Berthelot's pipit populations in respect to malaria prevalence. SNPs were considered candidates for associations with population-level malaria prevalence if they ranked in the highest 1% of Bayes factor values (> 6.68, threshold indicated by the vertical red line) and 10% of Spearman's ρ (threshold indicated by the horizontal red line). Twenty-three SNPs were identified as candidates (in red).

Table 3.1. SNPs identified as being associated with population-level malaria prevalence using Bayenv, ordered by Bayes factor (BF) values. SNPs highlighted in blue text were previously also identified as being associated with pox prevalence (Sheppard *et al.*, 2022) using the same genetic dataset in a Bayenv analysis.

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
444s109	156.0	24:3526218	*HSPA8*	6,319 bp upstream	heat shock protein family A (Hsp70) member 8	ATP-dependent molecular chaperone, aiding protein folding
1833s86	45.0	19:7248322	–	–	–	–
1796s91	42.6	Z:421273	*SMAD2*	In gene (intron)	SMAD family member 2	Transcriptional modulator and downstream effector of the transforming growth factor (TGF)-β signalling pathway
1063s41	30.2	1A:217173	*UBE2H*	In gene (intron)	ubiquitin conjugating enzyme E2 H	Catalyses the attachment of ubiquitin to other proteins in the ubiquitin/proteasome degradation pathway
451s74	17.8	8:23610708	*C8A*	In gene (intron)	complement C8 alpha chain	Complement system component, initiates membrane penetration and coordinates formation of the membrane attack complex (MAC), an important antibacterial immune effector
			C8B	8,722 bp downstream	complement C8 beta chain	
4354s39	14.0	1:98523884	*SCAF4*	440 bp downstream	SR-related CTD-associated factor 4	RNA binding protein, involved in mRNA processing potentially through termination of transcription
			SOD1	1,606 bp downstream	superoxide dismutase 1	Copper/zinc binding antioxidant enzyme
1916s23	13.2	Z:50640938	*CWC27*	In gene (intron)	CWC27 spliceosome associated cyclophilin	Pre-mRNA splicing factor recruited by the spliceosome

Table 3.1. (Continued).

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
886s27	13.1	11:8338772	–	–	–	–
2902s50	12.1	11:13961995	*MYLK3*	7,175 bp upstream	myosin light chain kinase 3	Kinase that phosphorylates myosin chains and plays a role in cardiac performance and calcium signalling
1644s50	11.2	18:10178831	–	–	–	–
909s118	11.1	21:5744237	–	–	–	–
2862s117	11.0	11:1265660	*MLYCD*	3,598 bp downstream	malonyl-CoA decarboxylase	Catalyses the conversion of malonyl-CoA to acetyl-CoA during fatty acid metabolism
1322s30	10.2	6:13842170	*DUPD1*	349 bp downstream	dual specificity phosphatase and pro isomerase domain containing 1	Dephosphorylates phospho-serine/threonine and phosphotyrosine residues, modulating downstream cell signalling pathways
			KAT6B	2,756 bp downstream	lysine acetyltransferase 6B	Catalyses the acetylation of histones and plays a key role in the regulation of transcription of multiple genes
3937s74	10.1	11:10604516	*ZCCHC14*	In gene (intron)	zinc finger CCHC domain-containing 14	May bind nucleic acids, proteins and zinc, and play a role in RNA metabolism
24s107	9.8	11:12916411	*ZNF423*	In gene (intron)	zinc finger protein 423	DNA-binding, multifunctional transcription factor involved in signal transduction
4155s99	9.2	22:2584411	–	–	–	–
2037s39	8.5	9:5918598	–	–	–	–

Table 3.1. (Continued).

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
162s103	8.4	1A:30828068	*XRCC6BP1*	986 bp downstream	Mitochondrial inner membrane protease ATP23	Metallopeptidase involved in ATP synthase assembly and DNA repair
			RPAP3	8,853 bp upstream	RNA polymerase II-associated protein 3	Helps the assembly of RNA polymerase II and functions in the regulation of transcription and apoptosis
146s79	8.3	2:4280785	*CRHR2*	In gene (intron)	corticotropin releasing hormone receptor 2	Binds corticotrophin-releasing factor which plays a role in coordinating the endocrine, autonomic, and behavioural responses to stress and immune challenge
508s46	8.0	2:67296436	–	–	–	–
315s88	7.2	Z:10812737	*LOC100223460*	In pseudogene	–	–
269s113	7.1	15:384062	*ADGRD1*	In gene (intron)	adhesion G protein-coupled receptor D1	Involved in cell surface receptor signalling
3997s13	6.8	14:9421650	*ZNF598*	2,333 bp downstream	zinc finger protein 598	Initiates ribosome quality control (RQC) pathway when a ribosome has stalled during translation
			SYNGR3	2,904 bp downstream	synaptogyrin-3	May play a role in regulated exocytosis

* Chromosome location and base position based on *Taeniopygia_guttata*-3.2.4 genome assembly.

† Within 10 kbp of focal SNP.

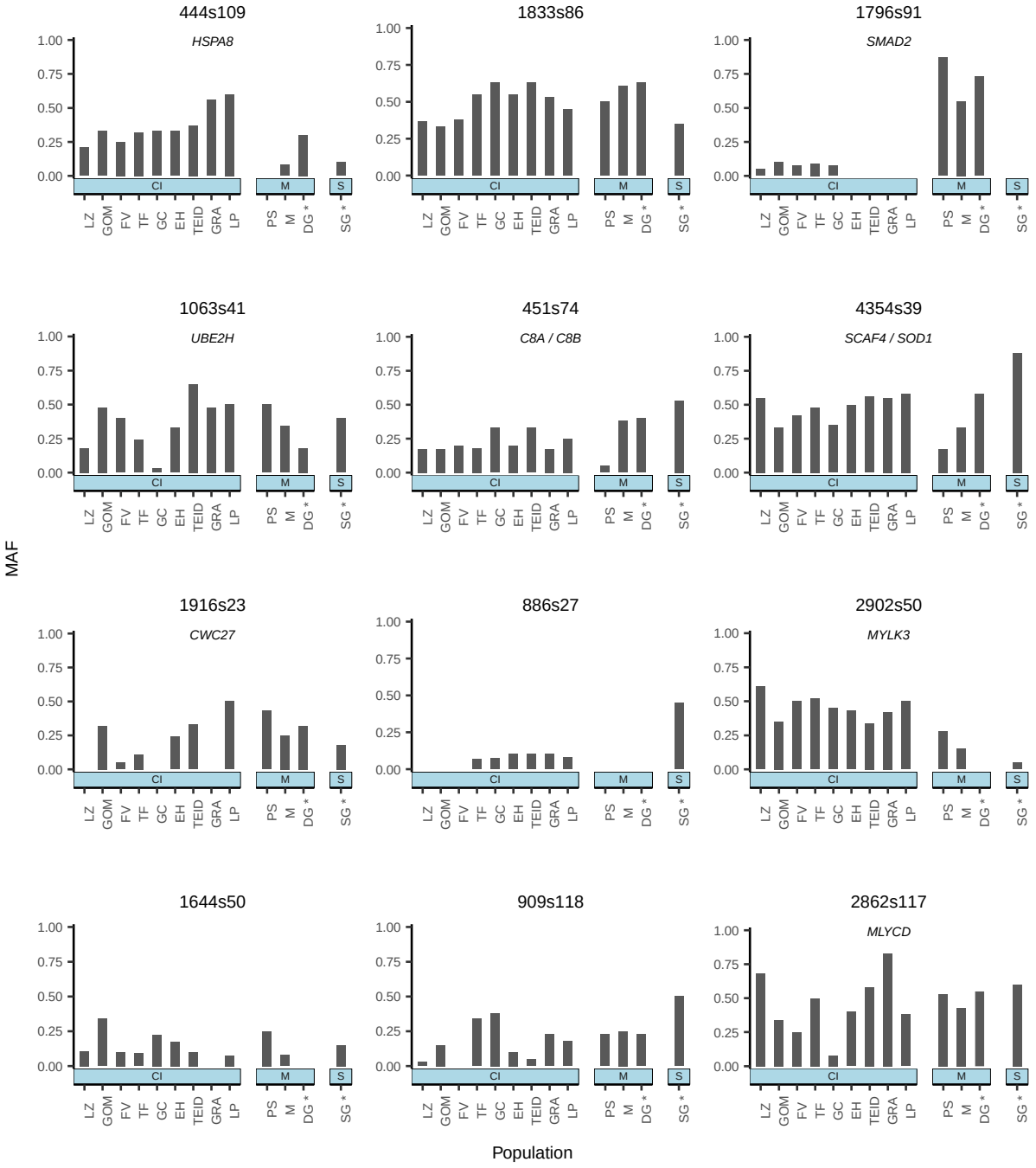

Figure 3.4. Minor allele frequency (MAF) of 23 candidate SNPs identified by testing for associations with population-level malaria prevalence in Bayenv. Bayenv corrects for neutral population structure among allele frequencies in the analysis. Nearby genes are noted below the SNP names. Populations are grouped by archipelago (CI = Canary Islands; M = Madeira; S = Selvagens) then ordered according to malaria prevalence (highest–lowest). Malaria-free populations are indicated by an asterisk. Acronyms: LZ = Lanzarote; GOM = La Gomera; FV = Fuerteventura; TF = Tenerife; GC = Gran Canaria; EH = El Hierro; TEID = El Teide; GRA = La Graciosa; LP = La Palma; PS = Porto Santo; M = Madeira; DG = Deserta Grande; SG = Selvagem Grande.

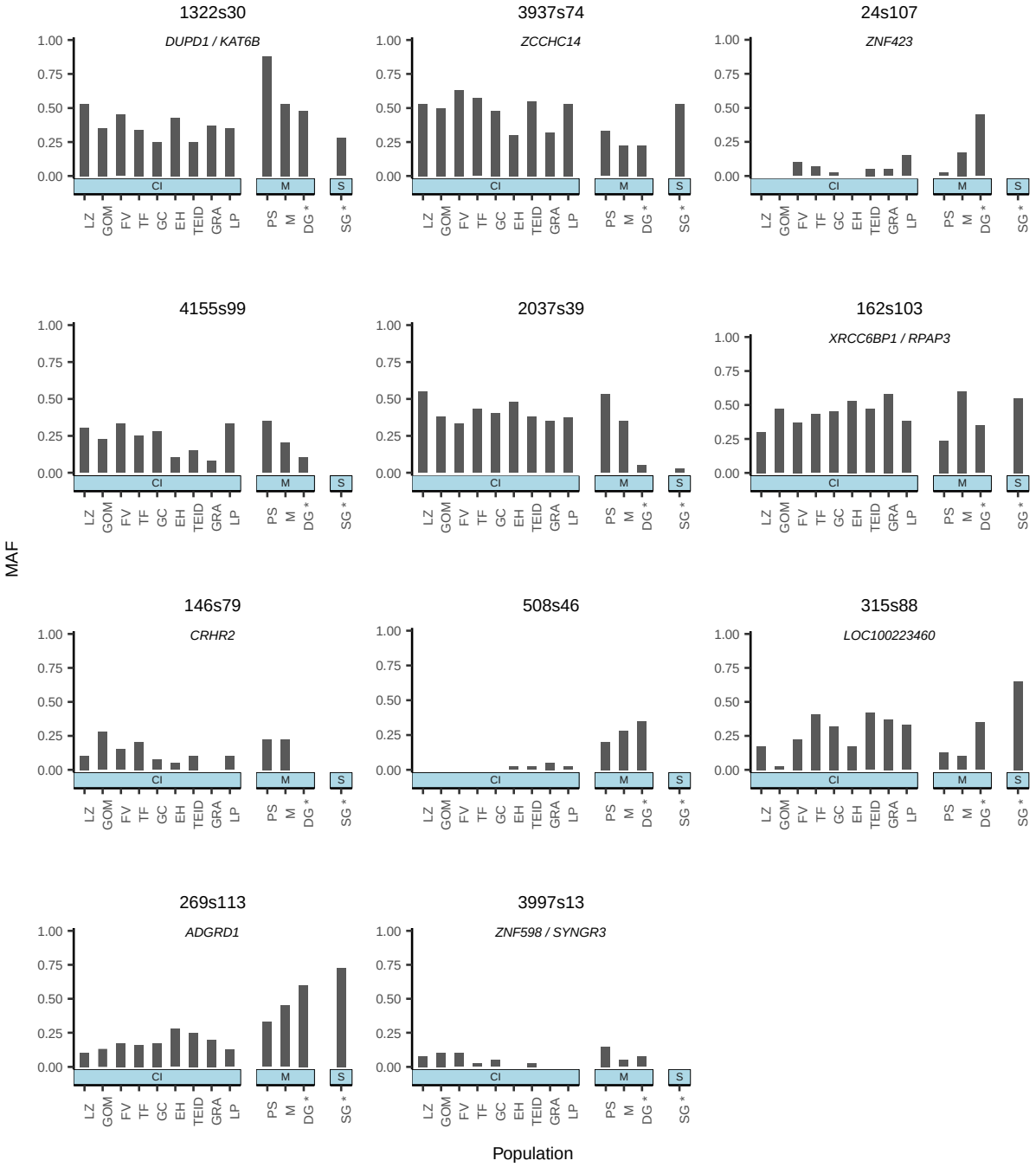

Figure 3.4. (Continued). Minor allele frequency (MAF) of 23 candidate SNPs identified by testing for associations with population-level malaria prevalence in Bayenv. Bayenv corrects for neutral population structure among allele frequencies in the analysis. Nearby genes are noted below the SNP names. Populations are grouped by archipelago (CI = Canary Islands; M = Madeira; S = Selvagens) then ordered according to malaria prevalence (highest–lowest). Malaria-free populations are indicated by an asterisk. Acronyms: LZ = Lanzarote; GOM = La Gomera; FV = Fuerteventura; TF = Tenerife; GC = Gran Canaria; EH = El Hierro; TEID = El Teide; GRA = La Graciosa; LP = La Palma; PS = Porto Santo; M = Madeira; DG = Deserta Grande; SG = Selvagem Grande.

When the same Bayenv analysis was performed using 'residual malaria', 18 SNPs exhibited a strong association. However, several of these associations appeared to be influenced by the most divergent, malaria-free population(s) (Figure S3.2). These SNPs were located in proximity to or within 13 candidate genes (Table S3.3). Notably, two of these SNPs, corresponding to three candidate genes, demonstrated significant associations that were shared between the analyses conducted with 'raw' prevalence values and 'residual malaria'. Specifically, eight SNPs were located within genes (34.8%), three were located 10 kbp up- or downstream of a gene(s) (16.7%), one was located within 10 kbp of a pseudogene (5.5%), and six were not near genes (33.3%).

3.5. Discussion

Applying a broad multi-population GEA approach, we identify candidate SNPs associated with avian malaria across the range of Berthelot's pipit. Population-level allele frequency variation at these candidate SNPs correlated with the local prevalence of malaria after controlling for genome-wide divergence due to neutral structure. These results reveal genes potentially involved in a host's response to malaria and suggest that local adaptation to spatially variable pathogen pressures, such as malaria, may be an important driver of genetic variation in Berthelot's pipit.

Spatially variable selective pressure from avian malaria has been shown to shape host genetic structure across populations (Foster *et al.*, 2007), drive variation in immunity and health (Names, Schultz, Hahn, *et al.*, 2021) and drive the evolution of variable tolerance (Atkinson *et al.*, 2013). In the present study, the main GEA analysis revealed 23 loci that showed strong evidence of association with population-level malaria prevalence, likely as a result of selective pressure. While Bayenv corrects for neutral population allele frequency patterns (Coop *et al.*, 2010; Günther & Coop, 2013), the methods used here are correlative so inferences can be confounded by collinearity along the axis of interest, whether with an additional environmental variable (discussed below in relation to pox) or the principal axes of neutral population structure. From previous work, malaria prevalence seems to be independent of genetic structure across Berthelot's pipit populations; rather, other biogeographical and environmental factors likely explain the distribution of avian malaria

(Spurgin *et al.*, 2012). For instance, Porto Santo experiences the highest levels of malaria prevalence yet other populations within the same archipelago, between which there is little genetic structure, are completely free of avian malaria (Figure 3.1). Hence, it is unlikely that prior distribution of susceptible alleles through founder effects would drive associations between malaria and specific loci. Instead, the identified genomic associations are consistent with locally divergent selection across the species' range and provide support for the hypothesis that there is adaptation to the local pathogen regime.

A set of 20 malaria-associated candidate genes were located within 10 kbp of the 23 associated SNPs, many of which have not been previously linked to divergence across this system. The molecular mechanisms that underlie the host response to malaria have predominantly been investigated in humans and murine models, as well as captive and natural populations of birds. We did not identify some of the immune-related genes commonly reported to influence malaria-related phenotypes in these taxa such as *G6PD*, *ABO*, *CD40LG*, *TNF*, *ATP2B4*, beta globins, interleukins, toll-like receptors or MHC (Bonneaud *et al.*, 2006; Longley *et al.*, 2011; Sepil *et al.*, 2013; Rockett *et al.*, 2014; Ravenhall *et al.*, 2018; MalariaGEN, 2019). This may in part be due to a lack of genomic resolution to evaluate these in our study (see below) or the differences in specific host-pathogen interactions. However, we found a few genes that function with the innate immune system and inflammatory response (*C8A*, *C8B*, *SMAD2*). For example, all complement pathways culminate in the formation of the membrane attack complex – an important innate immune effector that transverses pathogen cell membranes and facilitates pathogen clearance – of which complement 8 (*C8*) is a key component (Bayly-Jones *et al.*, 2017). Studies in both humans and mice have revealed the complement system is activated against malaria (Silver *et al.*, 2010; Kurtovic *et al.*, 2020) with elevated levels of membrane attack complex formation observed during infection (Roestenberg *et al.*, 2007). In birds, other genes of the complement system have been shown to be differentially expressed during infection with malaria (Videvall *et al.*, 2015) and there is evidence that elevated complement activity might act to control parasitaemia (Ellis *et al.*, 2015). Also, *SMAD2* encodes an intracellular transcription factor, activated by the transforming growth factor β (*TGF-β*) signalling pathway, which together regulate inflammation and interact with other immune system pathways. *TGF-β* itself was inferred to be under selection from malaria in low-elevation

populations of Hawai'i ‘amakihi (*Chlorodrepanis virens*) (Cassin-Sackett *et al.*, 2019). Given the dual role inflammation plays in malaria clearance and associated pathological changes (Popa & Popa, 2021), genes involved in its regulation may represent important mechanisms of malaria tolerance. Overall, these genes correspond with previous findings that suggest both the innate immune system and inflammatory response are commonly initiated in response to malaria infection, thus these genes seem appropriate candidates for further study into responses to malaria infection.

Recent transcriptomic studies have uncovered genes that are upregulated or downregulated during infectious disease (e.g., Rosenblum *et al.*, 2012; Huang *et al.*, 2013; Chu *et al.*, 2016; Boštjančić *et al.*, 2022; McNew *et al.*, 2022). Those examining gene expression changes in the host response to malaria have identified genes that were related to a number of functions beyond those strictly within the immune system, including the stress response, cell death regulation, metabolism, cell signalling, transcriptional regulation and telomerase activity (Videvall *et al.*, 2015, 2020; Quin *et al.*, 2017). In the present study, we detected associations with malaria prevalence within two of the same genes that were previously overexpressed in infected avian hosts (Videvall *et al.*, 2015), *UBE2H* and *CWC27*. Both of these genes have been associated with metabolic processes; *UBE2H* is involved in ubiquitin-dependent protein catabolism, targeting proteins for degradation, and can also affect antigen processing and presentation (Stewart *et al.*, 2016), and *CWC27* is a spliceosome component likely involved in mRNA processing and protein metabolic processes (Busetto *et al.*, 2020). Among the malaria-associated variants identified in the present study, many were also located near genes with links to protein and RNA metabolism (*HSPA8*, *SOD1*, *MYLK3*, *DUPD1*, *KAT6B*, *ZCCHC14*, *ZNF423*, *XRCC6BP1*) and a single variant was near a gene involved in fatty-acid oxidative metabolism (*MLYCD*). Immune functions incur high energetic and metabolic costs (Lochmiller & Deerenberg, 2000; Ots *et al.*, 2001; Martin *et al.*, 2003) and there is some evidence that energy costs of infection may impact body condition in Berthelot’s pipit (Spurgin *et al.*, 2012), thus metabolic adaptations could influence host fitness.

During malaria infection, the production of reactive oxygen species (ROS) occurs as a result of phagocytosis (which is beneficial for pathogen clearance), as well as host inflammatory

signalling, host metabolic changes induced by infection and the degradation of haemoglobin by the pathogen itself (Vasquez *et al.*, 2021). Thus the production of ROS is a key component of the host defence response, however high levels can lead to non-specific oxidative damage to host lipids, proteins and DNA (Szabó, 2003). The imbalance between oxidants and antioxidant defences is referred to as oxidative stress and is a common occurrence across a variety of taxa experiencing *Plasmodium* infections (e.g., humans: Bilgin *et al.*, 2012; Narsaria *et al.*, 2012; monkeys: Srivastava *et al.*, 1992; mice: Nneji *et al.*, 2013; Scaccabarozzi *et al.*, 2018 and birds: van de Crommenacker *et al.*, 2012; Delhaye *et al.*, 2016; Jiménez-Peñuela *et al.*, 2023). Antioxidant defences include both limiting oxidation and repairing cellular damage (Pamplona & Costantini, 2011). Among the candidate genes we identified in this study, *HSPA8*, an isoform of heat shock protein 70 (*HSP70*), contributes to the protein damage stress response (Sottile & Nadin, 2018). *HSP70* has previously been shown to be highly differentiated between malaria-naïve and exposed populations of Hawai'i ‘amakihi (passerine birds) and therefore inferred to be under selection (Cassin-Sackett *et al.*, 2019). Another candidate gene, *XRCC6BP1*, is potentially involved in the repair of DNA double-strand breaks (Fischer & Meese, 2007) but as far as we are aware, has not been linked to malaria response previously. A variant near to *SOD1* was also identified. This gene encodes a well-known antioxidant enzyme that catalyses superoxide radicals, thereby regulating levels of ROS and limiting their potential toxicity (Fridavich, 1995). Superoxide dismutase has been repeatedly linked to malaria infection across a range of taxa, including evidence for lower antioxidant activity in birds (Jiménez-Peñuela *et al.*, 2023), decreasing levels in monkeys (Srivastava *et al.*, 1992) and the downregulation of gene expression (Reuling *et al.*, 2018) and links with parasitaemia in humans (Andrade *et al.*, 2010). There is also evidence that polymorphism within *SOD1* may be associated with malaria susceptibility (Fernandes *et al.*, 2015). Therefore, it is plausible that variation within or near these genes may mediate stress, minimise tissue damage and affects the host’s ability to tolerate infection by malaria.

Many of the genes identified in our study are regulatory (e.g., *HSPA8*, *SMAD2*, *C8A*, *C8B*, *SOD1*, *MYLK3*, *MLYCD*, *KAT6B*, *ZNF423*, *RPAP3*, *CRHR2*, *ADGRD1*) which suggests that mechanisms for pathogen tolerance may be an important part of the response to malaria infection in Berthelot’s pipit. Furthermore, our findings suggest malaria-mediated selection may drive the evolution of parts of the genome involved in diverse functions related to

metabolic processes, stress response and transcriptional regulation, as well as the innate immune system and inflammatory response. This result underscores the importance of genome-wide approaches for the identification of candidate genes that underlie host-pathogen interactions, even when their relevance may not be immediately apparent. Some of the genes we have identified have multifarious roles in broad cellular processes (Table 3.1), therefore understanding the exact mechanism by which these genes influence the host's malaria response is not possible here. While much of the variation that we have identified seems to be linked with appropriate candidate genes that may putatively confer defence against malaria, we stress that validating an adaptive role requires further exploration through functional, experimental, and genomic investigations. If the GEA has identified genes involved in local adaptation in response to malaria, then future research should establish the association between gene-specific variants, infection status, and individual host fitness. This can be achieved through extensive within-population studies or – although challenging in non-model wild species like Berthelot's pipit – experimental infection studies.

General signatures of selection in the genome of Berthelot's pipit have previously been investigated at the population and archipelago level using differentiation-based analyses (Armstrong *et al.*, 2018; Martin *et al.*, 2021), with a single overlapping candidate gene region identified between studies at these different scales. In contrast to differentiation-based methods, GEA methods have the capability to identify the specific environmental variables driving selection, providing a more direct insight into the selective pressures at play. Here, we have specifically looked at loci that align with malaria prevalence across the geographic range. One of the genes identified in the current study (*ZNF598*), which codes for a zinc finger protein, was inferred to be under divergent selection between archipelagos (Armstrong *et al.*, 2018). This family of proteins can interact with DNA, RNA and proteins and are involved in numerous cellular functions including transcriptional regulation, ubiquitin-mediated protein degradation and regulation of apoptosis (Cassandri *et al.*, 2017), but the link to malaria protection is unclear. Thus, the most strongly divergent SNPs among populations of Berthelot's pipit (as identified in differentiation-based analyses) may not be shaped by pathogen pressures because we only detected a signal for one of the same genes using the GEA.

We detected genetic associations with population-level malaria prevalence within our RAD-generated SNP set. While RAD-seq is a cost-effective way to generate genome-wide marker sets, it is a reduced representation sequencing approach so potentially important genes in unrepresented regions could have gone undetected. This could explain why we did not detect immune genes identified as being relevant for malaria resistance in other taxa (mentioned above), or additional immune genes overall since only a small fraction of avian immune genes can be evaluated with this dataset (Sheppard *et al.*, 2022). Equally, some identified variants were not located within genes, which could either reflect the limited density of the marker set or potentially important functional non-coding regions. Even when associated variants were identified within genes, some lack research (e.g., *ZCCHC14*) making it difficult to interpret their potential function during infection. It is also important to note that while Bayenv is a powerful tool, it is not without weaknesses. False positives can derive from confounding correlations between environmental factors, which may be especially difficult to separate at the between population-level (see below), or even strong genetic drift. We employed a rather conservative approach, selecting only the candidates that ranked above high percentiles. While this could increase the number of false negatives, the aim of this study was to detect the top candidate gene set that merits future within-population investigation, rather than every possible association with malaria. Future research aimed at detecting additional variants under malaria-mediated selection, or indeed corroborating the candidate genes identified in the current study, may utilise alternative statistical approaches. Specifically, harnessing the capabilities of whole genome sequencing data would facilitate a comprehensive identification of genetic variants. The broad coverage provided by whole genome data enhances statistical power, offering a significant advantage in capturing subtle or weak signals of selection. Furthermore, exploiting museum specimens could prove valuable for characterizing historical polymorphic SNPs that are close to or have reached fixation in contemporary populations due to positive or purifying selection.

Concurrent infections of avian malaria and poxvirus have been documented in a number of host species (e.g, Atkinson *et al.*, 2005; Alley *et al.*, 2010; Ha *et al.*, 2012; Samuel *et al.*, 2018), including Berthelot's pipits (Spurgin *et al.*, 2012). Although high pathogen prevalence itself might suggest a strong likelihood of coinfection, our previous research demonstrated non-random cooccurrence within the pipit populations of Tenerife and Porto Santo through

individual-based analyses (Sheppard *et al.*, 2022). In that same study, we used a Bayenv analysis to identify genetic associations with pox prevalence differences between populations. Here, we show that patterns of both malaria and pox prevalence are similarly structured at the population level, thus making it technically difficult to isolate the effects of either pathogen in this natural system. Due to the univariate nature of Bayenv, the model lacks the capability to account for shared variance among multiple predictors. Therefore, caution must be exercised when interpreting loci correlated to multiple environmental variables. While some identified genes may be associated with pox rather than malaria, and vice versa, there is variability allowing the detection of additional candidates involved in malaria protection. In the current study, we identified 15 novel candidate genes (17 variants) putatively involved in malaria response in Berthelot's pipit that were not identified in the previous pox analysis. Conversely, seven genes (8 variants) identified in the previous pox analysis were not associated with malaria prevalence in the current study, and only five genes (6 variants) were identified in both studies (*HSPA8*, *SMAD2*, *UBE2H*, *CWC27*, and *MLYCD*).

To better discern malaria's individual contribution as a driver of genetic variation, we also tested for genomic associations with 'residual malaria' prevalence. Beyond the three initially identified genes with 'raw' prevalence values (*C8A*, *C8B* and *CRHR2*), the analysis identified an additional 10 candidate genes (16 SNPs). As expected, none of these genes, including the overlapping three, were identified in the previous pox study (Sheppard *et al.*, 2022). A literature search did not reveal logical associations with malaria infection for these new genes (Table S3.3), however a few have been identified as outliers/candidates in prior pipit studies using differentiation-based outlier tests (*WDR72*) (Armstrong *et al.*, 2018; Martin *et al.*, 2021) or in individual-level associations with other unrelated traits such as bill length (*PLEKHN1*) (Armstrong *et al.*, 2018). Many of these outliers can be attributed to a combination of factors, including the demographic history of certain populations aligning with the gradient of malaria prevalence across populations. The 'residual' analysis, with reduced variation in pathogen data, has constrained power, potentially resulting in the dominance of drift-driven false positives. Notably, smaller populations like Selvagem Grande and Deserta Grande, having undergone extreme bottlenecking and being the only malaria-free populations, may be driving these new associations when power is limited (refer to

Figure S3.2). Considering the analysis with 'raw' malaria prevalence had significantly greater power, associations with residual malaria did not yield a substantially improved candidate gene set. However, future in-depth analyses examining SNPs associated with malaria in this species should undoubtedly include *C8A*, *C8B*, and *CRHR2*.

Reasons for the association between malaria and pox in Berthelot's pipit are unknown but could include common risk factors that mediate exposure, such as host or vector densities, behaviours, and climatic/habitat conditions (Johnson & Buller, 2011; Hellard *et al.*, 2015). For example, the geographic distribution of malaria and pox infections might be shaped by similar anthropogenic factors at very local scales, as has been shown in other studies (Carrete *et al.*, 2009; González-Quevedo *et al.*, 2014; Padilla *et al.*, 2017) and both pathogens can be transmitted among wild birds via mosquito vectors (Akey *et al.*, 1981; Forrester, 1991). Alternatively, the co-occurrence of these pathogens may reflect pathogen induced immunosuppression facilitating subsequent infection by a second pathogen (synergistic interaction) (Schat & Skinner, 2013; Clark *et al.*, 2016). Many pathogenic infections, including malaria, are known to induce immunosuppression of the host response, increasing host susceptibility to other pathogens (Cox, 2001; Mabbott, 2018). An investigation of avian pox and malaria in Hawaiian birds by Samuel *et al.* (2018) concluded from epidemiological data that synergistic interactions or age-related acquisition were likely explanations for positive associations between these two pathogens. We believe the latter is unlikely in our system since age was not a significant predictor of malaria (Spurgin *et al.*, 2012) or pox infection status (Sheppard *et al.*, 2022).

We have been operating under the assumption that the covariation of SNP frequencies with malaria prevalence reflects pathogen-mediated selection. It's important to note, however, that the causal relationship might also be reversed; populations with high frequencies of specific protective SNPs could potentially be more resistant to malaria. The mechanisms identified in this study, such as the regulation of inflammation, metabolism, stress response, and cell death, likely represent broad defence mechanisms that could be activated during infections by various pathogens. If these mechanisms respond similarly to pox infection, then the cooccurrence patterns of avian malaria and pox across Berthelot's pipit populations could, theoretically, be attributed to spatial variation in underlying resistance or tolerance to

multiple pathogens. Nevertheless, we consider this scenario unlikely, given that the spatial cooccurrence patterns align with biogeographical rules (Spurgin *et al.* 2012) and are also likely to be influenced by additional environmental factors (González-Quevedo *et al.*, 2014). Regardless, the confounding factors that exist within the data presented here make it difficult to tease apart the initial drivers of pathogen cooccurrence.

Using the literature across taxa, we categorised genes identified in the GEA pipit studies (Sheppard *et al.*, 2022 and the present study; 27 in total) that might be overlap due to the collinearity between pox and malaria prevalence across populations. This categorisation was based on whether these genes have previously been directly linked to malaria or pox, or both (Table 3.2). Among these, five genes exhibited associations with both pox and malaria in Berthelot's pipit (Table 3.1 and 3.2). Two of these had previously been linked with both malaria and pox in other species and taxa, two had been linked with malaria only, and none had been linked exclusively to pox only in other species or taxa. For genes identified exclusively in either of our two pipit studies (i.e., for an association with malaria prevalence only, or for an association with pox prevalence only), only malaria-associated candidates could be linked with malaria and/or pox in other species or taxa. However, a lack of research and annotation for many genes is likely to be an issue. For example, it seems genes with easily identifiable functional pathways, such as the heat shock response, complement pathway, and corticosterone stress response, could clearly be linked to both pathogens. The greater number of genes linked to malaria does not necessarily imply a higher involvement of candidates in malaria response compared to pox response. It might reflect an imbalance in research efforts into malaria and pox viruses. Ideally, we would leverage the gradient of malaria and pox prevalence across populations of Berthelot's pipits (Figure 3.2) to determine the most suitable populations to try and isolate the effect of each pathogen. In many populations where one pathogen appears to be absent, the other is present but at low infection rates (*ca*. 1–5%). Consequently, selective pressure from that pathogen might be too weak to detect any associations. Outlier populations with significantly higher prevalence of one pathogen compared to the other, such as La Gomera and Tenerife (Figure 3.2), might provide the best conditions to disentangle these associations and further study the identified candidates. Indeed, extensive within-population investigations focused on the

candidate genes identified here, incorporating infection data for both pathogens, could help to verify these relationships.

For both pathogens, the molecular identification of strains was limited to a single genetic marker, the partial sequence of either the mitochondrial cytochrome *b* gene (malaria) or the *4b* core protein gene (pox). In some cases, even within a lineage, there can be considerable variation and population structure at other genes throughout the genome, including those potentially linked to infection success (Lauron *et al.*, 2014; Hellgren *et al.*, 2015; Huang *et al.*, 2019). It is therefore possible that undetected population structure in the pathogens among archipelagos or islands could drive the evolution of host defence at a finer geographic scale than considered in the current study, and as such, we would be unable to detect these archipelago/population-specific malaria-associated loci.

3.5.1. Conclusions

Our study provides correlational evidence that heterogenous pathogen-mediated selective pressure among populations of Berthelot's pipit drives spatial patterns of host genetic variation. It also highlights the methodological challenges involved in teasing apart the separate selective effects of different co-occurring pathogens (avian malaria and pox) in the same natural host system. Despite this, our findings suggest that genes linked to stress response, transcriptional regulation, metabolic processes and the innate immune system, may be involved in malaria response, or pathogen response more broadly, in this system. Any variation associated with malaria in this study needs to be followed up with extensive within-population investigations, for which we now have proposed a best candidate gene set.

Table 3.2. Evidence from the literature for links between the candidate genes identified in Berthelot's pipit and malaria and/or poxviruses in other species/taxa. Candidate genes listed here were identified by testing for associations with population-level avian malaria (in this study) or avian pox (Sheppard *et al.*, 2022) prevalence in Bayenv. Genes are first ordered according to the associated pathogen(s) and then by decreasing Bayes factor value.

Candidate gene	Associated pathogen	Links with poxviruses	Links with malaria
HSPA8	Pox & Malaria	HSP70 upregulated during poxvirus infection in mouse and human models (Jindal & Young, 1992; Sedger & Ruby, 1994; Brum *et al.*, 2003; Kowalczyk *et al.*, 2005; Cheng *et al.*, 2018)	HSP70 highly differentiated between malaria-naïve and exposed populations of a passerine bird (Cassin-Sackett *et al.*, 2019)
SMAD2	Pox & Malaria	–	Regulates inflammatory response as part of the TGF-β pathway; *TGF-β* itself inferred to be under divergent selection from avian malaria (Cassin-Sackett *et al.*, 2019)
UBE2H	Pox & Malaria	–	Overexpressed in malaria infected passerine, humans and mice (Videvall *et al.*, 2015)
CWC27	Pox & Malaria	Cyclophilins play a role against poxviruses in cell cultures (Castro *et al.*, 2003; Zhou *et al.*, 2021)	Overexpressed in malaria infected passerine (Videvall *et al.*, 2015)
MLYCD	Pox & Malaria	–	–
C8A	Malaria	Complement-related genes differentially expressed in pox infected passerine bird (McNew *et al.*, 2022). Complement may play a role in poxvirus response in mouse models (Miller *et al.*, 1995)	Complement system is activated against malaria (Silver *et al.*, 2010; Kurtovic *et al.*, 2020) and may act to control parasitaemia in passerine birds (Ellis *et al.*, 2015). Complement-related genes differentially expressed in malaria infected passerine (Videvall *et al.*, 2015)
C8B	Malaria		
SCAF4	Malaria	–	–

Table 3.2. (Continued).

Candidate gene	Associated pathogen	Links with poxviruses	Links with malaria
SOD1	Malaria	*Leporipoxvirus* disrupts host *SOD1* activity (Teoh *et al.*, 2003, 2005; Cao *et al.*, 2022)	Lower activity in infected birds (Jiménez-Peñuela *et al.*, 2023). Decreasing levels during course of infection in monkeys (Srivastava *et al.*, 1992). Downregulated in human infections (Reuling *et al.*, 2018) and associated with parasitaemia (Andrade *et al.*, 2010). Polymorphism may be associated with malaria susceptibility (Fernandes *et al.*, 2015).
MYLK3	Malaria	–	–
DUPD1	Malaria	–	–
KAT6B	Malaria	–	–
ZCCHC14	Malaria	–	–
ZNF423	Malaria	–	–
XRCC6BP1	Malaria	–	–
RPAP3	Malaria	–	–
CRHR2	Malaria	–	Modulates the release of corticosterone, the main glucocorticoid stress hormone in birds, for which high levels may result in higher malaria parasite load in a passerine bird (Names, Schultz, Krause, *et al.*, 2021)
ADGRD1	Malaria	–	–
ZNF598	Malaria	Co-opted by poxviruses in cell cultures for replication and the synthesis of viral proteins (DiGiuseppe *et al.*, 2018; Sundaramoorthy *et al.*, 2021)	–
SYNGR3	Malaria	–	–

Table 3.2. (Continued).

Candidate gene	Associated pathogen	Links with poxviruses	Links with malaria
MTHFD1L	Pox	–	–
ABLIM3	Pox	–	–
OAT	Pox	–	–
MFSD2A	Pox	–	–
MYCL	Pox	–	–
AKAP13	Pox	–	–

3.6. References

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 of the United States of America*, *101*, 3490–3494.

Akey, B. L., Nayar, J. K., & Forrester, D. J. (1981). Avian pox in Florida Wild Turkeys: *Culex nigripalpus* and *Wyeomyia vanduzeei* as experimental vectors. *Journal of Wildlife Diseases*, *17*, 597–599.

Alley, M. R., Ha, H. J., Howe, L., Hale, K. A., & Cash, W. (2010). Concurrent avian malaria and avipox virus infection in translocated South Island saddlebacks (*Philesturnus carunculatus carunculatus*). *New Zealand Veterinary Journal*, *58*, 218–223.

Anderson, R. M., & May, R. M. (1979). Population biology of infectious diseases: Part I. *Nature*, *280*, 361–367.

Andrade, B. B., Reis-Filho, A., Souza-Neto, S. M., Raffaele-Netto, I., Camargo, L. M. A., Barral, A., & Barral-Netto, M. (2010). Plasma superoxide dismutase-1 as a surrogate marker of vivax malaria severity. *PLoS Neglected Tropical Diseases*, *4*, e650.

Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution*, *9*, 12482–12502.

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

Asghar, M., Hasselquist, D., & Bensch, S. (2011). Are chronic avian haemosporidian infections costly in wild birds? *Journal of Avian Biology*, *42*, 530–537.

Asghar, M., Hasselquist, D., Hansson, B., Zehtindjiev, P., Westerdahl, H., & Bensch, S. (2015). Hidden costs of infection: Chronic malaria accelerates telomere degradation and senescence in wild birds. *Science*, *347*, 436–438.

Atkinson, C. T., Lease, J. K., Dusek, R. J., & Samuel, M. D. (2005). Prevalence of pox-like lesions and malaria in forest bird communities on Leeward Mauna Loa Volcano, Hawaii. *The Condor*, *107*, 537–546.

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.

Atkinson, C. T., Woods, K. L., Dusek, R. J., Sileo, L. S., & Iko, W. M. (1995). Wildlife disease and conservation in Hawaii: Pathogenicity of avian malaria (*Plasmodium relictum*) in experimentally infected Iiwi (*Vestiaria coccinea*). *Parasitology*, *111*, 59–69.

Bayly-Jones, C., Bubeck, D., & Dunstone, M. A. (2017). The mystery behind membrane insertion: A review of the complement membrane attack complex. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*, 20160221.

Beaumont, M. A. (2005). Adaptation and speciation: What can Fst tell us? *Trends in Ecology*

and Evolution, *20*, 435–440.

Bilgin, R., Yalcin, M. S., Yucebilgic, G., Koltas, I. S., & Yazar, S. (2012). Oxidative stress in vivax malaria. *Korean Journal of Parasitology*, *50*, 375–377.

Blair, L. M., Granka, J. M., & Feldman, M. W. (2014). On the stability of the Bayenv method in assessing human SNP-environment associations. *Human Genomics*, *8*, 1.

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

Boštjančić, L. L., Francesconi, C., Rutz, C., Hoffbeck, L., Poidevin, L., Kress, A., Jussila, J., *et al.* (2022). Host-pathogen coevolution drives innate immune response to *Aphanomyces astaci* infection in freshwater crayfish: Transcriptomic evidence. *BMC Genomics*, *23*, 600.

Brooks, M. E., Kristensen, K., Van Benthem, K. J., Magnusson, A., Berg, C. W., Nielsen, A., Skaug, H. J., *et al.* (2017). glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. *The R Journal*, *9*, 378–400.

Brouwer, L., Barr, I., Van De Pol, M., 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.

Brum, L. M., Lopez, M. C., Varela, J. C., Baker, H. V., & Moyer, R. W. (2003). Microarray analysis of A549 cells infected with rabbitpox virus (RPV): A comparison of wild-type RPV and RPV deleted for the host range gene, SPI-1. *Virology*, *315*, 322–334.

Busetto, V., Barbosa, I., Basquin, J., Marquenet, É., Hocq, R., Hennion, M., Paternina, J. A., *et al.* (2020). Structural and functional insights into CWC27/CWC22 heterodimer linking the exon junction complex to spliceosomes. *Nucleic Acids Research*, *48*, 5670–5683.

Cao, J., Teoh, M., Moon, M., & McFadden, G. (2022). *Leporipoxvirus* Cu-Zn superoxide dismutase homologs inhibit cellular superoxide dismutase, but are not essential for virus replication or virulence. *Virology*, *296*, 125–135.

Carrete, M., Serrano, D., Illera, J. C., López, G., Vögeli, M., Delgado, A., & Tella, J. L. (2009). Goats, birds, and emergent diseases: Apparent and hidden effects of exotic species in an island environment. *Ecological Applications*, *19*, 840–853.

Cassandri, M., Smirnov, A., Novelli, F., Pitolli, C., Agostini, M., Malewicz, M., Melino, G., *et al.* (2017). Zinc-finger proteins in health and disease. *Cell Death Discovery*, *3*, 1–12.

Cassin-Sackett, L., Callicrate, T. E., & Fleischer, R. C. (2019). Parallel evolution of gene classes, but not genes: Evidence from Hawai'ian honeycreeper populations exposed to avian malaria. *Molecular Ecology*, *28*, 568–583.

Castro, A. P. V., Carvalho, T. M. U., Moussatché, N., & Damaso, C. R. A. (2003). Redistribution of cyclophilin A to viral factories during vaccinia virus infection and its incorporation into mature particles. *Journal of Virology*, *77*, 9052–9068.

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

Charlesworth, D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, *2*, e64.

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.

Cheng, W., Jia, H., Wang, X., He, X., Jin, Q., Cao, J., & Jing, Z. (2018). Ectromelia virus upregulates the expression of heat shock protein 70 to promote viral replication. *International Journal of Molecular Medicine*, *42*, 1044–1053.

Chu, Z. J., Wang, Y. J., Ying, S. H., Wang, X. W., & Feng, M. G. (2016). Genome-wide host-pathogen interaction unveiled by transcriptomic response of Diamondback moth to fungal infection. *PLoS ONE*, *11*, e0152908.

Ciloglu, A., Ellis, V. A., Bernotienė, R., Valkiūnas, G., & Bensch, S. (2019). A new one-step multiplex PCR assay for simultaneous detection and identification of avian haemosporidian parasites. *Parasitology Research*, *118*, 191–201.

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.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

Cox, F. E. G. (2001). Concomitant infections, parasites and immune responses. *Parasitology*, *122*, S23–S28.

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.

Daszak, P., Cunningham, A. A., & Hyatt, A. D. (2000). Emerging infectious diseases of wildlife - Threats to biodiversity and human health. *Science*, *287*, 443–449.

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

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

Delhaye, J., Jenkins, T., & Christe, P. (2016). *Plasmodium* infection and oxidative status in breeding great tits, *Parus major*. *Malaria Journal*, *15*, 531.

DiGiuseppe, S., Rollins, M. G., Bartom, E. T., & Walsh, D. (2018). ZNF598 plays distinct roles in interferon-stimulated gene expression and poxvirus protein synthesis. *Cell Reports*, *23*, 1249.

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

Douma, J. C., & Weedon, J. T. (2019). Analysing continuous proportions in ecology and evolution: A practical introduction to beta and Dirichlet regression. *Methods in Ecology and Evolution*, *10*, 1412–1430.

Ellis, V. A., Cornet, S., Merrill, L., Kunkel, M. R., Tsunekage, T., & Ricklefs, R. E. (2015). Host

immune responses to experimental infection of *Plasmodium relictum* (lineage SGS1) in domestic canaries (*Serinus canaria*). *Parasitology Research*, *114*, 3627–3636.

Fernandes, R. C., Hasan, M., Gupta, H., Geetha, K., Rai, P. S., Hande, M. H., D'Souza, S. C., *et al.* (2015). Host genetic variations in glutathione-S-transferases, superoxide dismutases and catalase genes influence susceptibility to malaria infection in an Indian population. *Molecular Genetics and Genomics*, *290*, 1155–1168.

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

Fischer, U., & Meese, E. (2007). Glioblastoma multiforme: The role of DSB repair between genotype and phenotype. *Oncogene*, *26*, 7809–7815.

Forrester, D. J. (1991). The ecology and epizootiology of avian pox and malaria in Wild Turkeys. *Bulletin of the Society for Vector Ecology*, *16*, 127–148.

Foster, J. T., Woodworth, B. L., Eggert, L. E., Hart, P. J., Palmer, D., Duffy, D. C., & Fleischer, R. C. (2007). Genetic structure and evolved malaria resistance in Hawaiian honeycreepers. *Molecular Ecology*, *16*, 4738–4746.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. (2013). Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*, 1687–1699.

Fridavich, I. (1995). Superoxide radical and superoxide dismutases. *Annual Review of Biochemistry*, *64*, 97–112.

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.

Garroway, C. J., Radersma, R., Sepil, I., Santure, A. W., De Cauwer, I., Slate, J., & Sheldon, B. C. (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.

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*, 1555–1579.

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., 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.

Gunther, T., & Coop, G. (2018). *A Short Manual for Bayenv2.0*.

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele

frequencies. *Genetics*, *195*, 205–220.

Ha, H. J., Banda, M., Alley, M. R., Howe, L., & Gartrell, B. D. (2012). The seroprevalence of avipoxvirus and its association with avian malaria (*Plasmodium* spp.) infection in introduced passerine birds in the southern regions of the North Island of New Zealand. *Avian Diseases*, *57*, 109–115.

Hancock, A. M., Witonsky, D. B., Ehler, E., Alkorta-Aranburu, G., Beall, C., Gebremedhin, A., Sukernik, R., *et al.* (2010). Human adaptations to diet, subsistence, and ecoregion are due to subtle shifts in allele frequency. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 8924–8930.

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*.

Hellard, E., Fouchet, D., Vavre, F., & Pontier, D. (2015). Parasite-parasite interactions in the wild: How to detect them? *Trends in Parasitology*, *31*, 640–652.

Hellgren, O., Atkinson, C. T., Bensch, S., Albayrak, T., Dimitrov, D., Ewen, J. G., Kim, K. S., *et al.* (2015). Global phylogeography of the avian malaria pathogen *Plasmodium relictum* based on MSP1 allelic diversity. *Ecography*, *38*, 842–850.

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

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., *et al.* (2016). Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *American Naturalist*, *188*, 379–397.

Huang, X., Rapševičius, P., Chapa-Vargas, L., Hellgren, O., & Bensch, S. (2019). Within-lineage divergence of avian haemosporidians: A case study to reveal the origin of a widespread *Haemoproteus* parasite. *Journal of Parasitology*, *105*, 414–422.

Huang, Y., Li, Y., Burt, D. W., Chen, H., Zhang, Y., Qian, W., Kim, H., *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.

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.

Jiménez-Peñuela, J., Ferraguti, M., Martínez-De La Puente, J., Soriguer, R. C., & Figuerola, J. (2023). Oxidative status in relation to blood parasite infections in house sparrows living along an urbanization gradient. *Environmental Pollution*, *316*, 120712.

Jindal, S., & Young, R. A. (1992). Vaccinia virus infection induces a stress response that leads

to association of Hsp70 with viral proteins. *Journal of Virology*, *66*, 5357–5362.

Johnson, P. T. J., & Buller, I. D. (2011). Parasite competition hidden by correlated coinfection: Using surveys and experiments to understand parasite interactions. *Ecology*, *92*, 535–541.

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

Kowalczyk, A., Guzik, K., Slezak, K., Dziedzic, J., & Rokita, H. (2005). Heat shock protein and heat shock factor 1 expression and localization in vaccinia virus infected human monocyte derived macrophages. *Journal of Inflammation*, *2*, 12.

Kurtovic, L., Boyle, M. J., Opi, D. H., Kennedy, A. T., Tham, W. H., Reiling, L., Chan, J. A., *et al.* (2020). Complement in malaria immunity and vaccines. *Immunological Reviews*, *293*, 38–56.

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.

Lapointe, D. A., Atkinson, C. T., & Samuel, M. D. (2012). Ecology and conservation biology of avian malaria. *Annals of the New York Academy of Sciences*, *1249*, 211–226.

Lauron, E. J., Oakgrove, K. S., Tell, L. A., Biskar, K., Roy, S. W., & Sehgal, R. N. M. (2014). Transcriptome sequencing and analysis of *Plasmodium gallinaceum* reveals polymorphisms and selection on the apical membrane antigen-1. *Malaria Journal*, *13*, 382.

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.

Lochmiller, R. L., & Deerenberg, C. (2000). Trade-offs in evolutionary immunology: Just what is the cost of immunity? *Oikos*, *88*, 87–98.

Loiseau, C., Richard, M., Garnier, S., Chastel, O., Julliard, R., Zoorob, R., & Sorci, G. (2009). Diversifying selection on MHC class I in the house sparrow (*Passer domesticus*). *Molecular Ecology*, *18*, 1331–1340.

Longley, R., Smith, C., Fortin, A., Berghout, J., McMorran, B., Burgio, G., Foote, S., *et al.* (2011). Host resistance to malaria: Using mouse models to explore the host response. *Mammalian Genome*, *22*, 32–42.

Lotterhos, K. E., & Whitlock, M. C. (2015). The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, *24*, 1031–1046.

Mabbott, N. A. (2018). The influence of parasite infections on host immunity to co-infection with other pathogens. *Frontiers in Immunology*, *9*, 2579.

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

MalariaGEN. (2019). Insights into malaria susceptibility using genome-wide data on 17,000

individuals from Africa, Asia and Oceania. *Nature Communications*, *10*, 5732.

Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and within-archipelago adaptation between island bird populations. *Royal Society Open Science*, *8*, 201146.

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging island populations of a bird. *Molecular Ecology*, *32*, 1972–1989.

Martin, L. B., Scheuerlein, A., & Wikelski, M. (2003). Immune activity elevates energy expenditure of house sparrows: A link between direct and indirect costs? *Proceedings of the Royal Society of London B*, *270*, 153–158.

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.

McNew, S. M., Loyola, D. C., Yepez, J., Andreadis, C., Gotanda, K., Saulsberry, A., & Fessl, B. (2022). Transcriptomic responses of Galápagos finches to avian poxvirus infection. *Molecular Ecology*, *31*, 5552–5567.

Miller, C. G., Justus, D. E., Jayaraman, S., & Kotwal, G. J. (1995). Severe and prolonged inflammatory response to localized cowpox virus infection in footpads of C5-deficient mice: Investigation of the role of host complement in poxvirus pathogenesis. *Cellular Immunology*, *162*, 326–332.

Mukhin, A., Palinauskas, V., Platonova, E., Kobylkov, D., Vakoliuk, I., & Valkiunas, G. (2016). The strategy to survive primary malaria infection: An experimental study on behavioural changes in parasitized birds. *PLoS ONE*, *11*, e0159216.

Musa, S., Mackenstedt, U., Woog, F., & Dinkel, A. (2022). Untangling the actual infection status: detection of avian haemosporidian parasites of three Malagasy bird species using microscopy, multiplex PCR, and nested PCR methods. *Parasitology Research*, *121*, 2817–2829.

Names, G. R., Schultz, E. M., Hahn, T. P., Hunt, K. E., Angelier, F., Ribout, C., & Klasing, K. C. (2021). Variation in immunity and health in response to introduced avian malaria in an endemic Hawaiian songbird. *Animal Conservation*, *25*, 455–466.

Names, G. R., Schultz, E. M., Krause, J. S., Hahn, T. P., Wingfield, J. C., Heal, M., Cornelius, J. M., *et al.* (2021). Stress in paradise: Effects of elevated corticosterone on immunity and avian malaria resilience in a Hawaiian passerine. *The Journal of Experimental Biology*, *224*, 242951.

Narsaria, N., Mohanty, C., Das, B. K., Mishra, S. P., & Prasad, R. (2012). Oxidative stress in children with severe malaria. *Journal of Tropical Pediatrics*, *58*, 147–150.

Nneji, C. M., Adaramoye, O. A., Falade, C. O., & Ademowo, O. G. (2013). Effect of chloroquine, methylene blue and artemether on red cell and hepatic antioxidant defence system in mice infected with *Plasmodium yoelii nigeriensis*. *Parasitology Research*, *112*, 2619–2625.

Ots, I., Kerimov, A. B., Ivankina, E. V., Ilyina, T. A., & Hõrak, P. (2001). Immune challenge affects basal metabolic activity in wintering great tits. *Proceedings of the Royal Society of London B*, *268*, 1175.

Padilla, D. P., Illera, J. C., Gonzalez-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.

Palinauskas, V., Valkiunas, G., Bolshakov, C. V., & Bensch, S. (2008). *Plasmodium relictum* (lineage P-SGS1): Effects on experimentally infected passerine birds. *Experimental Parasitology*, *120*, 372–380.

Pamplona, R., & Costantini, D. (2011). Molecular and structural antioxidant defenses against oxidative stress in animals. *American Journal of Physiology - Regulatory Integrative and Comparative Physiology*, *301*, 843–863.

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, *6*, 445–464.

Paterson, S., Vogwill, T., Buckling, A., Benmayor, R., Spiers, A. J., Thomson, N. R., Quail, M., *et al.* (2010). Antagonistic coevolution accelerates molecular evolution. *Nature*, *464*, 275–278.

Pigeault, R., Cozzarolo, C., Glaizot, O., & Christe, P. (2020). Effect of age, haemosporidian infection and body condition on pair composition and reproductive success in Great Tits *Parus major*. *Ibis*, *162*, 613–626.

Popa, G. L., & Popa, M. I. (2021). Recent advances in understanding the inflammatory response in malaria: A review of the dual role of cytokines. *Journal of Immunology Research*, *2021*, 7785180.

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

Quin, J. E., Bujila, I., Chérif, M., Sanou, G. S., Qu, Y., Homann, M. V., Rolicka, A., *et al.* (2017). Major transcriptional changes observed in the Fulani, an ethnic group less susceptible to malaria. *ELife*, *6*, 29156.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Råberg, L., Graham, A. L., & Read, A. F. (2009). Decomposing health: Tolerance and resistance to parasites in animals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 37–49.

Råberg, L., Sim, D., & Read, A. F. (2007). Disentangling genetic variation for resistance and tolerance to infectious diseases in animals. *Science*, *318*, 812–814.

Ravenhall, M., Campino, S., Sepúlveda, N., Manjurano, A., Nadjm, B., Mtove, G., Wangai, H., *et al.* (2018). Novel genetic polymorphisms associated with severe malaria and under selective pressure in North-eastern Tanzania. *PLoS Genetics*, *14*, e1007172.

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*,

24, 4348–4370.

Reuling, I. J., de Jong, G. M., Yap, X. Z., Asghar, M., Walk, J., van de Schans, L. A., Koelewijn, R., *et al.* (2018). Liver injury in uncomplicated malaria is an overlooked phenomenon: An observational study. *EBioMedicine*, *36*, 131–139.

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.

Rockett, K. A., Clarke, G. M., Fitzpatrick, K., Hubbart, C., Jeffreys, A. E., Rowlands, K., Craik, R., *et al.* (2014). Reappraisal of known malaria resistance loci in a large multicenter study. *Nature Genetics*, *46*, 1197–1204.

Roestenberg, M., McCall, M., Mollnes, T. E., van Deuren, M., Sprong, T., Klasen, I., Hermsen, C. C., *et al.* (2007). Complement activation in experimental human malaria infection. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, *101*, 643–649.

Rosenblum, E. B., Poorten, T. J., Settles, M., & Murdoch, G. K. (2012). Only skin deep: Shared genetic response to the deadly chytrid fungus in susceptible frog species. *Molecular Ecology*, *21*, 3110–3120.

Samuel, M. D., Woodworth, B. L., Atkinson, C. T., Hart, P. J., & LaPointe, D. A. (2018). The epidemiology of avian pox and interaction with avian malaria in Hawaiian forest birds. *Ecological Monographs*, *88*, 621–637.

Scaccabarozzi, D., Deroost, K., Corbett, Y., Lays, N., Corsetto, P., Salè, F. O., Van Den Steen, P. E., *et al.* (2018). Differential induction of malaria liver pathology in mice infected with *Plasmodium chabaudi* AS or *Plasmodium berghei* NK65. *Malaria Journal*, *17*, 18.

Schat, K. A., & Skinner, M. A. (2013). 'Avian immunosuppressive diseases and immunoevasion' in Schat, K. A., Kaspers, B., & Kaiser, P. (eds) *Avian Immunology* (2nd edn). Cambridge: Academic Press, 275–297.

Schoville, S. D., Bonin, A., François, O., Lobreaux, S., Melodelima, C., & Manel, S. (2012). Adaptive genetic variation on the landscape: Methods and cases. *Annual Review of Ecology, Evolution, and Systematics*, *43*, 23–43.

Sedger, L., & Ruby, J. (1994). Heat shock response to vaccinia virus infection. *Journal of Virology*, *68*, 4685–4689.

Sehgal, R. N. M. (2015). Manifold habitat effects on the prevalence and diversity of avian blood parasites. *International Journal for Parasitology: Parasites and Wildlife*, *4*, 421–430.

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.

Sheppard, E. C., Martin, C. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2022). Genomic associations with pox virus across divergent populations in Berthelot's pipit. *Molecular Ecology*, *31*, 3154–3173.

Silver, K. L., Higgins, S. J., Mcdonald, C. R., & Kain, K. C. (2010). Complement driven innate immune response to malaria: Fuelling severe malarial diseases. *Cellular Microbiology*,

12, 1036–1045.

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*, 16–34.

Sottile, M. L., & Nadin, S. B. (2018). Heat shock proteins and DNA repair mechanisms: An updated overview. *Cell Stress and Chaperones*, *23*, 303–315.

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

Srivastava, P., Puri, S. K., Dutta, G. P., & Pandey, V. C. (1992). Status of oxidative stress and antioxidant defences during *Plasmodivm knowlesi* infection and chloroquine treatment in *Macaca mulatta*. *International Journal for Parasitology*, *22*, 243–245.

Stewart, M. D., Ritterhoff, T., Klevit, R. E., & Brzovic, P. S. (2016). E2 enzymes: More than just middle men. *Cell Research*, *26*, 423–440.

Sundaramoorthy, E., Ryan, A. P., Fulzele, A., Leonard, M., Daugherty, M. D., & Bennett, E. J. (2021). Ribosome quality control activity potentiates vaccinia virus protein synthesis during infection. *Journal of Cell Science*, *134*, 257188.

Szabó, C. (2003). Multiple pathways of peroxynitrite cytotoxicity. *Toxicology Letters*, *140–141*, 105–112.

Teoh, M. L. T., Turner, P. V., & Evans, D. H. (2005). Tumorigenic poxviruses up-regulate intracellular superoxide to inhibit apoptosis and promote cell proliferation. *Journal of Virology*, *79*, 5799–5811.

Teoh, M. L. T., Walasek, P. J., & Evans, D. H. (2003). *Leporipoxvirus* Cu,Zn-superoxide dismutase (SOD) homologs are catalytically inert decoy proteins that bind copper chaperone for SOD. *The Journal of Biological Chemistry*, *278*, 33175–33184.

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

Valkiunas, G. (2005). Avian Malaria Parasites and Other Haemosporidia. Boca Raton: CRC Press.

van de Crommenacker, J., Richardson, D. S., Koltz, A. M., Hutchings, K., & Komdeur, J. (2012). Parasitic infection and oxidative status are associated and vary with breeding activity in the Seychelles warbler. *Proceedings of the Royal Society B: Biological Sciences*, *279*, 1466–1476.

Vasquez, M., Zuniga, M., & Rodriguez, A. (2021). Oxidative stress and pathogenesis in malaria. *Frontiers in Cellular and Infection Microbiology*, *11*, 768182.

Videvall, E., Cornwallis, C. K., Palinauskas, V., Valkiunas, G., & Hellgren, O. (2015). The avian transcriptome response to malaria infection. *Molecular Biology and Evolution*, *32*, 1255.

Videvall, E., Palinauskas, V., Valkiūnas, G., & Hellgren, O. (2020). Host transcriptional responses to high- and low-virulent avian malaria parasites. *American Naturalist*, *195*, 1079–1084.

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.

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

Yorinks, N., Atkinson, T., & Carter, T. (2000). Effects of malaria on activity budgets of experimentally infected juvenile apapane (*Himatione sanguinea*). *The Auk*, *117*, 731–738.

Zhou, Y., Gao, F., Lv, L., Wang, S., He, W., Lan, Y., Li, Z., *et al.* (2021). Host factor cyclophilin B affects Orf virus replication by interacting with viral ORF058 protein. *Veterinary Microbiology*, *258*, 109099.

Zueva, K. J., Lumme, J., Veselov, A. E., Kent, M. P., Lien, S., & Primmer, C. R. (2014). Footprints of directional selection in wild Atlantic salmon populations: Evidence for parasite-driven evolution? *PLoS ONE*, *9*, 91672.

Zueva, K. J., Lumme, J., Veselov, A. E., Kent, M. P., & Primmer, C. R. (2018). Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). *Marine Genomics*, *39*, 26–38.

3.7. Supplements

Table S3.1. Sampling periods for avian malaria (*Plasmodium*) infection across the 13 populations of Berthelot's pipit across Macaronesia and sampling distribution for ddRAD-seq data.

Archipelago	Island	Year	Month(s)	*n*	*n* ddRAD-seq data
Madeira	Deserta Grande	2006	Sept–Oct	31	18
		2009	Apr, Sept–Oct	4	2
	Madeira	2005	Apr	1	
		2006	Sept	32	
		2009	Mar–Apr, Sept–Oct	29	20
		2016	Apr–May	29	
	Porto Santo	2006	Sept	31	3
		2009	Mar	30	17
		2016	May–June	129	
Selvagens	Selvagem Grande	2005	Apr	52	
		2009	Mar	42	20
Canary Islands	La Graciosa	2006	Mar	24	2
		2009	Jan	26	18
	Lanzarote	2006	Mar	13	
		2008	Nov	1	
		2009	Jan	30	20
		2019	Apr	83	
		2020	Feb–Apr	113	
	Fuerteventura	2006	Feb–Mar	12	2
		2008	Nov	1	
		2009	Jan	30	18
	Gran Canaria	2006	Jan–Feb	31	
		2009	Jan–Feb	33	20
	El Teide	2006	Mar–Apr	30	
		2008	Oct	3	
		2009	Apr–May, Aug	22	20
		2011	Mar–Apr	30	
	Tenerife	2005	May	7	
		2006	Feb–Mar	25	8
		2009	Jan–Aug	35	14
		2010	Apr–May	96	
		2011	Jan–May	358	
	La Gomera	2006	Feb	30	7
		2009	Jan, Apr	20	13
	La Palma	2006	Jan–Feb	28	3
		2009	Feb, July	22	17
	El Hierro	2006	Jan–Feb	31	
		2009	Feb	30	20

Table S3.2. Prevalence of avian malaria (*Plasmodium*) in 13 populations of Berthelot's pipit across Macaronesia. Prevalence was estimated from field data collected from 2005–2020 (the number of individuals identified as PCR positive divided by the total number of individuals caught throughout the entire period). Malaria residuals were extracted from a zero-inflated beta regression model between population-level avian pox and malaria. Values were standardised by subtracting the mean and dividing by the standard deviation across populations. Standardised variables were then used as input for genotype environment association analyses.

Archipelago	**Island**	***n***	**Malaria prevalence %**	**Standardised prevalence**	**Residual malaria**	**Standardised residual malaria**
Madeira	Deserta Grande	35	0.00	-0.90	-0.076	-0.75
	Madeira	91	1.10	-0.85	-0.065	-0.67
	Porto Santo	190	66.32	1.76	0.102	0.52
Selvagens	Selvagem Grande	94	0.00	-0.90	-0.076	-0.75
Canary Islands	La Graciosa	50	2.00	-0.82	-0.056	-0.60
	Lanzarote	240	57.92	1.43	0.193	1.17
	Fuerteventura	43	44.19	0.88	0.012	-0.12
	Gran Canaria	64	28.13	0.23	-0.077	-0.75
	El Teide	85	2.35	-0.80	-0.066	-0.67
	Tenerife	521	36.08	0.55	0.225	1.39
	La Gomera	50	46.00	0.95	0.342	2.23
	La Palma	50	2.00	-0.82	-0.056	-0.60
	El Hierro	61	4.92	-0.70	-0.027	-0.40

Table S3.3. SNPs identified as being associated with 'residual malaria' prevalence using Bayenv, ordered by Bayes factor (BF) values. Residuals were extracted from a zero-inflated beta regression model between avian pox and malaria prevalence.

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
115s49	81.6	15:4213461	*FICD*	2,410 bp upstream	adenosine monophosphate-protein transferase FICD	Modulates the activity of specific protein chaperones through AMPylation, contributing to protein homeostasis
1343s47	76.7	24:2805042	*GRIK4*	In gene (intron)	glutamate receptor ionotropic, kainate 4	Encodes a receptor subunit essential for glutamate-gated transmembrane ion transport and intercellular synaptic signalling
1710s69	19.5	21:4658591	*CASZ1*	In gene (intron)	zinc finger protein castor homolog 1	Acts as a transcription factor, playing a critical role in regulating gene expression and cellular development. Notably, it has been linked to variations in blood pressure among humans and vision in fish
146s79	18.5	2:4280785	*CRHR2*	In gene (intron)	corticotropin releasing hormone receptor 2	Binds corticotrophin-releasing factor, modulating hormonal release and coordinating physiological changes, thereby mediating the physiological responses to stress
3891s33	15.0	13:5125125	–	–	–	–
315s88	14.1	Z:10812737	*LOC100223460*	In pseudogene	Uncharacterised	–
2996s115	13.4	4A:5948905	–	–	–	–
3664s59	9.2	5:36381483	–	–	–	–

Table S3.3. (Continued).

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
451s74	6.2	8:23610708	*C8A* *C8B*	In gene (intron) 8,722 bp downstream	complement C8 alpha chain complement C8 beta chain	Encodes subunits of the complement system component essential for the formation of the membrane attack complex (MAC). The MAC serves as an important immune effector, creating pores in the membranes of target cells, leading to their destruction
647s104	6.0	4A:16665555	*LOC100221727*	9,542 bp upstream	ras-related GTP-binding protein A	Regulates the mTORC1 signalling cascade based on amino acid availability, thereby playing a role in modulating cell growth and metabolism
86s75	5.6	21:5714453	*PLEKHN1*	In gene (intron)	pleckstrin homology domain-containing family N member 1	Enables phospholipid binding activity, positively regulates apoptotic processes, and is suggested to play a potential role in inflammatory processes
343s91	5.4	4A:5651803	*EDA*	In gene (intron)	ectodysplasin-A	Plays a crucial role in signalling pathways that regulate developmental processes, notably influencing the formation of ectodermal tissues across various taxa (e.g., teeth, hair follicles, feathers, scales, armour)
932s117	5.1	3:73368053	–	–	–	–

Table S3.3. (Continued).

SNP	BF	Genomic location*	Candidate gene(s)†	Distance from gene	Gene name	Putative function
1157s37	4.9	22:655112	*SLC23A2*	In gene (intron)	solute carrier family 23 member 2	Sodium/ascorbate cotransporter, crucial for maintaining intracellular levels of vitamin C, potentially supporting the immune system and contributing to cellular protection against oxidative stress
2850s103	4.7	10:8024284	*WDR72*	8,709 bp upstream	WD repeat-containing protein 72	Role and function generally unknown; some link to enamel formation in humans. Notably well-conserved in birds but selective pressures are unclear
1695s90	4.5	1:64912593	–	–	–	–
3269s98	4.5	2:58237224	–	–	–	–
3055s27	4.3	15:13669057	*RTN4R*	In gene (intron)	reticulon-4 receptor	Binds proteins associated with the inhibition of neural regeneration, thereby modulating nerve cell responses, contributing to neural development and regeneration

* Chromosome location and base position based on *Taeniopygia_guttata*-3.2.4 genome assembly

† Within 10 kbp of focal SNP

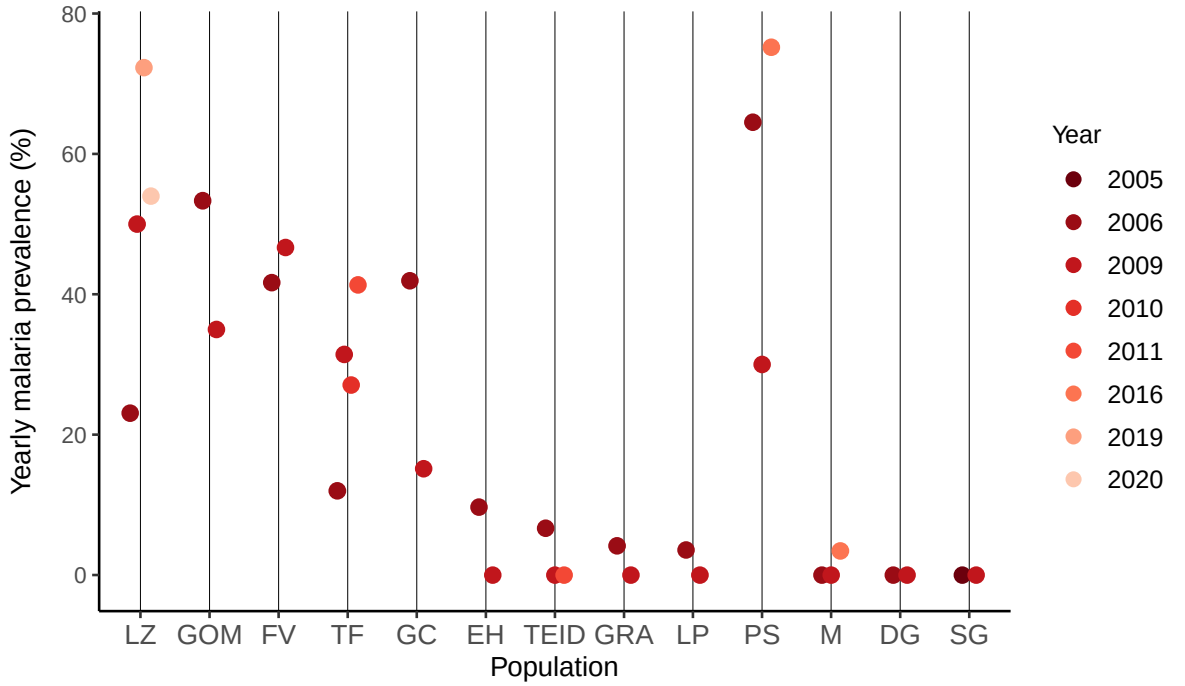

Figure S3.1. Dot plot showing temporal patterns of malaria (*Plasmodium*) prevalence across island populations of Berthelot's pipit. All populations have been sampled at least twice. Prevalence values only included where $n > 10$. LZ = Lanzarote; GOM = La Gomera; FV = Fuerteventura; TF = Tenerife; GC = Gran Canaria; EH = El Hierro; TEID = El Teide; GRA = La Graciosa; LP = La Palma; PS = Porto Santo; M = Madeira; DG = Deserta Grande; SG = Selvagem Grande.

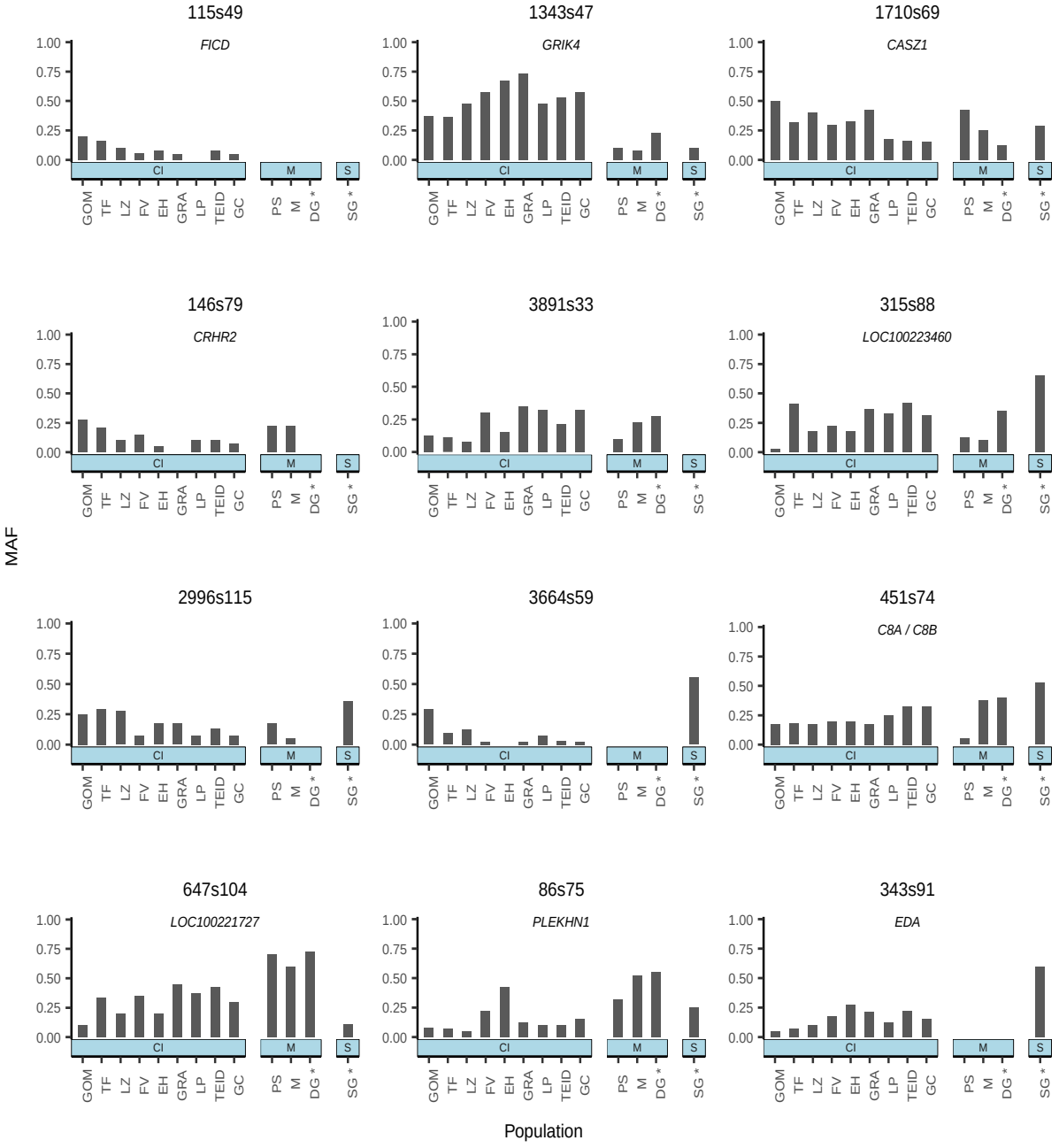

Figure S3.2. Minor allele frequency (MAF) of 18 candidate SNPs identified by testing for associations with 'pox-corrected' residual malaria across Berthelot's pipit populations in Bayenv. Residuals were extracted from a zero-inflated beta regression model between avian pox and malaria prevalence. Bayenv corrects for neutral population structure among allele frequencies in the analysis. Nearby genes are noted below the SNP names. Populations are grouped by archipelago (CI = Canary Islands; M = Madeira; S = Selvagens) then ordered according to residual malaria (highest–lowest). Malaria-free populations are indicated by an asterisk. Acronyms: GOM = La Gomera; TF = Tenerife; LZ = Lanzarote; FV = Fuerteventura; EH = El Hierro; GRA = La Graciosa; LP = La Palma; TEID = El Teide; GC = Gran Canaria; PS = Porto Santo; M = Madeira; DG = Deserta Grande; SG = Selvagem Grande.

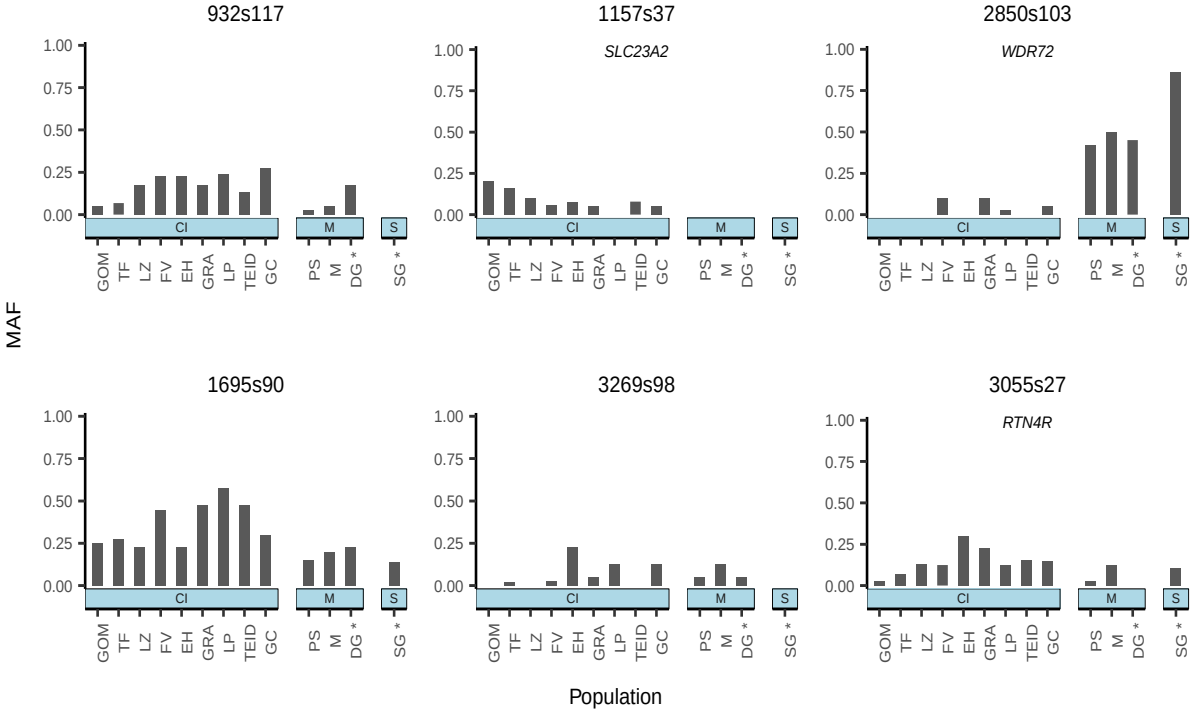

Figure S3.2. (Continued). Minor allele frequency (MAF) of 18 candidate SNPs identified by testing for associations with 'pox-corrected' residual malaria across Berthelot's pipit populations in Bayenv. Residuals were extracted from a zero-inflated beta regression model between avian pox and malaria prevalence. Bayenv corrects for neutral population structure among allele frequencies in the analysis. Nearby genes are noted below the SNP names. Populations are grouped by archipelago (CI = Canary Islands; M = Madeira; S = Selvagens) then ordered according to residual malaria (highest–lowest). Malaria-free populations are indicated by an asterisk. Acronyms: GOM = La Gomera; TF = Tenerife; LZ = Lanzarote; FV = Fuerteventura; EH = El Hierro; GRA = La Graciosa; LP = La Palma; TEID = El Teide; GC = Gran Canaria; PS = Porto Santo; M = Madeira; DG = Deserta Grande; SG = Selvagem Grande.

Chapter 4

Genomic erosion in oceanic island populations of an endemic bird: A spatio-temporal analysis

Century-old specimens of Berthelot's pipit at the Natural History Museum, London

4.1. Abstract

Understanding the impact of historical demographic events and ongoing genetic processes on biodiversity patterns is important for informing conservation strategies and advancing our understanding of evolution. This is particularly so in island species, which typically exist in small, isolated populations with limited genetic diversity, making them particularly susceptible to genetic erosion and environmental pressures. We combined historical (museomics) and contemporary genomic data to examine spatio-temporal patterns of genomic variation and population differentiation in Berthelot's pipits (*Anthus berthelotii*) across 10 endemic island populations in three archipelagos. Our findings show a significant decline in genetic diversity (autosomal heterozygosity) across all populations over a period of *ca.* 120 years, averaging a 20% loss of genetic variation. Temporal analyses using PCA and F_{ST} indicate ongoing differentiation within populations over this period, likely influenced by genetic drift and bottleneck effects. Among the archipelagos, the smallest and most isolated (Selvagens), which is also the most recent site of pipit colonisation, exhibited the lowest genetic diversity and greatest differentiation from other populations. This pattern may reflect historical founder events, local demographic fluctuations and potentially more recent environmental pressures, including anthropogenic effects. In the Canary Islands, different populations exhibited considerable similarity in genetic diversity at each time point, and low differentiation from each other, suggesting the possibility of shared environmental pressures and/or sufficient intra-archipelago gene flow that may have mitigated isolation effects. Our study underscores the potential of historical versus contemporary genomic comparisons to reveal the interplay between demographic history, environmental changes, and genomic dynamics. Furthermore, it highlights the importance of ongoing genomic monitoring in seemingly stable populations to better understand the evolutionary trajectories and genetic health of isolated populations.

4.2. Introduction

The rapid acceleration of human-induced changes in natural ecosystems (Venter *et al.*, 2016; Otto, 2018) underscores the need to investigate how historical demographic events and ongoing population genetic processes may shape existing biodiversity patterns. While anthropogenic pressures, including climate change and land use alterations, are key contributors to wild population declines (Hughes *et al.*, 1997; Ceballos & Ehrlich, 2002; Ceballos *et al.*, 2017), complex demographic histories, founder effects, and genetic drift may also drive changes in small and fragmented populations (Caughley, 1994; Gómez-Sánchez *et al.*, 2018). These processes, along with possible inbreeding effects (i.e., increased homozygosity), can ultimately reduce genetic diversity, adversely affect fitness and adaptive potential, and thus impact the survival of populations (Reed & Frankham, 2003; Frankham, 2005; Willi *et al.*, 2006; Charlesworth & Willis, 2009; Bozzuto *et al.*, 2019).

Indicators of 'genomic erosion' – the reduction of genetic diversity and genetic health within a declining population – include metrics such as genetic diversity, levels of inbreeding, and genetic load (Bosse & van Loon, 2022). The increased feasibility of generating genome-scale data for non-model organisms now allows for more precise assessments of these indicators in wild populations compared to traditional marker-based estimates. The transition to whole-genome approaches also provides greater resolution, particularly in evaluating inbreeding dynamics, allowing for a detailed exploration of patterns of runs of homozygosity (ROH) throughout the genome (Curik *et al.*, 2014). These continuous stretches of homozygous nucleotides in an individual's genome arise when an offspring inherits identical chromosomal segments from both parents because of a shared common ancestor (Broman & Weber, 1999; Gibson *et al.*, 2006). The frequency of ROH in offspring increases with a higher degree of shared parental ancestry. Thus, in populations with small effective population sizes, or after instances of consanguineous mating, individuals are more likely to be related by ancestry, thereby increasing the occurrence of ROH. Consequently, the proportion of the autosomal genome in ROH (F_{ROH}) reflects the extent of inbreeding within the population (McQuillan *et al.*, 2008). Over time, recombination acts to break up these ROH, such that the length distribution of ROH in genomes can be used to determine the timing of inbreeding (e.g., Kirin *et al.*, 2010; Kardos *et al.*, 2017; Foote *et al.*, 2021).

Relatively more numerous and longer ROH indicate interbreeding between more recent common ancestors with limited time for recombination to act, signifying recent inbreeding. Conversely, relatively numerous shorter ROH, derived from more distant ancestors, indicate a historical bottleneck.

Whole-genome studies focusing on natural contemporary populations offer valuable insights into past evolutionary events, yet they only provide a snapshot of the current genomic landscape, limiting the ability to infer past processes (DeFaveri & Merilä, 2015; Ryman *et al.*, 2019; Nadachowska-Brzyska *et al.*, 2021, 2022). A more accurate assessment, including better resolution of the historical processes shaping evolutionary shifts, requires establishing a baseline of the original genome-wide parameters with a temporal comparison (Díez-del-Molino *et al.*, 2018; Jensen *et al.*, 2022). While relatively long-term studies have been feasible for some wild populations (e.g., Stoffel *et al.*, 2021; Pemberton *et al.*, 2022), museum specimens increasingly serve as valuable sources of historical genomic data due to advancements in sequencing technologies, extraction methods and bioinformatic resources (Burrell *et al.*, 2015; Billerman & Walsh, 2019; Tsai *et al.*, 2020; Irestedt *et al.*, 2022). The use of museum specimens introduces additional ancestral generations, enabling a higher resolution investigation of evolutionary timescales and encompassing significant anthropogenic or environmental changes over the last century or so, proving particularly valuable for studies of genetic change through longer time frames. The use of historical DNA, from museum specimens dating up to approximately 200 years ago (hDNA; Raxworthy & Smith, 2021), has begun to facilitate a deeper understanding of evolutionary histories, processes of genomic erosion, and population declines (e.g., Dussex *et al.*, 2019; Van Der Valk *et al.*, 2019; Sánchez-Barreiro *et al.*, 2021; Femerling *et al.*, 2023), as well as adaptation and selection (Mikheyev *et al.*, 2015; Parejo *et al.*, 2020) across various species.

So far, the study of temporal genomic change in wild populations has primarily focused on threatened populations (Sánchez-Barreiro *et al.*, 2021; Jackson *et al.*, 2022), extinct versus extant species (e.g., Dussex *et al.*, 2019; Irestedt *et al.*, 2019), or socio-economically relevant species (e.g., Saha *et al.*, 2024). However, the relevance of undertaking such studies extends to species in oceanic island populations. Island endemics often exist in small populations and typically experience a population bottleneck during their initial founding event; they

therefore typically show reduced genetic diversity compared to mainland counterparts (Frankham, 1997; Leroy *et al.*, 2021). The recurrent nature of such systems across island archipelagos introduces the possibility of multiple founding events and sequential bottlenecks, exacerbating the decline in genetic diversity compared to the original source population (Clegg *et al.*, 2002; Lambert *et al.*, 2005; Pruett & Winker, 2005). Reduced effective population sizes and prolonged isolation – common characteristics of island populations – will also intensify genetic drift and inbreeding (Frankham, 1997; Clegg *et al.*, 2002), driving genomic erosion. Even after demographic recovery from a population bottleneck, the persisting effects of historical genetic drift may cause a continued decline in genetic diversity for many generations, known as drift debt (Gilroy *et al.*, 2017; Jackson *et al.*, 2022; Pinto *et al.*, 2023). All these processes outlined above, along with inter-island variation in ecological pressures, may drive genomic divergence between populations over space, and within populations over time (Nosil & Feder, 2012; Sendell-Price *et al.*, 2020). Therefore, investigating temporal genomic changes among multiple oceanic island populations of the same species will contribute towards a better understanding of genomic processes in the face of environmental challenges and population declines. Nevertheless, studies employing whole-genome data across multiple populations and time points in such wild populations are rare, if not absent.

Berthelot's pipit (*Anthus berthelotii*) provides an excellent system with which to better understand the interplay between demographic, environmental, and genomic dynamics within relatively small, isolated island populations. This small passerine bird is endemic to three archipelagos in the Macaronesian area of the north Atlantic but is currently listed as least concern (BirdLife International, 2024). Its evolutionary history involves a split from its sister species in mainland Africa (*ca*. 2.5 million years ago) (Voelker, 1999; Martin *et al.*, 2023), with populations in the Madeiran and Selvagen archipelagos independently deriving from the Canary Islands *ca.* 50,000 and 8,500 years ago, respectively (Illera *et al.*, 2007; Spurgin *et al.*, 2014; Martin *et al.*, 2023). The most recently established populations exhibit the lowest genetic diversity, with no evidence of gene flow between archipelagos, indicative of a history marked by founder effects and associated bottlenecks (Spurgin *et al.*, 2014; Armstrong *et al.*, 2018; Martin *et al.*, 2021, 2023). Adding to the complexity of its evolutionary landscape, past and present populations likely differ radically in human and

pathogen-mediated impacts, for example, ranging from none to a high prevalence of identified pathogens in some populations (Spurgin *et al.*, 2012; Sheppard *et al.*, 2022, 2024). This species – found in diverse open habitats such as grasslands, scrublands, and rocky terrains, spanning sea level to elevations exceeding 2,000 m (Clarke, 2006) – has recent historical population sizes that remain largely undocumented. However, there has been a two-fold increase in the resident human population in Madeira between 1890 and 2011 and a six-fold increase in the Canary Islands over a comparable period (De Oliveira, 2013; Instituto Canario de Estadística, 2024). These demographic shifts, coupled with a massive subsequent rise in tourism, urban expansion, and agricultural intensification, have likely led to widespread changes in land use (Otto *et al.*, 2007). Furthermore, the introduction of invasive mammal species – which pose severe threats to native species across islands globally (Traveset *et al.*, 2009; Medina *et al.*, 2011; Bellard *et al.*, 2016) – are likely to have affected pipit numbers (e.g., Olivera *et al.*, 2010). Thus, Berthelot's pipits will have encountered significant alterations in their environment and population numbers over this time. Finally, proximity to anthropogenic water sources and poultry farms heightens the risk of infectious diseases in the pipit (González-Quevedo *et al.*, 2016). Thus, it is likely that there has been a significant increase in overall pathogen infection levels in this species over the last century due to increased human activity, which may in turn have impacted population numbers. Exploring the genomic history of Berthelot’s pipit populations can elucidate the evolutionary and ecological importance of these changes.

In this study, we hypothesise that Berthelot's pipit populations, which share a common demographic origin but have independent demographic trajectories, will have experienced changes in genomic diversity and structure over the past century due to an interplay of demographic factors and environmental and anthropogenic induced pressures. To quantitatively assess these changes, we generated whole-genome resequencing data for both historical (1883–1920) and contemporary (2009) samples, spanning most of the species' geographic range. Specifically, we aimed to elucidate spatio-temporal patterns in population structure and differentiation, genetic diversity and inbreeding to better understand genomic dynamics within and between island populations. By examining these dynamics across populations with varied colonisation histories, bottleneck severities, degrees of isolation, and environmental pressures, this study seeks to provide a basis for

disentangling the relative influence of historical versus recent processes on the evolution of these island populations.

4.3. Materials and Methods

4.3.1. Sample collection, DNA extraction, library preparation and sequencing of historical individuals

We conducted whole-genome resequencing on historical Berthelot's pipit samples, sourced from specimens preserved as dried skins in various museums: Natural History Museum (NHM) in Tring, UK; National Museums Liverpool, UK; American Museum of Natural History (AMNH), USA; and Naturalis Biodiversity Center in Leiden, Netherlands. The selection of samples (Table S4.1) was based on criteria such as sample locality, collection availability, and sample collection date (range 1883–1920). In total, we included 29 toepad samples from 10 discrete island populations spanning three Macaronesian archipelagos: the Canary Islands, the Madeiran archipelago, and the Selvagens (refer to Table S4.1 for population details). The selected populations collectively encompass the great majority of the Berthelot's pipit's geographic range, as well as varying in terms of bottleneck history, isolation and contemporary anthropogenic and pathogen pressures, allowing us to explore the genomic landscape in the context of different island environments and historical influences.

DNA extraction and library preparation followed the detailed protocols outlined in Irestedt *et al.* (2022), specifically optimised for avian museum toepad samples. In short, DNA was extracted with the QIAamp DNA Micro Kit (Qiagen) and the library preparation process – based on a modified version of Meyer and Kircher's protocol (2010) – included blunt-end repair, adapter ligation, adapter fill-in, and four independently indexed PCRs. USER enzyme (New England Biolabs) treatment was used during library preparation to remove uracil residues, significantly reducing deamination patterns that are commonly observed in historical DNA (Briggs *et al.*, 2010) and avoiding the need to trim read ends in downstream bioinformatic steps. The four libraries per individual were pooled in equimolar ratios before being sent to SciLifeLab Stockholm (Sweden) for paired-end sequencing (2 × 100 bp) on the

Illumina NovaSeq6000 platform. Sequencing was performed in two batches (comprising 11 and 18 samples, respectively), each using a single flow cell lane.

4.3.2. Sample collection, DNA extraction, library preparation and sequencing of contemporary individuals

Contemporary samples were collected from individuals caught across the species' geographic range between 2006 and 2009 and consisted of blood preserved in ethanol: detailed sampling protocols are outlined in previous work (Illera *et al.*, 2007; Spurgin *et al.*, 2012). Samples were selected for the present study (n = 45 in total) in equal sex ratios and from the same 10 islands as above, providing modern counterparts to the analysed historical samples (see Table S4.1). DNA extraction followed the salt extraction protocol described by Richardson *et al.* (2001). Subsequently, DNA concentration was determined using a Qubit 4.0 Fluorometer (Invitrogen) and the Qubit dsDNA HS Assay Kit (Invitrogen).

Among these contemporary samples, 11 had been prepared and sequenced for a previous whole-genome Berthelot's pipit study (Martin *et al.*, 2023). For this subset, low Input Transposase Enabled (LITE) libraries were constructed for each individual and pooled across four lanes for paired-end sequencing (2 × 150 bp) on an Illumina HiSeq4000 platform. The remaining 34 individuals, not previously sequenced, were selected for first-time inclusion in this study. Libraries were constructed using the Illumina TruSeq DNA PCR-free preparation method and were sent to SciLifeLab Stockholm (Sweden) for paired-end sequencing (2 × 150 bp) on a single Illumina NovaSeq6000 lane.

Raw sequence reads from both historical and contemporary samples were processed using the Nextflow pipeline *nf-polish* (https://github.com/MozesBlom/nf-polish) (v22.10.6, Di Tommaso *et al.*, 2017). Prior to processing, quality summaries were obtained for the raw sequence data using FastQC v0.11.8 (Andrews, 2010). The pipeline then comprised removal of PCR duplicates using Super-Deduper v1.3.3 (https://github.com/s4hts/HTStream), adapter trimming using Trimmomatic v0.39 (Bolger *et al.*, 2014), merging of overlapping paired reads

with PEAR v0.9.11 (Zhang *et al.*, 2014), quality trimming using Trimmomatic v0.39 (Bolger *et al.*, 2014), and removal of low-complexity reads using a custom Python script.

We employed the white wagtail (*Motacilla alba*) – specifically the Motacilla_alba_V1.0_pri assembly (https://www.ncbi.nlm.nih.gov/assembly/GCF_015832195.1/) – as a reference because it represents a high-quality assembly from a closely related species (divergence time approximately 12.3 Mya, Harris *et al.*, 2018). This selection of a closely related species for the reference genome was chosen due to the assembly's greater contiguity (N50 = 8.5 Mbp) in comparison to the existing relatively poor-quality reference assembly of Berthelot's pipit (N50 = 0.36 Mbp), which is comprised of short read data (Armstrong *et al.*, 2018). The *M. alba* assembly is derived from a single male and scaffolded into 32 chromosomes, including both mitochondrial and Z scaffolds.

Cleaned sequencing reads were mapped to the wagtail reference genome using the Nextflow pipeline *nf-µmap* (https://github.com/IngoMue/nf-umap) (Di Tommaso *et al.*, 2017). This included mapping of merged and paired reads with BWA-MEM2 v2.2.1 (Vasimuddin *et al.*, 2019), post-mapping processing and merging of BAM files by individual using SAMtools v1.13 (Li *et al.*, 2009), and an assessment of DNA damage patterns using DamageProfiler v1.1 (Neukamm *et al.*, 2021).

To further assess data quality, we examined summary statistics, such as the average coverage per sample, using the Qualimap summary report (v2.2.2, Okonechnikov *et al.*, 2016) outputted from the mapping pipeline (Table S4.2). We then applied filtering to the BAM files, creating subsets for autosomal, mitochondrial, and sex-related scaffolds and removing reads with low mapping quality (-q 30). The average coverage for each of these subsets was derived from the total number of bases mapped, which was generated from SAMtools *stats* (Li *et al.*, 2009). Given the avian heterogametic sex determination system, whereby females possess ZW chromosomes, we calculated Z-autosomal coverage ratio and assigned sex to each sample based on the expectation that females should exhibit approximately half the coverage compared to males. Unplaced scaffolds were excluded from downstream analyses because these regions may represent poorly assembled or repetitive segments. Additionally, we omitted the Z-scaffolds and mitochondrial scaffolds due to their

distinct inheritance patterns, evolutionary processes and genomic dynamics compared to autosomal DNA. The remaining autosomal reference genome comprised 30 scaffolds, covering 89% of the genome (950.54 Gbp).

4.3.3. Genotype likelihoods and SNP calling

Given the variable coverage observed across historical samples, we avoided directly calling genotypes. Instead, downstream analyses were conducted within a probabilistic framework based on genotype likelihoods (GLs), using ANGSD v0.934 (Korneliussen *et al.*, 2014). This approach incorporates the inherent uncertainty about true genotypes in low–medium coverage historical data. GLs for all methods were computed using the GATK model (-GL 2) (McKenna *et al.*, 2010). Quality filters included the removal of reads mapping to multiple positions (-uniqueOnly 1) and minimum mapping and base quality scores of 30 (-minMapQ 30 -minQ 30). Similarly, depth thresholds were used to ensure adequate coverage for accurate genotype calling while avoiding regions prone to ambiguous or incorrect read mapping, such as repetitive regions, often indicated by exceptionally high coverage. Individuals were excluded from the analysis at a given site if their per-sample depth was below 5 or exceeded 50 (-setMinDepthInd 5 -setMaxDepthInd 50), allowing for approximately 10% missing data for each site (-minInd 67). Initial estimations involved GLs for sites on the autosomes, including non-variable sites. For SNP calling, major and minor alleles were inferred and a *p*-value threshold of 1 × 10−6 was employed (-do MinorMajor 1 -doMaf 1 -SNP_pval 1e–6). Sites with a significant *p*-value for being triallelic were excluded (-skipTriallelic 1), and only sites with minor allele frequency above 5% were retained (-minMaf 0.05).

Technical differences between historical and contemporary data, and the potential impact these may have in terms of systematic biases in population genetic diversity estimates and inferences of population structure, were examined extensively. This examination focused on three sources of potential bias: coverage, deamination, and read lengths, and as such, various sets of GLs were created for conducting subsequent analyses. Initially, the impact of varying coverage levels was explored by down-sampling genomes to different extents. Specifically, samples with average autosomal coverage exceeding 10X (after filtering for

mapping quality) were down-sampled to approximately 10X average coverage per sample, matching the average full coverage for historical samples. Additionally, all samples were also down-sampled to approximately 3X average coverage (the lowest of any sample) to create a dataset based on the same coverage across all individuals, and an intermediate level was implemented where samples with average autosomal coverage exceeding 5X (after filtering for mapping quality) were down-sampled to 5X. For all three versions, SAMtools *view* was used with a uniquely set -s filter for each sample to remove a proportion of reads and achieve the target coverage per sample. Subsequently, all analyses were rerun using GLs derived from the down-sampled genomes, with minor adjustments for each dataset: for the 5X dataset the minimum depth per site for individuals was set to three, and for the 3X dataset this filter was set to one.

To address potential biases stemming from post-mortem damage in historical samples, additional quality control measures were applied. In addition to the USER enzyme treatment during library preparation and assessing the frequency of C-to-T substitutions at read ends, GLs were generated as previously described while excluding transition sites using the ANGSD filter -rmTrans 1. Furthermore, historical and contemporary data differ in read length and type, with historical reads effectively becoming short single-end reads after merging read pairs. This discrepancy may result in varying levels of reference bias and alignment error, as such reads are less likely to confidently map, especially if they contain non-reference alleles. To assess this, we calculated the proportion of reads with lower mapping quality scores – a measure previously proposed as a proxy for challenging-to-map sites and successfully used to correct batch effects in Atlantic cod (*Gadus morhua*) (Lou & Therkildsen, 2022) – for each SNP. This calculation was performed after generating total sequencing depth (across all individuals) counts per site using ANGSD (-dumpCounts 1). Quality filters were applied as before, with mapping quality set to 0 and then 30 (without depth filters). Sites were removed from the initial sets of GLs where more than 10% of reads had mapping quality scores lower than 30. In total, 4% of the variable GL set (29,759 sites) and 0.7% of the full GL set (3,781,575 sites) was removed.

The majority of analyses presented utilised the higher equal coverage dataset (10X down-sampled, with historical samples below this threshold excluded) to ensure comparability

between historical and contemporary samples without compromising accuracy. The other two adjustments made to address potential systematic bias (i.e., removing transitions and lower mapping quality sites) had minimal impact on the results.

4.3.4. Population structure and differentiation

Population structure was assessed through principal component analysis (PCA) using filtered GLs of SNPs. Given that correlated SNPs can violate the assumption of independence in many population genetics analyses, we assessed pairwise linkage disequilibrium (LD) and pruned the SNP dataset accordingly. To do this, LD was estimated from GLs prior to PCA using ngsLD v1.1.0 (Fox *et al.*, 2019). LD estimation was constrained to pairs of SNPs on the same chromosome within a maximum distance of 500 kbp (--max_kb_dist 500), chosen to ensure that LD had sufficiently decayed to background levels. The rate of LD decay was visualised by randomly subsampling approximately 1% of the squared correlation coefficients (r^2) for all pairwise SNP combinations from each chromosome (83,231,565 in total) and was plotted using the 'fit_LDdecay.R' script provided in ngsLD with exhaustive fitting (--fit_level 100). Based on this analysis and rate of decay of LD observed, GLs were pruned using the script 'prune_ngsLD.py' provided in ngsLD, with a threshold distance of 20 kbp and an r^2 value of 0.1.

Two PCAs were then performed on the remaining SNPs in the combined dataset of historical and contemporary samples using PCAngsd v1.10 (Meisner & Albrechtsen, 2018). This process generated covariance matrices – one for the total dataset and another focusing on the largest archipelago, the Canary Islands. Subsequent analysis steps, such as obtaining the eigenvalues and eigenvectors, and visualisation, were carried out using R v4.3.1 (R Core Team, 2023).

Population differentiation was quantified with pairwise autosomal F_{ST} estimates both between island populations (spatially) and between each time window within each island population (temporally). Initially, unfolded site frequency likelihoods specific to each 'population' (categorised into historical and contemporary samples for each island) were estimated for the LD-pruned SNP dataset using ANGSD (-doSAF 1 -sites) with the same

quality and depth filters as when computing GLs. Then, maximum likelihood estimates of the folded 2D site frequency spectrum (SFS) for all pairs of populations were obtained using realSFS within ANGSD (- fold 1). The folding was performed due to uncertainties arising from the use of the white wagtail genome for polarising ancestral/derived alleles. Finally, the realSFS *fst index* and *stats* functions were used to retrieve values of the weighted F_{ST} estimate – an estimate more suitable for smaller sample sizes and population-specific SNPs (Bhatia *et al.*, 2013). To maximise sample size for this analysis, all individuals were included (*n* = 74), but only sites that were covered by a minimum of five reads in every individual (i.e., no missing data) were used.

4.3.5. Individual heterozygosity

To estimate autosomal heterozygosity, ANGSD was initially used to estimate unfolded site allele frequency likelihoods (-doSaf 1) for each sample. The input included a list of sites (-sites) that passed the initial quality and depth filtering during the estimation of GLs for all sites, and we applied the same read quality filters as when computing those GLs. A folded SFS was then generated individually for each sample using realSFS within ANGSD. Heterozygosity was calculated by dividing the count of heterozygous sites by the sample-specific total number of sites.

To test for significant differences between each time window across all populations and within each island population (i.e., temporal changes), non-parametric one-sided Wilcoxon rank sum tests were employed using the wilcox.test function within R v4.3.1 (R Core Team, 2023). Significant differences in heterozygosity among contemporary and historical populations (i.e., spatial differences) were initially assessed using Kruskal-Wallis tests across all populations. These were followed up with non-parametric pairwise Wilcoxon rank sum tests, applying a Benjamini–Hochberg correction (Haynes 2013), to evaluate differences between each pair of populations.

To quantify the magnitude of change in heterozygosity within each population across the two time periods, we calculated the difference between the mean contemporary

heterozygosity value and the mean historical value and then scaled this difference to express it as a proportion relative to the historical value.

4.4. Results

4.4.1. Sequencing and mapping metrics for historical versus contemporary samples

Raw read counts for historical samples averaged 382.23 million per sample (± 159.55 SD), while contemporary samples showed an average of 159.55 million reads (± 29.86 SD). Following data cleaning and mapping processes, the per-sample average dropped to 94.31 million (± 44.87 SD) for historical samples and 117.22 million (± 21.04 SD) for contemporary samples. This reduction primarily stems from merging overlapping read pairs (accounting for 85% and 20% of mapped reads on average in historical and contemporary samples, respectively), driven by the prevalence of short DNA fragments in historical samples. However, the average percentage of cleaned reads that could be mapped to the reference genome was slightly lower among historical samples (85.6% ± 5.8 SD) than contemporary samples (95.4% ± 0.3 SD), likely due to slightly lower endogenous read counts in historical samples. After trimming and merging, the mean fragment length per individual was 124 bp and 159 bp for historical and contemporary samples, respectively (using 100 bp paired-end sequencing for historical and 150 bp paired-end sequencing for contemporary). Further details can be found in Table S4.2.

There was only a marginal increase in the frequency of C-to-T and G-to-A misincorporations near the ends of historical reads (Figure S4.1A–D), indicating that the USER enzyme treatment successfully removed this characteristic signal of DNA degradation. The final coverage across autosomal scaffolds ranged from 3X to 25X (Table S4.2), with higher mean coverage in contemporary samples (16.6X ± 3.2 SD) compared to historical samples (9.7X ± 4.9 SD) (Table S4.3). Mitochondrial coverage displayed an opposing pattern, with historical samples showing substantially higher mean coverage (234.8X) compared to contemporary samples (22.5X). This observation likely reflects copy number differences and greater resistance to degradation in mitochondrial DNA compared to nuclear DNA. Z-autosome coverage ratios revealed robust sex determination consistency across contemporary

samples, aligning with previous PCR-based sex assignments (Spurgin *et al.*, 2012). In historical samples, sex determination mostly concurred with morphological assessments conducted by museums, except for three specimens which appear to have been misassigned historically (Figure S4.2). Of the total dataset, comprising both historical and contemporary samples, 36 individuals were identified as male and 38 as female.

Following down-sampling to 10X, and additional quality and depth filtering, the final GL dataset comprised a total of 398,630,637 autosomal sites. Within the variable dataset used for PCA and F_{ST} estimation after LD pruning (Figure S4.3), 619,854 biallelic sites remained.

4.4.2. Within-archipelago differentiation between historical and contemporary individuals

Using GLs for PCA, principal components one and two delineated both historical and contemporary Berthelot's pipits into three main distinct clusters, corresponding to the archipelagic divisions of the Canary Islands, Madeira, and the Selvagens (Figure 4.1A). Some separation between historical and contemporary samples from the Selvagens was observed, with the historical samples slightly closer to the other archipelagos than contemporary samples. However, discerning substructure within other archipelagos was not possible at this level.

To further untangle within-archipelago level structure, a separate PCA was conducted for the Canary Islands, the largest archipelago. This revealed a continuous cline in the genetic variation of both historical and contemporary individuals, corresponding to an east-west geographic gradient, with individuals from smaller populations at the most extreme ends of the Canary Island geographic range (i.e., El Hierro, La Gomera and La Graciosa) most distinct (Figure 4.1B). Notably, this PCA also showed separation between historical and contemporary individuals across the axes. Consistent results were obtained across GL methods, ruling out differences in coverage (Figures S4.4A–F), post-mortem DNA damage (Figures S4.5A&B), and potential differences in reference bias/mis-mapping (Figure S4.6A&B).

Similar patterns were evident in pairwise F_{ST} values, indicating limited within-archipelago differentiation across both historical and contemporary populations (Figure 4.2). Small but meaningful differentiation was observed between the most geographically distant populations within the Canary Islands, El Hierro and La Graciosa (contemporary F_{ST} = 0.025), whereas island populations in closer proximity, such as those in the central Canary Islands, were even less differentiated (e.g., Tenerife versus Gran Canaria contemporary F_{ST} = 0.009). However, substantial between-archipelago differentiation was evident. Populations linked by a single between-archipelago founding event (e.g., Madeira versus Canary Islands, or Selvagens versus Canary Islands) exhibited moderate F_{ST} values (contemporary F_{ST} = 0.089–0.100), while those linked by two separate between-archipelago founding events, the Madeiran archipelago and Selvagens, demonstrated markedly greater differentiation (F_{ST} =0.182–0.183).

At each time point, levels of differentiation observed within the same archipelago were comparable, as evidenced by among contemporary populations comparisons (Canary Islands, F_{ST} = 0.003–0.025; Madeiran archipelago, F_{ST} = 0.005) versus among historical populations comparisons (Canary Islands, F_{ST} = 0.004–0.021; Madeiran archipelago, F_{ST} = -0.009) (Figure 4.2). However, there were indications of a marginal increase in between-archipelago differentiation over time, particularly in comparisons involving the Selvagens archipelago (versus Canary Islands, historical F_{ST} = 0.060–0.078, contemporary F_{ST} = 0.089–0.095; versus Madeira, historical F_{ST} = 0.085–0.098, contemporary F_{ST} = 0.091–0.100). Temporal comparisons within the same population exhibited similar F_{ST} values to within-archipelago spatial comparisons, suggesting some divergence, with the Selvagens showing the greatest divergence over time (F_{ST} = 0.030).

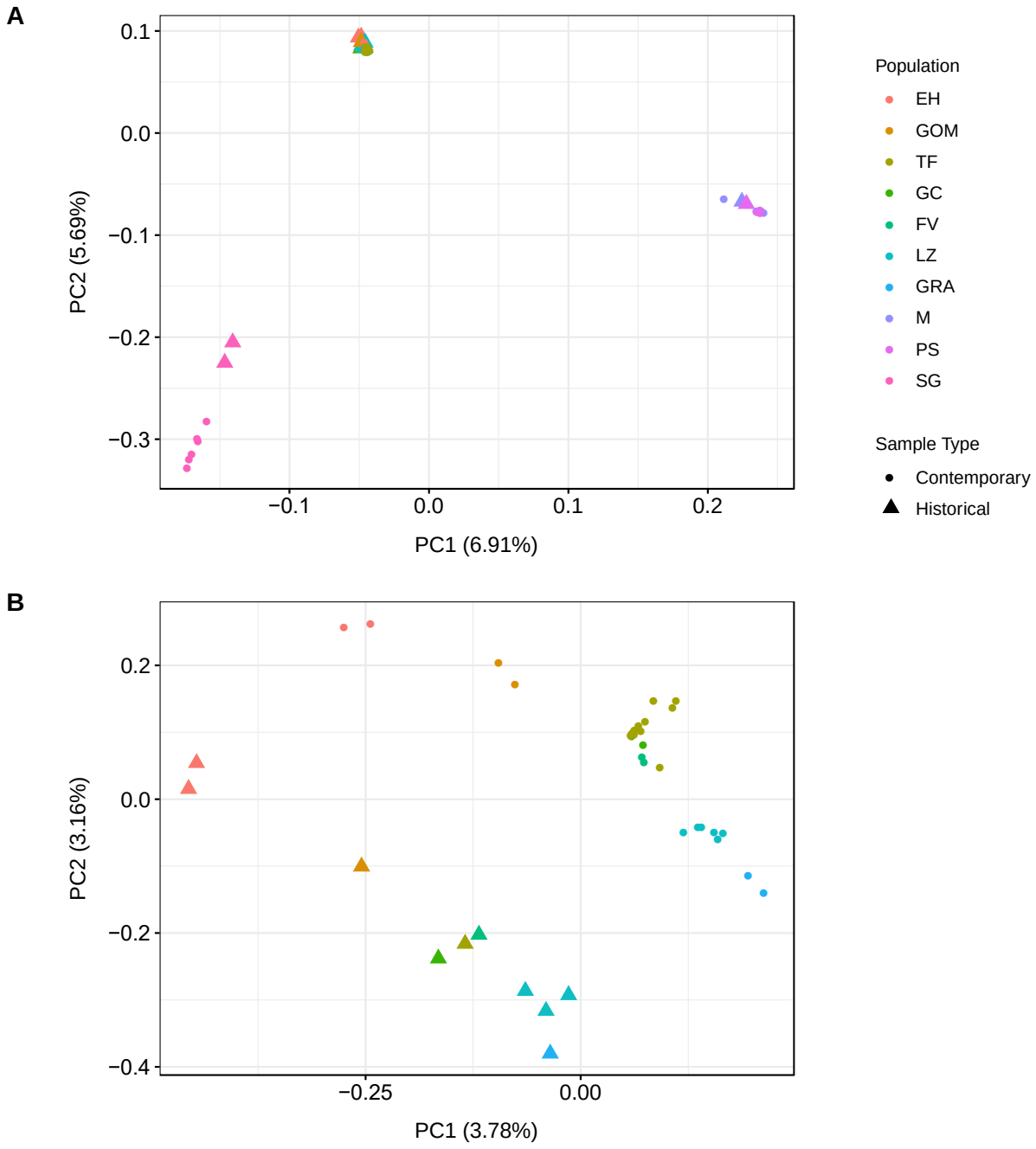

Figure 4.1. Principal component analyses of genomic variation in both historical and contemporary Berthelot's pipit samples, each with an average coverage of 10X following down-sampling, (**A**) among all populations, and (**B**) among populations within the largest archipelago only, the Canary Islands. Lower coverage historical individuals (< 10X average coverage) were excluded. The analysis is based on genotype likelihoods derived from 619,854 LD-pruned SNPs on the autosomes. The first two principal components are shown, each indicating the proportion of the observed variance explained (in brackets). Samples are colour-coded based on their respective island populations and shaped according to sample type (historical versus contemporary), see key in figure. Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

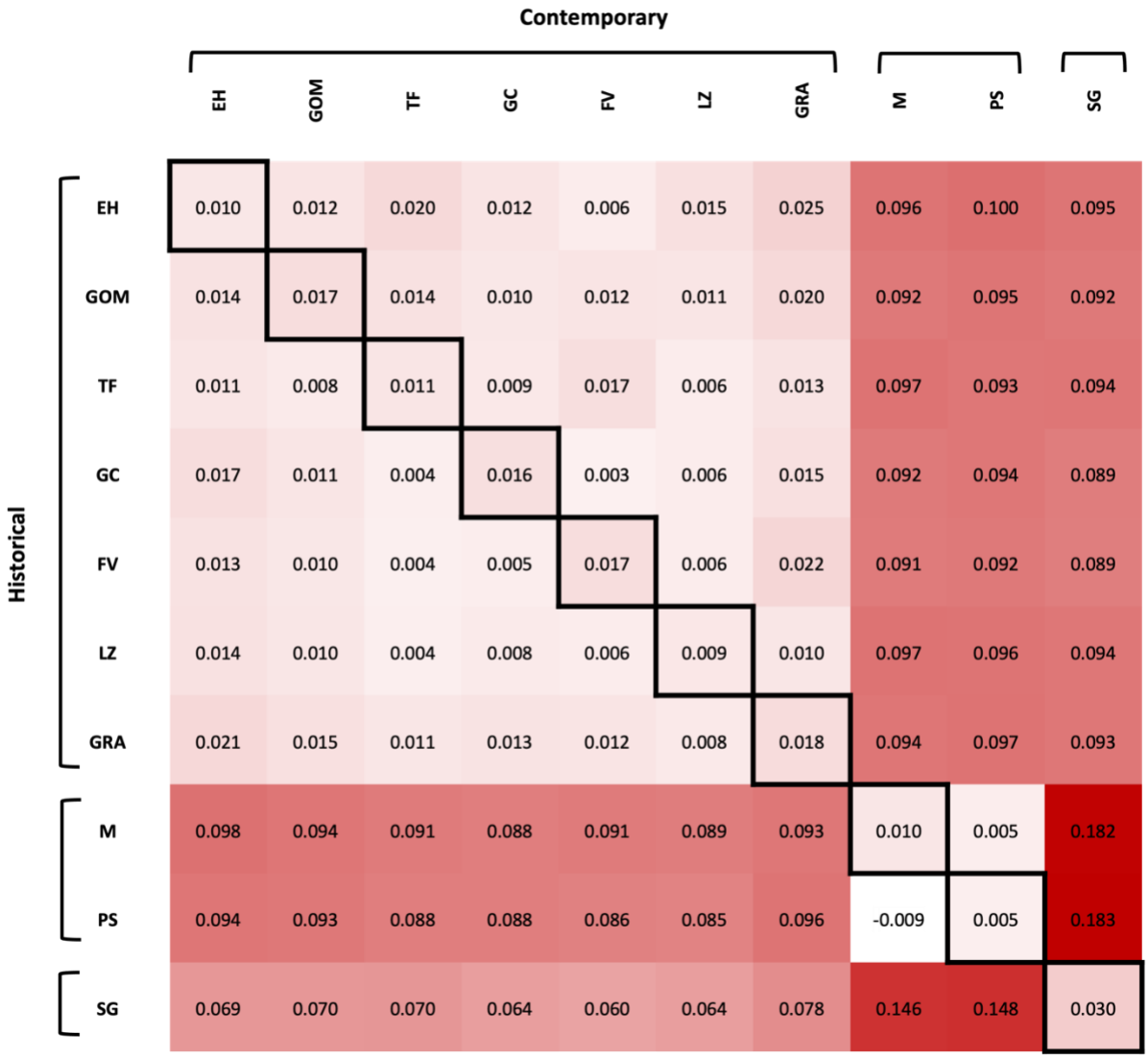

Figure 4.2. Temporal and spatial autosomal pairwise F_{ST} estimates among Berthelot's pipit populations, estimated with realSFS. The highlighted diagonal line represents temporal comparisons where F_{ST} values were obtained by comparing historical samples to contemporary samples from the same population. Above the diagonal are spatial comparisons among contemporary populations, while below the diagonal are spatial comparisons among historical populations. Estimates are derived from 79,176 LD-pruned sites present in every individual (i.e., with no 'missing' data). The analysis used the 10X down-sampled dataset, incorporating lower coverage historical individuals (with all sites covered by a minimum of five reads). Brackets delineate archipelago divisions, with colour scaled according to F_{ST} value. Canary Island Populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

4.4.3. Range-wide decline in autosomal heterozygosity over time in Berthelot's pipit

All populations of Berthelot's pipit experienced a decline in autosomal heterozygosity from the historical to the contemporary samples (one-sided Wilcoxon rank sum test, W = 553, n = 59, $p < 0.001$) (Figure 4.3). Lower coverage tended to inflate heterozygosity estimates (Figure S4.7), therefore, to ensure comparability across individuals, a subset of the dataset with an average coverage of 10X was retained after down-sampling. However, regardless of the dataset used – encompassing various coverages, removal of transitions, and exclusion of sites potentially affected by reference bias/mis-mapping – a decline in heterozygosity was consistently observed across all populations, spanning from historical to contemporary times (Figures S4.8A–E).

Heterozygosity differed significantly among contemporary populations (Kruskal-Wallis test, $X^2 = 34.735$, $n = 45$, $df = 9$, $p < 0.001$), but not among historical populations (Kruskal-Wallis test, $X^2 = 10.162$, $n = 14$, $df = 9$, $p = 0.34$). However, only one sample was used for most historical populations due to the stringent coverage threshold applied, thus leading to very low power. Where $n > 2$, pairwise Wilcoxon tests show that significant spatial differences among contemporary populations primarily stem from differences between archipelagos (Table S4.4).

Both historical and contemporary heterozygosity in each population was relatively consistent (within a time period) across the Canary Islands (Figure 4.3), despite variation in island sizes/population sizes, and land use pressures. Hence, this uniform pattern of heterozygosity across populations in the Canary Islands appears to have persisted over the past century and similar relative amounts of difference were observed (Table 4.1).

The most recently established and bottlenecked populations – Porto Santo, Madeira and Selvagem Grande – exhibited lower heterozygosity levels, whereby even historical heterozygosity estimates were lower than contemporary heterozygosity estimates in the Canary Islands, except for Porto Santo (Figure 4.3). The contemporary population of Selvagem Grande, which occupies the smallest island, had the lowest heterozygosity overall and also experienced the greatest reduction in heterozygosity from historical to

contemporary samples (Table 4.1) (with the exception of Porto Santo for which there is just one, extreme outlier sample comprising the historical data).

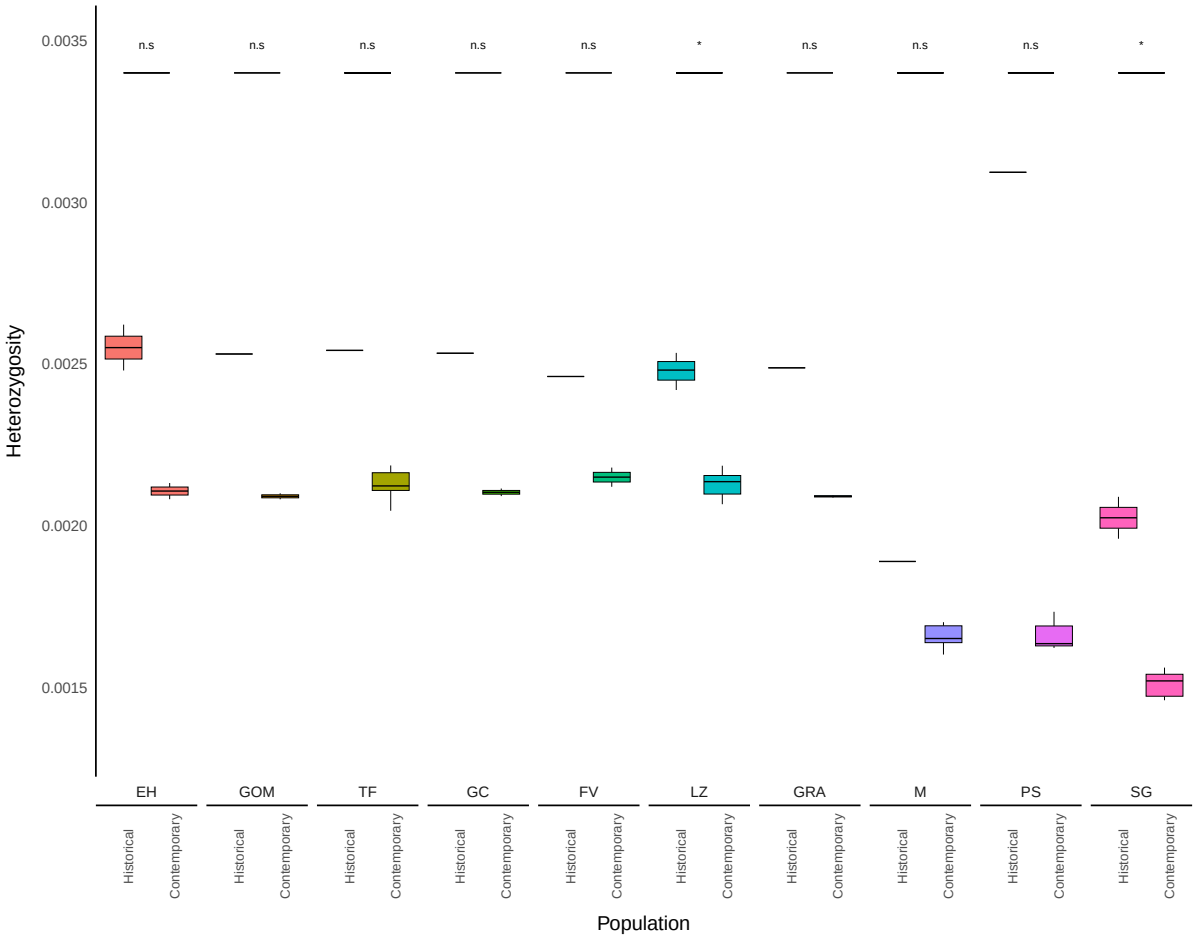

Figure 4.3. Spatio-temporal comparisons of individual autosomal heterozygosity in historical and contemporary Berthelot's pipit samples, each with an average coverage of 10X following down-sampling. Estimates are grouped according to island population and time window. Pairwise significance between time windows (historical versus contemporary for the same population) was assessed using one-sided Wilcoxon rank sum tests. Asterisks denote *p*-values < 0.05. Lower coverage historical individuals (< 10X average coverage) were excluded. Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande; Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa.

Table 4.1. Relative change in heterozygosity between historical and contemporary populations (2009) of Berthelot's pipits. Estimates are based on samples with an average coverage of 10X following down-sampling. Lower coverage historical individuals (< 10X average coverage) were excluded. Pairwise significance between time windows (historical versus contemporary) was assessed using one-sided Wilcoxon rank sum tests. Asterisks denote *p*-values < 0.05, though note differences in sample sizes.

Archipelago	Population	Collection year of oldest sample	*n* historical	*n* contemporary	Scaled Δ heterozygosity
Canary Islands	El Hierro	1890	2	2	-0.174
	La Gomera	1920	1	2	-0.174
	Tenerife	1888	1	12	-0.163
	Gran Canaria	1888	1	2	-0.170
	Fuerteventura	1910	1	2	-0.126
	Lanzarote	1913	3	6	-0.141*
	La Graciosa	1913	1	2	-0.160
Madeiran archipelago	Madeira	1900	1	6	-0.116
	Porto Santo	1894	1	5	-0.463
Selvagens	Selvagem Grande	1895	2	6	-0.264*

4.5. Discussion

Through a comparative examination of contemporary and historical whole-genome sequencing data spanning the 20th century, we have identified and quantified significant genomic shifts in Berthelot's pipit populations amidst a time of rapid environmental changes. The examination of data from multiple island populations allowed us to explore both spatial and temporal patterns of genomic diversity and divergence, offering valuable insights into the evolutionary dynamics within this system. Our results revealed a widespread decline in genetic diversity since the 1880s, likely indicative of population size reductions resulting in increased genetic drift. Differences in genetic diversity between archipelagos primarily reflect long-term founding events, yet our findings raise the possibility that additional, possibly recent pressures – such as anthropogenic influences – may have contributed to recent range-wide declines in genome-wide diversity, increased divergence between archipelagos, and temporal differentiation within populations. Despite

these temporal shifts, considerable similarity in genetic diversity within the Canary Islands suggests that sufficient inter-island migration may help homogenise genetic diversity levels across this larger archipelago.

Spatially, the spatial stepwise changes in autosomal heterozygosity between archipelagos, present in both historical and contemporary genomes (Figure 4.3), likely represent significant reductions in genetic diversity following initial colonisation events. These patterns highlight the enduring isolation and unique evolutionary trajectories of separate island groups over time. The comparisons of F_{ST} values among historical and contemporary populations reveal a similar pattern of greater differentiation between archipelagos but low differentiation within archipelagos (Figure 4.2). This within-archipelago pattern is further supported by a PCA, which reveals a gradual cline of genetic variation across the Canary Islands, rather than a clustered structure (Figure 4.1B). These findings imply a lack of complete population isolation, suggesting sustained gene flow among populations within archipelagos but not between them. These contemporary patterns align with the previous understandings of population structure from other, more limited, molecular markers, including reduced representation restriction-site associated DNA (RAD)-sequencing (Armstrong *et al.*, 2018; Martin *et al.*, 2021), microsatellites and mitochondrial DNA (Illera *et al.*, 2007; Spurgin *et al.*, 2014), as well as a very limited subset of whole genomes (Martin *et al.*, 2023).

In the current study, the genetic variation observed in Berthelot's pipit populations offers valuable insights that may have been overlooked without the inclusion of data obtained from museum specimens. While spatial patterns suggest that low variation in these populations is likely a consequence of historical colonisation processes, temporal contrasts reveal that autosomal heterozygosity was significantly higher approximately a century ago (Figure 4.3). Across all populations, we observed a loss of heterozygosity ranging from 12% to 46% of the historical variation over a span of 89 to 121 years (Table 4.1). This corresponds to a period of 24–33 or 43–59 generations, depending on estimates of generation length (Garcia-Del-Rey & Cresswell, 2007; Bird *et al.*, 2020). Even populations with initially limited genetic variation, such as Selvagem Grande and Madeira, experienced marked declines (Figure 4.3 and Table 4.1). Additionally, while we observed a weak increase in genetic

structure between archipelagos over time, genetic differentiation within archipelagos remained relatively constant. Temporal F_{ST} comparisons reveal patterns of spatial divergence among islands, with increased differentiation between Selvagens and other archipelagos. Notably, contemporary diversity levels were lower than those observed in some bird populations with recent bottlenecks (Wang *et al.*, 2022), and are comparable to estimates from other bottlenecked island endemic birds (Campana *et al.*, 2020). Overall, our contemporary heterozygosity estimates for Berthelot's pipit (0.0015–0.0021) fall within the lower range for 'least concern' bird species (0.00250 ± 0.00107) and are similar to the higher range for 'threatened' bird species (0.00118 ± 0.00085) (Li *et al.*, 2014).

Although direct comparisons across species must consider differences in life histories and ecological contexts, these reference points underscore the diversity loss observed in Berthelot's pipit. While populations naturally fluctuate, the consistent decline observed in our study suggests that such fluctuations alone are unlikely to account for the notable reduction in genetic diversity in Berthelot's pipits. This decline also aligns with previous temporal investigations globally, which have observed substantial losses in genetic diversity (allele diversity and expected heterozygosity) regardless of IUCN status, both in islands and continental mainland populations (Leigh *et al.*, 2019 and references therein). However, island species have been particularly impacted, experiencing a significantly higher loss of variation compared to mainland species, averaging a 27% decline in intraspecific genetic diversity (Leigh *et al.*, 2019). While these previous studies mainly used microsatellite markers, our whole-genome analysis reveals a comparable average loss of 20% in Berthelot's pipits. This proportional loss also mirrors some findings from temporal whole-genome studies in threatened species facing declines (Van Der Valk *et al.*, 2019; Sánchez-Barreiro *et al.*, 2021), although the remaining levels of variation observed in Berthelot's pipits are notably higher. These trends may thus reflect longer-term or directional pressures, rather than representing a snapshot of demographic variability alone. Our study therefore highlights the vulnerability of seemingly least concern island endemics to genomic erosion.

Attributing the observed loss of genetic diversity solely to a single cause is challenging, especially given the limited number of time points in our study. Genetic variation is unlikely

to decline evenly over time, and the initial time point may not capture variation before the causal factor started but rather reflect a time point in an already ongoing decline. Given these considerations, we propose three possible scenarios for the decline in genetic diversity, each with distinct implications for understanding the long-term persistence of Berthelot's pipit populations. First, recent environmental changes could be the primary driver, where rapid anthropogenic changes – including habitat alteration, land use shifts, and/or invasive species – have reduced population sizes, thus intensifying inbreeding and genetic drift, leading to the observed loss of heterozygosity. This is highly plausible given the drastic changes since the mid-twentieth century, when mass tourism, urban expansion and modern agriculture began reshaping the Canary Islands and Madeira, particularly the semi-arid coastal landscapes on which Berthelot's pipit rely (Otto *et al.*, 2007; Rodrigues, 2016). The spread of invasive species, like feral cats (*Felis silvestris catus),* have likely exacerbated this effect, as the ground-nesting, foraging habits of the pipit (Garcia-Del-Ray & Cresswell, 2007) make it especially vulnerable to predation (Medina & Nogales, 2009). For example, on Selvagem Grande, the eradication of invasive mammals in 2002 led to a rapid increase in pipit numbers (Olivera *et al.*, 2010), underscoring the potential for recent pressures to have driven population declines and heterozygosity loss.

Alternatively, a second scenario considers the possibility of a historical decline or environmental disturbance that predates our historical samples, which may have already lowered baseline genetic diversity by the time museum specimens were collected. For example, early North African settlers (*ca*. 2,000–2,500 years ago) introduced grazing animals, practiced small-scale crop cultivation, engaged in forest clearance, and altered fire regimes, all of which likely affected island flora and fauna, leading to species extinctions (e.g., Rando, 2002; Bocherens *et al*., 2006; de Nascimento *et al*., 2009) and ecosystem shifts (de Nascimento *et al*., 2016, 2020; Castilla-Beltrán *et al*., 2021). By the fifteenth century, European settlers further intensified landscape transformation, introducing intensive agriculture, widespread deforestation, urban development, and additional invasive species, which likely accelerated habitat fragmentation and ecosystem change (de Nascimento *et al*., 2020; Norder *et al*., 2020; Castilla-Beltrán *et al*., 2021). This scenario suggests that the observed declines reflect a long-term erosion of genetic diversity, initiated by these past disturbances rather than by recent changes alone. Such a trajectory would imply that

ongoing genetic loss is part of a prolonged process, with current diversity levels shaped by compounding effects over time.

Lastly, we consider the possibility of a gradual, continuous decline in genetic diversity, driven by small population sizes and persistent genetic drift since the species' colonisation of these islands. While this scenario is possible, it would require consistently small effective population sizes (N_e) over a prolonged period (c*a*. 8,500 years ago to 2.5 million years ago). Such conditions could theoretically result in long-term diversity loss (Wright, 1931), but they are less likely to explain the *ca*. 20% decline observed in just the past *ca*. 120 years. This rapid rate of loss suggests that other factors, such as pre-modern disturbances or more recent anthropogenic pressures, are more likely to have played a significant role in the observed genetic decline. Evidence of stabilising gene flow within archipelagos, as indicated by PCA clustering, low differentiation, and comparable levels of genetic variation among islands, further argues against genetic drift alone being the primary driver. While persistent drift may have contributed to diversity loss to some extent, especially in smaller, isolated populations, the magnitude and timing of the declines align more closely with more recent disturbances, though it is difficult to say for certain without further analysis. Testing the plausibility of these scenarios would require additional analyses, such as genomic simulations (e.g., Haller & Messer, 2019) or demographic reconstructions targeting the past 100 generations (see Nadachowska-Brzyska *et al*., 2022). Unlike deeper coalescent signals, which converge to shared ancestral lineages and are less impacted by small sample sizes, methods designed to detect recent N_e changes – such as GONE analysis (Santiago *et al*., 2020) – require larger sample sizes to more accurately capture weaker signals and subtle allele frequency shifts (Beichman, *et al*. 2018; Santiago *et al*. 2020). Thus, the current dataset, with 2–12 individuals per population, lacks the resolution necessary to robustly test for the possibility of recent demographic events.

We conducted rigorous tests to identify and mitigate any systematic biases arising from technical differences between historical and contemporary datasets that could explain the temporal differences in our data. Historical samples present unique challenges due to factors such as fragmentation and post-mortem damage (Stiller *et al.*, 2006; Sefc *et al.*, 2007), which directly affect the accuracy of sequencing and mapping processes. Additionally, coverage in

hDNA is typically lower, influenced by factors such as lower endogenous content and DNA quality (McDonough *et al.*, 2018; Tsai *et al.*, 2020). This lower coverage increases uncertainty regarding erroneous base calls (from sequencing errors or damage), mis-mapping of reads, and successful sampling of both alleles in a diploid organism (Nielsen *et al.*, 2012; Lou *et al.*, 2021). Such challenges can impact variant calling accuracy, thereby influencing population genomic inference (Parks & Lambert, 2015). Consequently, such biases have the potential to generate artificial differences between historical and contemporary data, complicating the interpretation of genuine biological patterns. In our current study, the primary axes of variation in the PCA were associated with differences between archipelagos (Figure 4.1A), likely reflecting historical archipelago founding events and subsequent isolation. However, the separate clustering of historical and contemporary samples at the within-archipelago level (Figure 4.1B) – though likely representing genuine genetic divergence differences between the time points – could also be the result of residual lower-level technical differences between the two DNA sources not entirely eliminated by our processing.

The ANGSD GL approach that we employed in this study, though not specifically designed for hDNA, effectively handles low-coverage sequencing data and accounts for uncertainty in genotype inference (Korneliussen *et al.*, 2014), therefore it is commonly used in whole-genome analyses of hDNA (Mikheyev *et al.*, 2015; Van Der Valk *et al.*, 2019; Pinsky *et al.*, 2021; Sánchez-Barreiro *et al.*, 2021; Wu *et al.*, 2022). However, there are indications that our heterozygosity estimates may be inflated in historical samples with lower average coverage and higher levels of missing data (Figure S4.7), although we have done our best to eliminate this. Our conclusions are therefore primarily based on a 'fair' comparison, utilising matched down-sampled coverages. Though, it is notable that inferences of spatio-temporal patterns in population structure, differentiation and diversity (heterozygosity) also remained consistent regardless of the coverage dataset used (Figures S4.4 & S4.8). Removing all transition sites, and therefore any deamination patterns, also had no discernible impact on inferences (Figures S4.8B). Nonetheless, we cannot exclude the potential impact of shorter read lengths and the prevalence of 'single-end' merged reads in historical samples (Table S4.2), as these factors may increase the likelihood of misalignment, particularly when reads do not have the reference allele (Leigh *et al.*, 2018; Lou & Therkildsen, 2022). We used a

correction technique for this informed by the Atlantic cod genome (Lou & Therkildsen, 2022) which had minimal impact on our conclusions (Figure S4.6 & S4.9).

The combined effects of variable read lengths (Table S4.2) and sequence divergence between the Berthelot's pipit genome and the white wagtail genome (used as the reference), may introduce bias into heterozygosity estimates (see also Z-autosome coverage ratios, Figure S4.2) (see Parks & Lambert, 2015). Nevertheless, this bias appears to be relatively minor as we are still able to resolve appropriate signatures of population differentiation. Studies using ancient DNA, which typically involves even shorter read lengths, have highlighted the potential impact of mapping against a reference genome of a related species on heterozygosity estimates (Orlando *et al.*, 2013; Gopalakrishnan *et al.*, 2017), though this tends to underestimate the variation present in samples, contrary to what would be needed to produce a similar effect in our study. Regardless, re-analysing our data using a species-specific reference genome and validating the consistency (or inconsistency) of results would help to identify any residual biases introduced by the mapping process. Unfortunately, the current draft reference genome for the Berthelot's pipit lacks contiguity, completeness and annotation (Armstrong *et al.*, 2018) which would introduce new limitations. These limitations highlight the need for a high-quality, *de novo* assembled pipit-specific genome, which is now underway by another laboratory group unconnected with the current work. Alternative potential correction strategies include read trimming to uniform lengths or restricting analyses to 'callable' conserved regions (Leigh *et al.*, 2018). However, it is very challenging to determine if the latter approach would be effective. Moreover, trimming contemporary reads to the same length could ultimately inflate all estimates.

Despite remaining potential reasons for biases between historical and contemporary samples, we anticipate minimal impact on heterozygosity estimates, especially in high-coverage individuals. This assumption is based on several factors: first, the primary partition of variation effectively and strongly resolved population differentiation; second, the fairly small difference in average read lengths between the two sample groups (*ca*. 20%); and third, that heterozygosity estimates are derived from autosomal sites, and only a fraction of these sites are likely to be affected by any remaining biases. However, such biases could

significantly affect analyses that focus on specific SNPs/regions or continuous tracts of the genome. The absence of a species-specific reference genome has thus far prevented efforts to test for any potential remaining biases accurately; consequently, we chose not to analyse ROH at this stage. Genetic diversity is just one indicator of genomic erosion (Bosse & van Loon, 2022) and given that we have shown a range-wide decline within Berthelot's pipit populations over the previous century, future work should focus on measures directly linked to fitness in this species, such as genome-wide estimates of inbreeding and an investigation of adaptive variation over the same geographic and temporal scales. For example, analysing the proportion and length distribution of ROH in historical and contemporary genomes can shed light on the timing and extent of inbreeding within populations, as well as the causes of genetic diversity loss (e.g., Dussex *et al.*, 2019; Van Der Valk *et al.*, 2019; Sánchez-Barreiro *et al.*, 2021; Femerling *et al.*, 2023). Long ROH segments typically indicate recent inbreeding, suggesting a response to immediate demographic pressures, while shorter segments may reflect historical inbreeding events and prolonged genetic drift. This information could help distinguish between the influences of recent human impacts and longer-standing isolation in Berthelot's pipit populations. Importantly, examining the relationship between inbreeding and fitness traits would further elucidate the consequences of declining genetic diversity and the potential risks associated with inbreeding depression for the survival of Berthelot's pipit.

Future work should also involve investigating how the F_{ST} (genetic differentiation) changes across space and time at immune genes (compared to genome-wide changes) in populations with varying type and intensities of pathogen prevalence. This analysis would shed light on the evolutionary dynamics of immune-related or adaptive genes more broadly. Examining temporal patterns in variation previously associated with pathogen-mediated selection in Berthelot's pipit (González-Quevedo & Phillips, 2016; Armstrong *et al.*, 2018, 2019; Sheppard *et al.*, 2022, 2024), as well as divergent selection more generally (Armstrong *et al.*, 2018; Martin *et al.*, 2021), would provide valuable insights into the evolution of adaptive variation over time in this species.

4.5.1. Conclusions

Our study documents a widespread decline in autosomal heterozygosity and increased differentiation among archipelago populations of Berthelot's pipit over the past century, coinciding with a period of extensive environmental change. These consistent patterns of change in genetic variation offer valuable insights into possible drivers of genetic diversity loss within populations, with pre-modern or recent anthropogenic impacts as plausible factors. Our findings enhance our understanding of the spatio-temporal dynamics in relatively small and isolated island populations that are currently classified as least concern. In the future, the development of a high-quality, species-specific reference genome for Berthelot's pipit promises to further improve the accuracy of genomic analyses in this system. Such advancements would enable us to investigate additional metrics, including known adaptive variation and measures of inbreeding such as runs of homozygosity, over comparable time scales. This, in turn, could provide critical insights into the potential fitness consequences of genomic erosion in these populations, while also enhancing our ability to identify the drivers of genetic decline in Berthelot's pipit.

4.6. References

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution*, *9*, 12482–12502.

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

Beichman, A., Huerta-Sanchez, E., & Lohmueller, K. (2018). Using genomic data to infer historic population dynamics of nonmodel organisms. *Annual Review of Ecology, Evolution, and Systematics*, *49*, 433–456.

Bellard, C., Genovesi, P., & Jeschke, J. M. (2016). Global patterns in threats to vertebrates by biological invasions. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20152454.

Bhatia, G., Patterson, N., Sankararaman, S., & Price, A. L. (2013). Estimating and interpreting F_{ST}: the impact of rare variants. *Genome Research*, *23*, 1514–1521.

Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: a review of methods, challenges, applications, and future directions. *Molecular Ecology Resources*, *19*, 1115–1130.

Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., Symes, A., *et al.* (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, *34*, 1252–1261.

BirdLife International. (2024). Species factsheet: *Anthus berthelotii*. Available at: https://datazone.birdlife.org/species/factsheet/berthelots-pipit-anthus-berthelotii.

Bocherens, H., Michaux, J., García-Talavera, F., & Van der Plicht, J. (2006). Extinction of endemic vertebrates on islands: the case of the giant rat *Canariomys bravoi* (Mammalia, Rodentia) on Tenerife (Canary Islands, Spain). *Comptes Rendus Palevol*, *5*, 885–891.

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

Bosse, M., & van Loon, S. (2022). Challenges in quantifying genome erosion for conservation. *Frontiers in Genetics*, *13*, 960958.

Bozzuto, C., Biebach, I., Muff, S., Ives, A. R., & Keller, L. F. (2019). Inbreeding reduces long-term growth of Alpine ibex populations. *Nature Ecology and Evolution*, *3*, 1359–1364.

Briggs, A. W., Stenzel, U., Meyer, M., Krause, J., Kircher, M., & Pääbo, S. (2010). Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. *Nucleic Acids Research*, *38*, e87.

Broman, K., & Weber, J. (1999). Long homozygous chromosomal segments in reference families from the centre d'Etude du polymorphisme humain. *The American Journal of Human Genetics*, *65*, 1493–1500.

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44.

Campana, M., Corvelo, A., Shelton, J., Callicrate, T., Bunting, K., Riley-Gillis, B., Wos, F., *et al*. (2020). Adaptive radiation genomics of two ecologically divergent Hawai'ian honeycreepers: The 'akiapōlā'au and the Hawai'i 'amakihi. *Journal of Heredity*, *111*, 21-32.

Castilla-Beltrán, A., de Nascimento, L., Fernández-Palacios, J. M., Whittaker, R. J., Willis, K. J., Edwards, M., & Nogué, S. (2021). Anthropogenic transitions from forested to human-dominated landscapes in southern Macaronesia. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2022215118.

Caughley, G. (1994). Directions in conservation biology. *Journal of Animal Ecology*, *63*, 215–244.

Ceballos, G., & Ehrlich, P. R. (2002). Mammal population losses and the extinction crisis. *Science*, *296*, 904–907.

Ceballos, G., Ehrlich, P. R., & Dirzo, R. (2017). Biological annihilation via the ongoing sixth mass extinction signaled by vertebrate population losses and declines. *Proceedings of the National Academy of Sciences of the United States of America*, *114*, E6089–E6096.

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

Clarke, T. (2006). *A Field Guide to the Birds of the Atlantic Islands: Canary Islands, Madeira, Azores, Cape Verde*. London: Bloomsbury Publishing Plc.

Clegg, S. M., Degnan, S. M., Kikkawa, J., Moritz, C., Estoup, A., & Owens, I. P. F. (2002). Genetic consequences of sequential founder events by an island-colonizing bird. *Proceedings of the National Academy of Sciences of the United States of America*, *99*, 8127–8132.

Curik, I., Ferenčaković, M., & Sölkner, J. (2014). Inbreeding and runs of homozygosity: A possible solution to an old problem. *Livestock Science*, *166*, 26–34.

de Nascimento, L., Nogué, S., Criado, C., Ravazzi, C., Whittaker, R. J., Willis, K. J., & Fernández-Palacios, J. M. (2016). Reconstructing Holocene vegetation on the island of Gran Canaria before and after human colonization. *Holocene*, *26*, 113–125.

de Nascimento, L., Nogué, S., Naranjo-Cigala, A., Criado, C., McGlone, M., Fernández-Palacios, E., & Fernández-Palacios, J. M. (2020). Human impact and ecological changes during prehistoric settlement on the Canary Islands. *Quaternary Science Reviews*, *239*, 106332.

de Nascimento, L., Willis, K. J., Fernández-Palacios, J. M., Criado, C., & Whittaker, R. J. (2009). The long-term ecology of the lost forests of La Laguna, Tenerife (Canary Islands). *Journal of Biogeography*, *36*, 499–514.

De Oliveira, I. T. , 2013. O arquipélago da Madeira: dinâmicas demográficas de 1890 a 2011. In *Santos, C., Matos, P.T.D.E. (Eds.), A Demografia Das Sociedades Insulares Portuguesas. Séculos XV A XXI* (pp. 105–135).

DeFaveri, J., & Merilä, J. (2015). Temporal stability of genetic variability and differentiation in the three-spined stickleback (*Gasterosteus aculeatus*). *PLoS ONE*, *10*, e0123891.

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*, 316–319.

Díez-del-Molino, D., Sánchez-Barreiro, F., Barnes, I., Gilbert, M. T. P., & Dalén, L. (2018). Quantifying temporal genomic erosion in endangered species. *Trends in Ecology & Evolution*, *33*, 176–185.

Dussex, N., Von Seth, J., Knapp, M., Kardailsky, O., Robertson, B. C., & Dalén, L. (2019). Complete genomes of two extinct New Zealand passerines show responses to climate fluctuations but no evidence for genomic erosion prior to extinction. *Biology Letters*, *15*, 20190491.

Femerling, G., Oosterhout, C. van, Feng, S., Bristol, R. M., Zhang, G., Groombridge, J., Gilbert, M. T. P., *et al.* (2023). Genetic load and adaptive potential of a recovered avian species that narrowly avoided extinction. *Molecular Biology and Evolution*, *40*, msad256.

Foote, A. D., Hooper, R., Alexander, A., Baird, R. W., Charles, |, Baker, S., Ballance, L., *et al.* (2021). Runs of homozygosity in killer whale genomes provide a global record of demographic histories. *Molecular Ecology*, *30*, 6162–6177.

Fox, E. A., Wright, A. E., Fumagalli, M., & Vieira, F. G. (2019). ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics*, *35*, 3855–3856.

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

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, *126*, 131–140.

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

Gibson, J., Morton, N. E., & Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics*, *15*, 789–795.

Gilroy, D. L., Phillips, K. P., Richardson, D. S., & van Oosterhout, C. (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.

Gómez-Sánchez, D., Olalde, I., Sastre, N., Enseñat, C., Carrasco, R., Marques-Bonet, T., Lalueza-Fox, C., *et al.* (2018). On the path to extinction: inbreeding and admixture in a declining grey wolf population. *Molecular Ecology*, *27*, 3599–3612.

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.

Gopalakrishnan, S., Samaniego Castruita, J. A., Sinding, M. H. S., Kuderna, L. F. K., Räikkönen, J., Petersen, B., Sicheritz-Ponten, T., *et al.* (2017). The wolf reference genome sequence (*Canis lupus lupus*) and its implications for *Canis spp*. population genomics. *BMC Genomics*, *18*, 495.

Haller, B. C., & Messer, P. W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. *Molecular Biology and Evolution*, 3*6*, 632–637.

Harris, R. B., Alström, P., Ödeen, A., & Leaché, A. D. (2018). Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (*Motacilla*). *Molecular Phylogenetics and Evolution*, *120*, 183–195.

Haynes, W. (2013). 'Benjamini-Hochberg method', in Dubitzky, W., Wolkenhauer, O., Cho, K. H., & Yokota, H. (eds) *Encyclopedia of Systems Biology.* New York: Springer, 78.

Hughes, J. B., Daily, G. C., & Ehrlich, P. R. (1997). Population diversity: Its extent and extinction. *Science*, *278*, 689–692.

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.

Instituto Canario de Estadística. (2024). Población, hogares y tamaño medio de los hogares según censos. Municipios por islas de Canarias y años.

https://www.gobiernodecanarias.org/istac/estadisticas/demografia/poblacion/cifrascensales/C00025A.html.

Irestedt, M., Ericson, P. G. P., Johansson, U. S., Oliver, P., Joseph, L., & Blom, M. P. K. (2019). No signs of genetic erosion in a 19th century genome of the extinct Paradise parrot (*Psephotellus pulcherrimus*). *Diversity, 11,* 58.

Irestedt, M., Thörn, F., Müller, I. A., Jønsson, K. A., Ericson, P. G. P., & Blom, M. P. K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources*, *22*, 2672–2684.

Jackson, H. A., Percival-Alwyn, L., Ryan, C., Albeshr, M. F., Venturi, L., Morales, H. E., Mathers, T. C., *et al.* (2022). Genomic erosion in a demographically recovered bird species during conservation rescue. *Conservation Biology*, *36*, e13918.

Jensen, E. L., Díez-del-Molino, D., Gilbert, M. T. P., Bertola, L. D., Borges, F., Cubric-Curik, V., de Navascués, M., *et al.* (2022). Ancient and historical DNA in conservation policy. *Trends in Ecology & Evolution*, *37*, 420–429.

Kardos, M., Qvarnström, A., Ellegren, H. (2017). Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics, 205,* 1319–1334*.*

Kirin, M., McQuillan, R., Franklin, C., Campbell, H., McKeigue, P., & Wilson, J. (2010). Genomic runs of homozygosity record population history and consanguinity. *PLoS ONE*, *5*, 13996.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, 356.

Lambert, D. M., King, T., Shepherd, L. D., Livingston, A., Anderson, S., & Craig, J. L. (2005). Serial population bottlenecks and genetic variation: translocated populations of the New Zealand Saddleback (*Philesturnus carunculatus rufusater*). *Conservation Genetics*, *6*, 1–14.

Leigh, D. M., Hendry, A. P., Vázquez-Domínguez, E., & Friesen, V. L. (2019). Estimated six per cent loss of genetic variation in wild populations since the industrial revolution. *Evolutionary Applications*, *12*, 1505–1512.

Leigh, D. M., Lischer, H. E. L., Grossen, C., & Keller, L. F. (2018). Batch effects in a multiyear sequencing study: False biological trends due to changes in read lengths. *Molecular Ecology Resources*, *18*, 778–788.

Leroy, T., Rousselle, M., Tilak, M. K., Caizergues, A. E., Scornavacca, C., Recuerda, M., Fuchs, J., *et al.* (2021). Island songbirds as windows into evolution in small populations. *Current Biology*, *31*, 1303–1310.e4.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., *et al.* (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*, 2078–2079.

Li, S., Li, B., Cheng, C., Xiong, Z., Liu, Q., Lai, J., Carey, H. V., *et al.* (2014). Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. *Genome Biology*, *15*, 557.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*, 5966–5993.

Lou, R. N., & Therkildsen, N. O. (2022). Batch effects in population genomic studies with low-coverage whole genome sequencing data: causes, detection and mitigation. *Molecular Ecology Resources*, *22*, 1678–1692.

Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and within-archipelago adaptation between island bird populations. *Royal Society Open Science*, *8*, 201146.

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging island populations of a bird. *Molecular Ecology*, *32*, 1972–1989.

McDonough, M. M., Parker, L. D., McInerney, N. R., Campana, M. G., & Maldonado, J. E. (2018). Performance of commonly requested destructive museum samples for mammalian genomic studies. *Journal of Mammalogy*, *99*, 789–802.

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., *et al.* (2010). The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*, 1297–1303.

McQuillan, R., Leutenegger, A.-L., Abdel-Rahman, R., Franklin, C. S., Pericic, M., Barac-Lauc, L., Smolej-Narancic, N., *et al.* (2008). Runs of homozygosity in European populations. *The American Journal of Human Genetics*, *83*, 359–372.

Medina, F. M., Bonnaud, E., Vidal, E., Tershy, B. R., Zavaleta, E. S., Josh Donlan, C., Keitt, B. S., *et al.* (2011). A global review of the impacts of invasive cats on island endangered vertebrates. *Global Change Biology*, *17*, 3503–3510.

Medina, F. M., & Nogales, M. (2009). A review on the impacts of feral cats (*Felis silvestris catus*) in the Canary Islands: implications for the conservation of its endangered fauna. *Biodiversity and Conservation*, *18*, 829–846.

Meisner, J., & Albrechtsen, A. (2018). Inferring population structure and admixture proportions in low-depth NGS data. *Genetics*, *210*, 719–731.

Meyer, M., & Kircher, M. (2010). Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols*, *5*, prot5448.

Mikheyev, A. S., Tin, M. M. Y., Arora, J., & Seeley, T. D. (2015). Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nature Communications*, *6*, 7991.

Nadachowska-Brzyska, K., Dutoit, L., Smeds, L., Kardos, M., Gustafsson, L., & Ellegren, H. (2021). Genomic inference of contemporary effective population size in a large island population of collared flycatchers (*Ficedula albicollis*). *Molecular Ecology*, *30*, 3965–3973.

Nadachowska-Brzyska, K., Konczal, M., & Babik, W. (2022). Navigating the temporal continuum of effective population size. *Methods in Ecology and Evolution*, *13*, 22–41.

Neukamm, J., Peltzer, A., & Nieselt, K. (2021). DamageProfiler: fast damage pattern calculation for ancient DNA. *Bioinformatics*, *37*, 3652.

Nielsen, R., Korneliussen, T., Albrechtsen, A., Li, Y., & Wang, J. (2012). SNP calling, genotype calling, and sample allele frequency estimation from new-generation sequencing data. *PLoS ONE*, *7*, e37558.

Norder, S. J., de Lima, R. F., de Nascimento, L., Lim, J. Y., Fernández-Palacios, J. M., Romeiras, M. M., Elias, R. B., *et al.* (2020). Global change in microcosms: Environmental and societal predictors of land cover change on the Atlantic Ocean Islands. *Anthropocene*, *30*, 100242.

Nosil, P., & Feder, J. L. (2012). Genomic divergence during speciation: Causes and consequences. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*, 332–342.

Okonechnikov, K., Conesa, A., & García-Alcalde, F. (2016). Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*, 32, 292–294.

Olivera, P., Menezes, D., Trout, R., Buckle, A., Geraldes, P., Jesus, J., Oliveira, P., *et al.* (2010). Successful eradication of the European rabbit (*Oryctolagus cuniculus*) and house mouse (*Mus musculus*) from the island of Selvagem Grande (Macaronesian archipelago), in the Eastern Atlantic. *Integrative Zoology*, *5*, 70–83.

Orlando, L., Ginolhac, A., Zhang, G., Froese, D., Albrechtsen, A., Stiller, M., Schubert, M., *et al.* (2013). Recalibrating *Equus* evolution using the genome sequence of an early Middle Pleistocene horse. *Nature*, *499*, 74–78.

Otto, R., Krüsi, B. O., & Kienast, F. (2007). Degradation of an arid coastal landscape in relation to land use changes in Southern Tenerife (Canary Islands). *Journal of Arid Environments*, *70*, 527–539.

Otto, S. P. (2018). Adaptation, speciation and extinction in the Anthropocene. *Proceedings of the Royal Society B: Biological Sciences*, *285*, 20182047.

Parejo, M., Wragg, D., Henriques, D., Charrière, J.-D., & Estonba, A. (2020). Digging into the genomic past of Swiss honey bees by whole-genome sequencing museum specimens. *Genome Biology and Evolution*, *12*, 2535–2551.

Parks, M., & Lambert, D. (2015). Impacts of low coverage depths and post-mortem DNA damage on variant calling: A simulation study. *BMC Genomics*, *16*, 19.

Pemberton, J. M., Kruuk, L. E. B., & Clutton-Brock, T. (2022). The unusual value of long-term studies of individuals: The example of the Isle of Rum Red Deer Project. *Annual Review of Ecology, Evolution, and Systematics*, *53*, 327–351.

Pinsky, M. L., Eikeset, A. M., Helmerson, C., Bradbury, I. R., Bentzen, P., Morris, C., Gondek-Wyrozemska, A. T., *et al.* (2021). Genomic stability through time despite decades of exploitation in cod on both sides of the Atlantic. *Proceedings of the National Academy of Sciences of the United States of America*, *118*, e2025453118.

Pinto, A. V., Hansson, B., Patramanis, I., Morales, H. E., & van Oosterhout, C. (2023). The impact of habitat loss and population fragmentation on genomic erosion. *Conservation Genetics*, *25*, 49–57.

Pruett, C. L., & Winker, K. (2005). Northwestern song sparrow populations show genetic effects of sequential colonization. *Molecular Ecology*, *14*, 1421–1434.

R Core Team (2023). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rando, J.C. (2002). New data of fossil birds from El Hierro (Canary Islands): probable causes of extinction and some biogeographical considerations. *Ardeola*, *49*, 39–49.

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*, 1049–1060.

Reed, D., & 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.

Rodrigues, M. (2016). Land-use in the Macaronesian islands of Portugal and Spain. *Journal of Maps*, *12*, 1258–1261.

Ryman, N., Laikre, L., & Hössjer, O. (2019). Do estimates of contemporary effective population size tell us what we want to know? *Molecular Ecology*, *28*, 1904–1918.

Saha, A., Kurland, S., Kutschera, V. E., Díez-del-Molino, D., Ekman, D., Ryman, N., & Laikre, L. (2024). Monitoring genome-wide diversity over contemporary time with new indicators applied to Arctic charr populations. *Conservation Genetics*, *25*, 513–531.

Sánchez-Barreiro, F., Gopalakrishnan, S., Ramos-Madrigal, J., Westbury, M. V., de Manuel, M., Margaryan, A., Ciucani, M. M., *et al.* (2021). Historical population declines prompted significant genomic erosion in the northern and southern white rhinoceros (*Ceratotherium simum*). *Molecular Ecology*, *30*, 6355–6369.

Santiago, E., Novo, I., Pardiñas, A. F., Saura, M., Wang, J., & Caballero, A. (2020). Recent demographic history inferred by high-resolution analysis of linkage disequilibrium. *Molecular Biology and Evolution*, *37*, 3642–3653.

Sefc, K. M., Payne, R. B., & Sorenson, M. D. (2007). Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics*, *8*, 879–884.

Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., van Doren, B. M., Underwood, V. L., Coulson, T., *et al.* (2020). The genomic landscape of divergence across the speciation continuum in island-colonising Silvereyes (*Zosterops lateralis*). *G3*, *10*, 3147–3163.

Sheppard, E. C., Martin, C. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2022). Genomic associations with pox virus across divergent populations in Berthelot's pipit. *Molecular Ecology*, *31*, 3154–3173.

Sheppard, E. C., Martin, C. A., Illera, J. C., Suh, A., Spurgin, L. G., & Richardson, D. S. (2024). Genotype-environment associations reveal genes potentially linked to avian malaria infection in populations of an endemic island bird. *Molecular Ecology*, *33*, e17329.

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.

Stiller, M., Green, R. E., Ronan, M., Simons, J. F., Du, L., He, W., Egholm, M., *et al.* (2006). Patterns of nucleotide misincorporations during enzymatic amplification and direct large-scale sequencing of ancient DNA. *Proceedings of the National Academy of Sciences of the United States of America*, *103*, 13578–13584.

Stoffel, M. A., Johnston, S. E., Pilkington, J. G., & Pemberton, J. M. (2021). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *Nature Communications*, *12*, 2972.

Traveset, A., Nogales, M., Alcover, J. A., Delgado, J. D., López-Darias, M., Godoy, D., Igual, J. M., *et al.* (2009). A review on the effects of alien rodents in the Balearic (western Mediterranean Sea) and Canary Islands (eastern Atlantic Ocean). *Biological Invasions*, *11*, 1653–1670.

Tsai, W. L. E., Schedl, M. E., Maley, J. M., & McCormack, J. E. (2020). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources*, *20*, 1220–1227.

Van Der Valk, T., Díez-del-Molino, D., Marques-Bonet, T., Guschanski, K., & Dalén, L. (2019). Historical genomes reveal the genomic consequences of recent population decline in eastern gorillas. *Current Biology*, *29*, 165–170.

Vasimuddin, M. D., Misra, S., Li, H., & Aluru, S. (2019). Efficient architecture-aware acceleration of BWA-MEM for multicore systems. *IEEE Parallel and Distributed Processing Symposium (IPDPS)*. Available at: https://doi.org/10.1109/ipdps.2019.00041.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., *et al.* (2016). Sixteen years of change in the global terrestrial human footprint and implications for biodiversity conservation. *Nature Communications*, *7*, 12558.

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

Wang, P., Hou, R., Wu, Y., Zhang, Z., Que, P., & Chen, P. (2022). Genomic status of yellow-breasted bunting following recent rapid population decline. *iScience*, *25*, 104501.

Willi, Y., Van Buskirk, J., & Hoffmann, A. (2006). Limits to the adaptive potential of small populations. *Annual Review of Ecology, Evolution, and Systematics*, *37*, 433–458.

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

Wu, M. Y., Lau, C. J., Xin Ng, E. Y., Baveja, P., Gwee, C. Y., Sadanandan, K., Ferasyi, T. R., *et al.* (2022). Genomes from historic DNA unveil massive hidden extinction and terminal endangerment in a tropical Asian songbird radiation. *Molecular Biology and Evolution*, *39*, msac189.

Zhang, J., Kobert, K., Flouri, T., & Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics*, *30*, 614–620.

4.7. Supplements

Table S4.1. Detail of the original island location and museum sources of the historical and contemporary Berthelot's pipit samples used in this study.

Population	**Locality**	**Collection year**	**Sample ID**	**Catalogue number (historical)**
Historical samples				
Canary Islands (CI)				
Tenerife (TF)	Santa Úrsula	1888	Aber14102	LivT14102
	Orotava	1888	Aber19051222101	NHM1905.12.22.101
	Santiago	1888	Aber19051222102	NHM1905.12.22.102
	Tenerife I.	1892	Aber572597	AMNH572597
	NA	1893	Aber19415307482	NHM1941.5.30.7482
	Road to Laguna, 2,000 ft	1895	Aber18957150	NHM1895.7.1.50
Lanzarote (LZ)	*NA*	1902	Aber19201030149	NHM1920.10.30.149
	NA	1903	Aber19201030150	NHM1920.10.30.150
	Yaiza	1913	Aber1913102238	NHM1913.10.22.38
	Haria, 1,100 ft	1913	Aber1918102236	NHM1918.10.22.36
	Playa Janubio, 50 ft	1913	Aber191981587	NHM1919.8.15.87
Gomera (GOM)	Valley near Hermigua, 1000 ft	1920	Aber19201030397	NHM1920.10.30.397
	Between San Sebastian and Hermigua, 3,000 ft	1920	Aber19201030398	NHM1920.10.30.398
La Graciosa (GRA)	*NA*	1913	Aber1913102241	NHM1913.10.22.41
	NA	1913	Aber1913102243	NHM1913.10.22.43
El Hierro (EH)	Tijadaya	1890	Aber15990	LivT15990
	Tijadaya	1890	Aber15998	LivT15998
Gran Canaria (GC)	Atalaya	1888	Aber13941	LivT13941
	Las Palmas	1888	Aber18986	LivT18986
Fuerteventura (FV)	Oliva	1910	Aber122983	RMNH.AVES.122983
Madeiran archipelago (M)				
Madeira (M)	Ribeira Brava	1883	Aber122979	RMNH.AVES.122979
	NA	1884	Aber1989664184	Liv1989.66.4184
	Madeira I.	1890	Aber572631	AMNH572631
	NA	1900	Aber122978	RMNH.AVES.122978
	Above Calheta, Madeira I.	1903	Aber572616	AMNH572616

Table S4.1. (Continued).

Population	Locality	Collection year	Sample ID	Catalogue number (historical)
Historical samples				
Madeiran archipelago (M)				
Porto Santo (PS)	*NA*	1894	Aber1896265	NHM1896.2.6.5
	NA	1895	Aber18957148	NHM1895.7.1.48
Selvagens (S)				
Selvagem Grande (SG)	*NA*	1895	Aber18957152	NHM1895.7.1.52
	NA	1895	Aber18957153	NHM1895.7.1.53
Contemporary samples				
Canary Islands (CI)				
Tenerife (TF)	Buenavista	2009	TF602	*NA*
	Los Silos	2009	TF606	*NA*
	El Medano	2009	TF609	*NA*
	El Fraile	2009	TF617	*NA*
	Guimar	2009	TF623	*NA*
	El Poris	2009	TF630	*NA*
	Las Cañadas	2009	TEID925	*NA*
	Las Cañadas	2009	TEID928	*NA*
	Las Cañadas	2009	TEID930	*NA*
	Las Cañadas	2009	TEID947	*NA*
	Las Cañadas	2009	TEID952	*NA*
	Las Cañadas	2009	TEID957	*NA*
Lanzarote (LZ)	Puerto Calero	2009	LZ676	*NA*
	Tias	2009	LZ678	*NA*
	Punta Grande	2009	LZ685	*NA*
	Teguise	2009	LZ687	*NA*
	Orzola	2009	LZ693	*NA*
	Haria	2009	LZ694	*NA*
Gomera (GOM)	Near airport	2009	GOM637	*NA*
	El Cercado	2009	GOM644	*NA*
La Graciosa (GRA)	*NA*	2009	GRA659	*NA*
	NA	2009	GRA671	*NA*
El Hierro (EH)	Las Puntas	2009	EH761	*NA*
	El Pinar	2009	EH779	*NA*

Table S4.1. (Continued).

Population	Locality	Collection year	Sample ID	Catalogue number (historical)
Contemporary samples				
Canary Islands (CI)				
Gran Canaria (GC)	Castilla del Romeral	2009	GC739	*NA*
	San Jose del Alomo	2009	GC811	*NA*
Fuerteventura (FV)	Cañada de la Mata	2009	FV718	*NA*
	Las Salinas	2009	FV722	*NA*
Madeiran archipelago (M)				
Madeira (M)	Ponta do Pargo	2009	M844	*NA*
	Ponta do Pargo	2009	M845	*NA*
	Ponta Sao Laurenco	2009	M849	*NA*
	Pol de Serra	2009	M858	*NA*
	Pol de Serra	2009	M861	*NA*
	Ponta Sao Laurenco	2009	M905	*NA*
Porto Santo (PS)	*NA*	2006	PS506	*NA*
	NA	2009	PS817	*NA*
	NA	2009	PS820	*NA*
	NA	2009	PS833	*NA*
	NA	2009	PS839	*NA*
Selvagens (S)				
Selvagem Grande (SG)	*NA*	2009	SG863	*NA*
	NA	2009	SG870	*NA*
	NA	2009	SG871	*NA*
	NA	2009	SG878	*NA*
	NA	2009	SG881	*NA*
	NA	2009	SG900	*NA*

Museum code: Liv = National Museums Liverpool, NHM = Natural History Museum, Tring, AMNH = American Museum of Natural History, RMNH = Naturalis Biodiversity Center in Leiden.

Table S4.2. Global processing and mapping statistics for historical and contemporary whole genome-resequencing data.

Sample ID	Sequencing batch	# Raw reads	# Cleaned reads	# Mapped reads	% Mapped reads	Avg. read length	Avg. coverage
Historical samples							
Aber122978	P24505	333,091,016	131,124,282	120,528,603	91.92	120	13
Aber122979	P24505	270,773,242	81,492,272	70,560,258	86.59	127	8
Aber122983	P24506	646,148,464	218,389,488	197,655,597	90.51	125	22
Aber13941	P24506	521,604,056	156,909,318	138,355,826	88.18	124	15
Aber14102	P24505	257,529,134	67,618,195	57,057,838	84.38	122	6
Aber15990	P24506	904,204,694	241,826,241	212,497,814	87.87	123	23
Aber15998	P24506	350,766,048	105,651,508	92,154,155	87.22	127	10
Aber18957148	P24506	442,522,478	100,724,501	80,689,959	80.11	121	8
Aber18957150	P24505	204,162,710	52,173,920	44,681,017	85.64	117	5
Aber18957152	P24505	593,617,456	190,801,097	171,144,529	89.7	126	19
Aber18957153	P24505	445,142,510	153,369,462	137,614,245	89.73	125	15
Aber1896265	P24506	487,197,286	119,653,949	98,490,631	82.31	135	10
Aber18986	P24506	425,641,634	121,984,431	84,544,314	69.31	125	9
Aber19051222101	P24505	373,463,198	118,384,762	105,217,522	88.88	126	11
Aber19051222102	P24505	225,402,512	75,580,261	68,131,968	90.15	124	7
Aber1913102238	P24505	289,763,474	105,833,338	96,482,853	91.16	126	11
Aber1913102241	P24506	366,178,016	55,181,573	37,611,922	68.16	126	4
Aber1913102243	P24506	415,711,412	120,733,837	105,364,198	87.27	123	11
Aber1918102236	P24505	329,578,246	120,103,429	108,874,254	90.65	126	12
Aber191981587	P24505	435,875,474	144,828,973	130,349,125	90.00	125	14
Aber19201030149	P24505	281,638,518	60,033,568	46,831,479	78.01	117	5

Table S4.2. (Continued).

Sample ID	Sequencing batch	# Raw reads	# Cleaned reads	# Mapped reads	% Mapped reads	Avg. read length	Avg. coverage
Historical samples							
Aber19201030150	P24505	279,376,090	87,153,122	76,389,978	87.65	128	8
Aber19201030397	P24506	357,957,936	83,730,881	68,356,047	81.64	125	7
Aber19201030398	P24506	624,710,548	122,857,707	105,659,322	86.00	122	11
Aber19415307482	P24505	223,815,974	56,806,346	47,104,857	82.92	120	5
Aber1989664184	P24505	266,600,970	73,093,853	64,099,240	87.69	125	7
Aber572597	P24505	228,368,892	56,663,268	49,158,260	86.76	126	5
Aber572616	P24505	240,367,572	69,491,789	61,021,697	87.81	126	7
Aber572631	P24505	263,471,770	68,553,700	58,430,736	85.23	120	6
Contemporary samples							
PS506	Previous study	160,411,954	127,353,597	122,083,821	95.86	159	17
TF602	P28102	195,888,224	152,336,204	145,164,912	95.29	157	20
TF606	Previous study	186,344,984	147,163,913	141,416,371	96.09	160	20
TF609	P28102	165,578,964	128,672,245	122,441,848	95.16	157	17
TF617	Previous study	198,939,116	155,405,486	149,345,408	96.10	160	21
TF623	P28102	264,372,600	180,364,685	171,435,118	95.05	170	26
TF630	P28102	112,133,004	88,743,645	84,790,510	95.55	157	12
GOM637	P28102	144,057,478	113,273,664	108,111,460	95.44	156	15
GOM644	P28102	185,015,124	142,607,004	135,805,022	95.23	156	19
GRA659	P28102	135,069,046	105,481,927	100,526,655	95.30	157	14
GRA671	P28102	125,672,140	99,506,243	94,720,961	95.19	155	13
LZ676	P28102	155,923,892	119,621,186	113,487,511	94.87	158	16

Table S4.2. (Continued).

Sample ID	Sequencing batch	# Raw reads	# Cleaned reads	# Mapped reads	% Mapped reads	Avg. read length	Avg. coverage
Contemporary samples							
LZ678	P28102	170,349,768	129,605,585	123,190,124	95.05	159	17
LZ685	P28102	137,548,564	103,269,222	98,205,291	95.10	157	14
LZ687	Previous study	239,064,654	186,741,921	178,732,523	95.71	159	25
LZ693	Previous study	200,350,286	157,912,747	151,868,283	96.17	160	22
LZ694	P28102	137,149,826	106,750,338	102,284,197	95.82	157	14
FV718	P28102	146,636,612	113,457,827	108,304,508	95.46	157	15
FV722	P28102	142,357,030	110,898,136	105,615,626	95.24	157	15
GC739	P28102	145,919,736	112,389,451	107,404,247	95.56	158	15
EH761	Previous study	174,351,340	134,402,578	128,480,623	95.59	162	19
EH779	Previous study	182,927,766	144,854,598	138,775,893	95.80	158	20
GC811	P28102	140,873,250	109,437,313	104,240,703	95.25	157	14
PS817	P28102	135,696,068	104,608,040	100,115,640	95.71	160	14
PS820	P28102	136,880,340	105,828,322	100,628,938	95.09	159	14
PS833	P28102	143,423,132	108,261,467	103,135,732	95.27	160	15
PS839	P28102	142,438,020	108,245,372	103,102,690	95.25	158	14
M844	P28102	165,026,630	127,081,287	121,734,221	95.79	159	17
M845	P28102	143,277,154	109,678,444	104,692,004	95.45	159	15
M849	Previous study	192,781,206	149,919,772	143,823,424	95.93	160	21
M858	P28102	129,570,756	100,552,390	96,051,045	95.52	158	13
M861	P28102	138,481,982	106,357,227	101,547,940	95.48	159	14
SG863	P28102	156,229,074	120,374,941	114,699,586	95.29	159	16

Table S4.2. (Continued).

Sample ID	Sequencing batch	# Raw reads	# Cleaned reads	# Mapped reads	% Mapped reads	Avg. read length	Avg. coverage
Contemporary samples							
SG870	P28102	153,794,016	116,689,091	111,050,209	95.17	160	16
SG871	P28102	147,744,782	111,887,134	106,658,577	95.33	160	15
SG878	Previous study	166,355,910	134,391,275	128,853,364	95.88	156	18
SG881	P28102	149,689,080	114,950,395	109,424,647	95.19	158	15
SG900	Previous study	183,823,472	143,054,935	136,470,758	95.40	160	20
M905	Previous study	174,879,396	130,265,192	124,412,392	95.51	166	18
TEID925	P28102	130,585,730	100,304,047	95,535,762	95.25	159	13
TEID928	P28102	120,526,500	93,122,592	88,429,993	94.96	159	12
TEID930	P28102	149,828,906	114,209,171	108,874,846	95.33	159	15
TEID947	P28102	133,468,920	101,477,409	96,890,083	95.48	159	14
TEID952	P28102	170,566,480	128,079,925	122,127,809	95.35	161	17
TEID957	P28102	167,838,810	126,283,245	120,197,780	95.18	160	17

Note: Samples labelled as "Previous study" were originally sequenced in Martin *et al.*, (2023) but are processed here in line with the current study. Batches with specific names correspond to newly generated data for this study.

Table S4.3. Scaffold-specific coverage for the historical and contemporary Berthelot's pipits used in this study following cleaning and mapping pipelines. Assigned sex inferred from Z-autosome coverage ratios.

Sample ID	**Avg. autosome coverage**	**Avg. mitochondrial coverage**	**Avg. Z coverage**	**Assigned sex**
Historical samples				
Aber122978	12.3	289.9	11.2	M
Aber122979	7.4	211.1	6.5	M
Aber122983	21.3	807.5	9.9	F
Aber13941	14.7	337.0	6.7	F
Aber14102	5.7	233.8	4.8	M
Aber15990	21.9	268.8	9.7	F
Aber15998	9.8	101.1	4.3	F
Aber18957148	7.9	115.2	6.7	M
Aber18957150	4.2	73.6	3.6	M
Aber18957152	18.2	407.0	8.5	F
Aber18957153	14.9	350.7	6.9	F
Aber1896265	9.0	121.3	3.9	F
Aber18986	8.6	179.4	7.2	M
Aber19051222101	11.2	212.0	5.2	F
Aber19051222102	7.0	147.0	6.2	M
Aber1913102238	10.1	99.9	9.1	M
Aber1913102241	3.2	112.7	2.6	M
Aber1913102243	10.6	202.0	9.1	M
Aber1918102236	11.6	136.6	5.4	F
Aber191981587	13.5	194.3	11.8	M
Aber19201030149	4.5	88.4	2.0	F
Aber19201030150	8.3	262.8	3.9	F
Aber19201030397	7.0	324.7	3.1	F
Aber19201030398	10.6	207.2	9.0	M
Aber19415307482	4.7	47.4	2.1	F
Aber1989664184	6.6	89.1	3.0	F
Aber572597	4.8	476.6	4.2	M
Aber572616	6.2	573.5	5.3	M
Aber572631	5.6	139.2	2.4	F
Contemporary samples				
EH761	19.0	25.7	9.6	F
EH779	19.5	29.3	18.8	M
FV718	14.8	21.0	13.7	M

Table S4.3. (Continued).

Sample ID	Avg. autosome coverage	Avg. mitochondrial coverage	Avg. Z coverage	Assigned sex
Contemporary samples				
FV722	14.9	25.9	7.5	F
GC739	14.8	19.0	14.2	M
GC811	14.6	19.9	7.5	F
GOM637	15.2	23.9	7.8	F
GOM644	18.4	23.1	17.4	M
GRA659	13.7	17.1	12.3	M
GRA671	13.2	19.2	6.7	F
LZ676	15.9	27.7	7.6	F
LZ678	17.4	28.8	8.7	F
LZ685	13.4	31.0	13.1	M
LZ687	25.2	53.4	24.5	M
LZ693	22.4	36.1	11.8	F
LZ694	14.0	15.7	13.8	M
M844	16.9	23.2	16.4	M
M845	14.9	20.8	7.6	F
M849	20.5	26.4	20.2	M
M858	13.6	16.8	7.1	F
M861	14.0	15.6	13.1	M
M905	18.8	16.5	9.4	F
PS506	17.8	41.2	9.4	F
PS817	13.9	11.8	13.5	M
PS820	14.3	18.4	7.1	F
PS833	14.7	15.0	7.6	F
PS839	14.2	21.3	13.2	M
SG863	16.3	22.1	8.2	F
SG870	15.3	15.8	14.1	M
SG871	14.8	16.0	14.2	M
SG878	17.9	22.6	17.7	M
SG881	15.5	16.8	7.9	F
SG900	19.8	26.6	10.0	F
TEID925	13.6	14.8	6.9	F
TEID928	12.5	13.3	6.0	F
TEID930	15.1	28.5	14.0	M
TEID947	13.4	17.0	12.7	M

Table S4.3. (Continued).

Sample ID	Avg. autosome coverage	Avg. mitochondrial coverage	Avg. Z coverage	Assigned sex
Contemporary samples				
TEID952	17.1	21.0	16.1	M
TEID957	17.2	13.4	8.8	F
TF602	19.7	18.5	18.6	M
TF606	20.8	26.3	10.9	F
TF609	17.2	13.9	8.7	F
TF617	21.3	41.0	21.1	M
TF623	26.0	25.7	13.0	F
TF630	11.6	13.4	11.1	M

Table S4.4. Results of non-parametric pairwise Wilcoxon rank sum tests after Benjamini–Hochberg correction to test for significant spatial differences in heterozygosity among contemporary populations of Berthelot's pipits. Significant *p*-values are highlighted in bold and sample sizes are shown above each population. Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

		n = 2	*n* = 2	*n* = 2	*n* = 2	*n* = 2	*n* = 6	*n* = 6	*n* = 5	*n* = 6
		EH	**FV**	**GC**	**GOM**	**GRA**	**LZ**	**M**	**PS**	**SG**
n = 2	**FV**	0.75								
n = 2	**GC**	1.00	0.54							
n = 2	**GOM**	0.75	0.54	0.75						
n = 2	**GRA**	1.00	0.54	0.75	1.00					
n = 6	**LZ**	0.75	0.94	0.75	0.62	0.62				
n = 6	**M**	0.18	0.18	0.18	0.18	0.18	**< 0.05**			
n = 5	**PS**	0.19	0.19	0.19	0.19	0.19	**< 0.05**	0.75		
n = 6	**SG**	0.18	0.18	0.18	0.18	0.18	**< 0.05**	**< 0.05**	**< 0.05**	
n = 12	**TF**	0.75	0.62	0.55	0.47	0.47	0.96	**< 0.01**	**< 0.01**	**< 0.01**

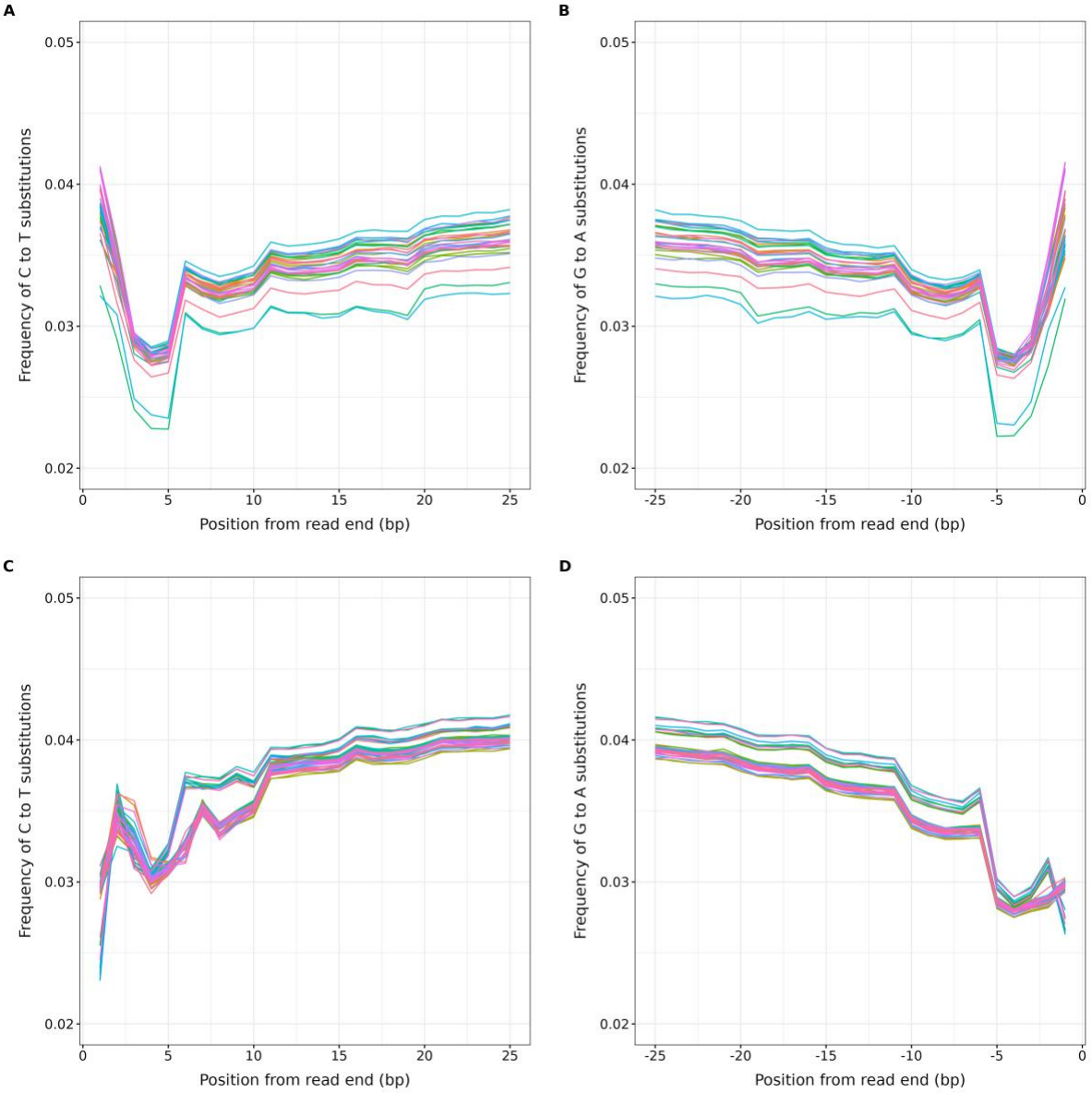

Figure S4.1. Comparison of DNA damage patterns in Berthelot's pipit genomes identified by DamageProfiler. Plots depict mean frequencies of transitions over the initial 25 positions from 5' and 3' read ends. Historical samples show uptick of (**A**) cytosine-to-thymine (C-to-T) substitutions at 5' read ends and (**B**) guanine-to-adenine (G-to-A) substitutions at 3' read ends, but this is minimal due to the application of USER enzyme treatment during library preparation. Note that the reference genome is derived from a white wagtail. Consequently, the contemporary genomes also show (**C**) C-to-T substitutions and (**D**) G-to-A substitutions, reflecting true nucleotide differences.

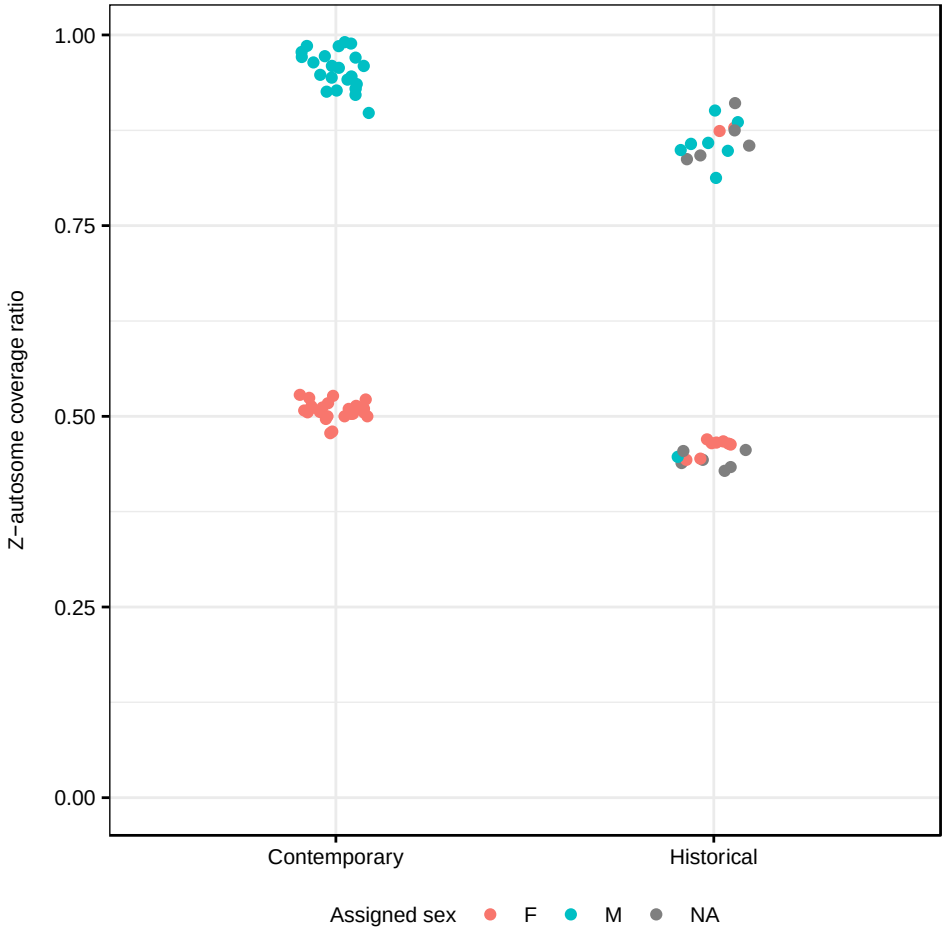

Figure S4.2. Z-autosomal coverage ratios in historical and contemporary Berthelot's pipit genomes. Each point represents a sample, with colours indicating the assigned sex based on previous PCR-based sex assignments (contemporary samples) or morphological determinations by museums (historical samples). As the heterogametic sex, females are expected to receive approximately half the coverage on the Z-related scaffolds.

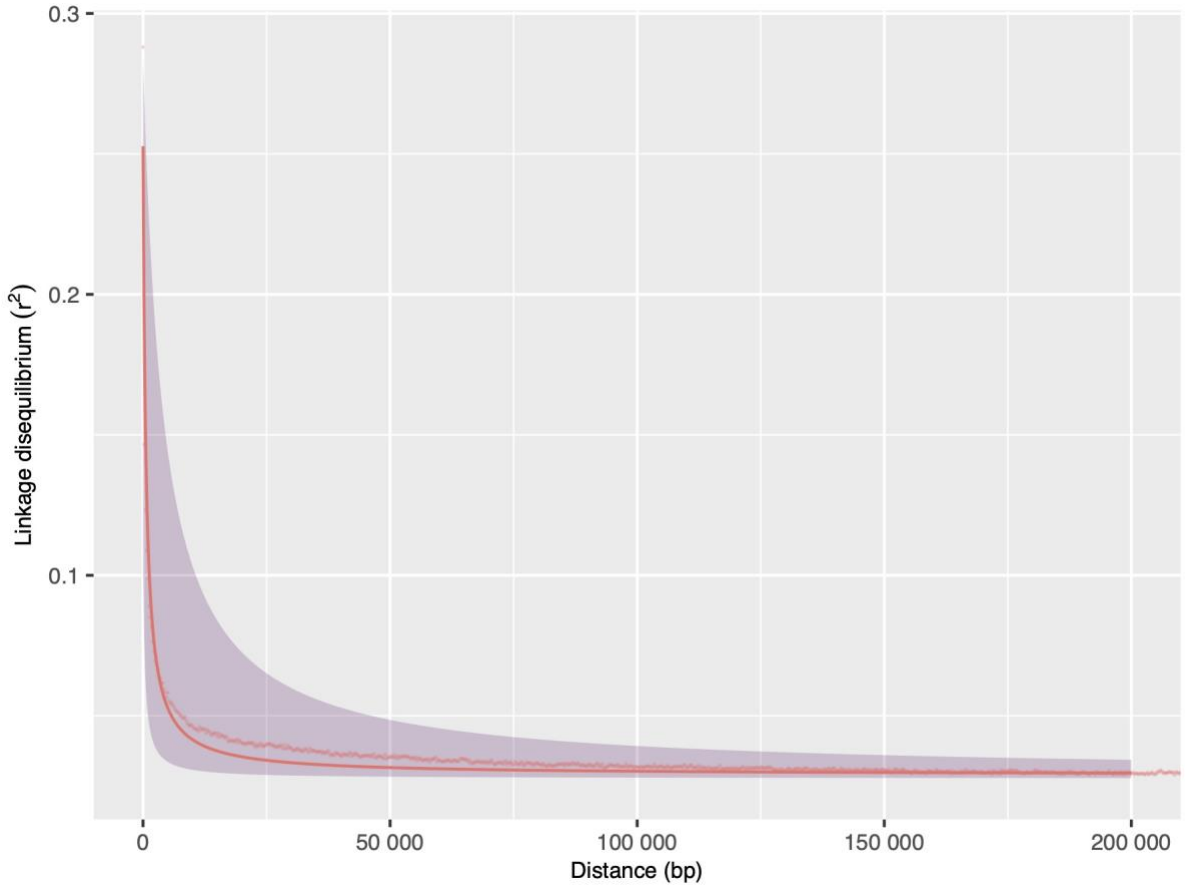

Figure S4.3. LD decay across 10 island populations of Berthelot's pipit. LD plotted using a random subsample consisting of approximately 1% of pairwise SNP comparisons from each chromosome. Maximum distance between SNPs for LD estimation: 500 kbp (only 0–200 kbp plotted). Red line represents model fit and shaded intervals represent 95% confidence level, derived from 100 bootstrap replicates.

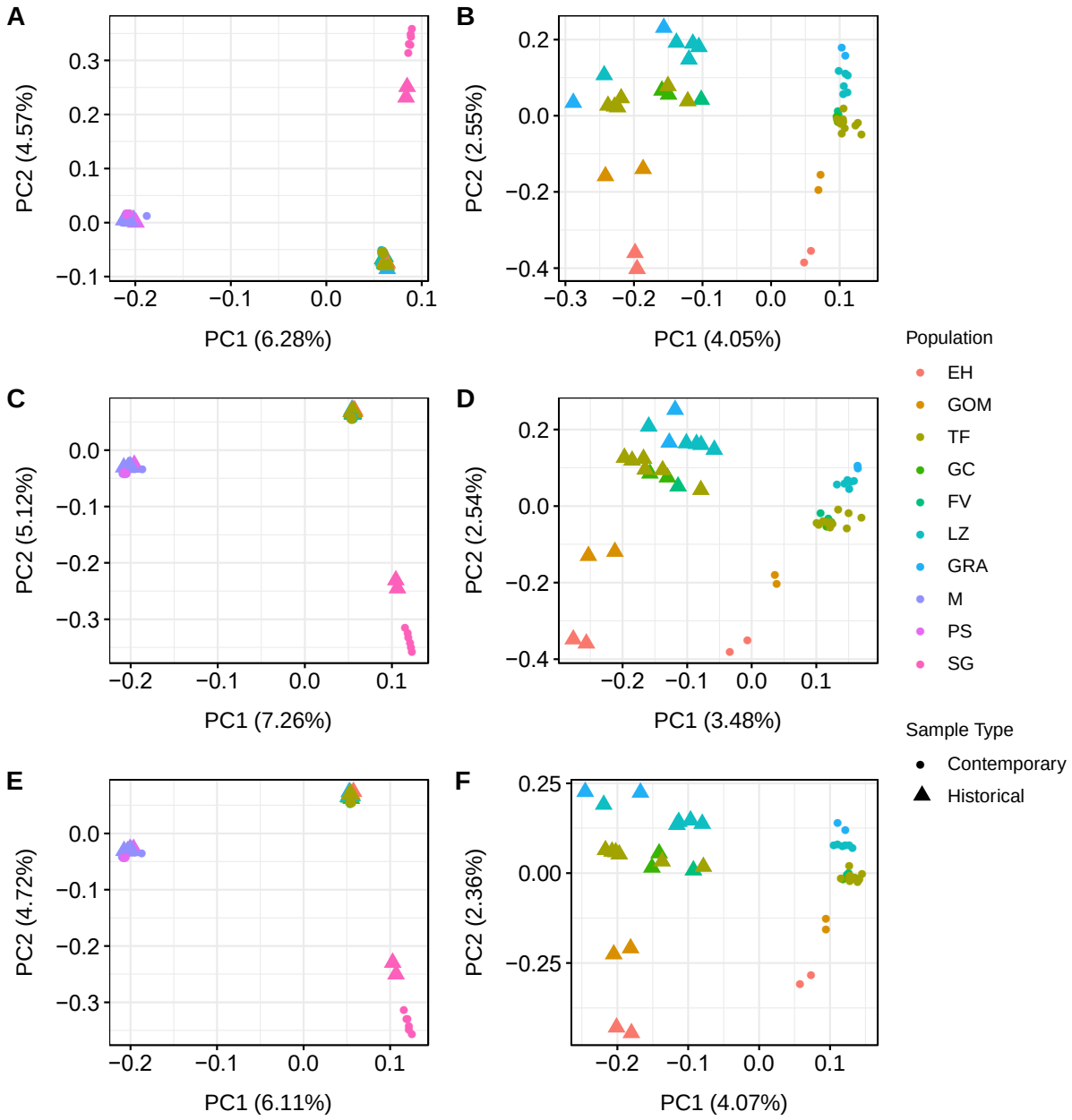

Figure S4.4. Principal component analyses of genomic variation in both historical and contemporary Berthelot's pipit samples, based on different average coverage. (**A**) Analysis is based on genotype likelihoods derived from 702,000 LD-pruned SNPs on the autosomes, with the full coverage dataset, among all populations, and (**B**) among populations within the largest archipelago only, the Canary Islands. (**C**) Analysis is based on genotype likelihoods derived from 431,199 LD-pruned SNPs on the autosomes, with an average of 5X coverage after down-sampling, among all populations, and (**D**) among populations within the largest archipelago only, the Canary Islands. (**E**) Analysis is based on genotype likelihoods derived from 705,782 LD-pruned SNPs on the autosomes, with an average of 3X coverage after down-sampling, among all populations, and (**F**) among populations within the largest archipelago only, the Canary Islands. For all PCAs, the first two principal components are shown, each indicating the proportion of the observed variance explained (in brackets). Samples are colour-coded based on their respective island populations and shaped according to sample type (historical versus contemporary). Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

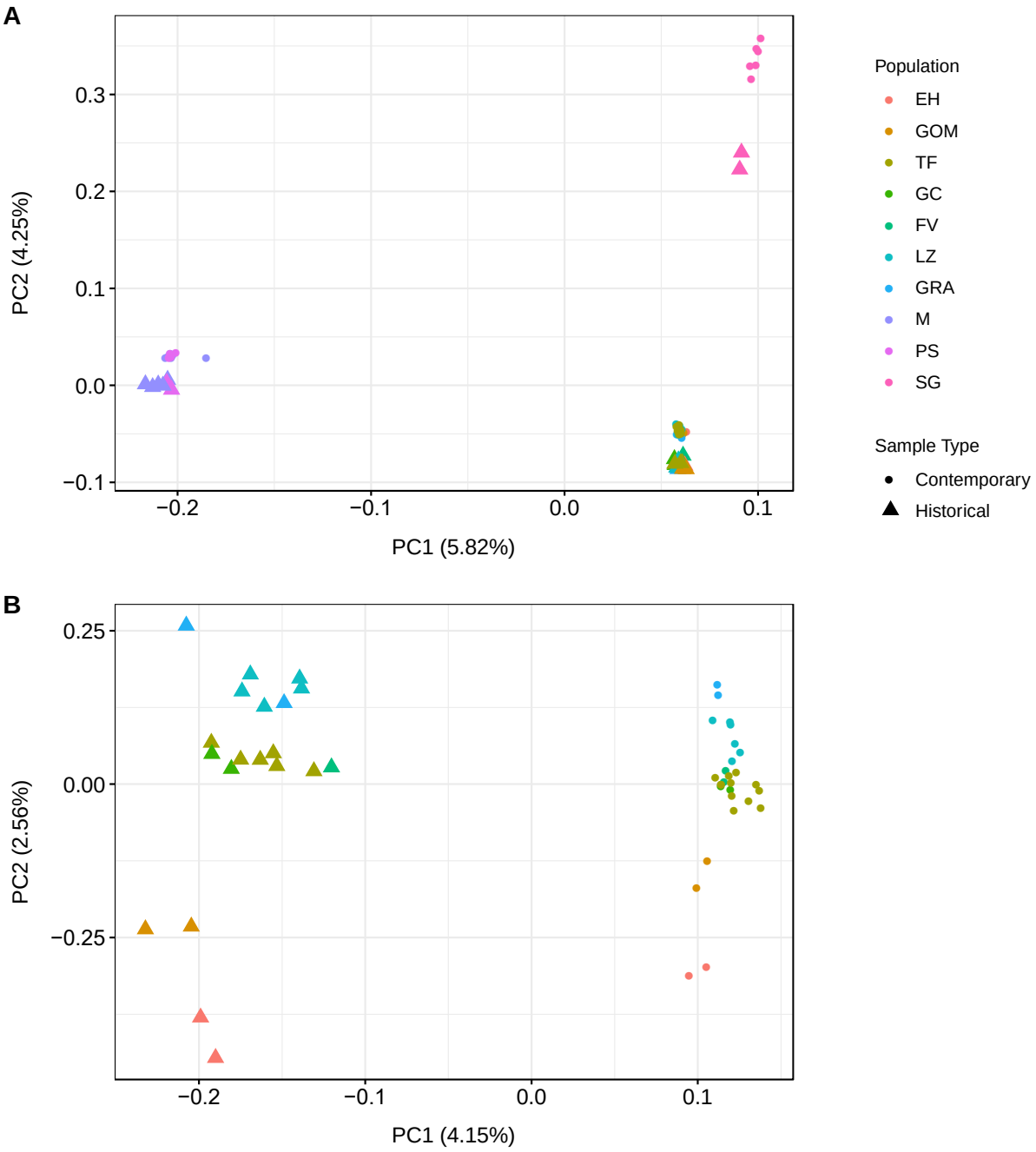

Figure S4.5. Principal component analyses of genomic variation in both historical and contemporary Berthelot's pipit samples, with all transitions removed, (**A**) among all populations, and (**B**) among populations within the largest archipelago only, the Canary Islands. The analysis is based on genotype likelihoods derived from 269,498 LD-pruned SNPs on the autosomes. The first two principal components are shown, each indicating the proportion of the observed variance explained (in brackets). Samples are colour-coded based on their respective island populations and shaped according to sample type (historical versus contemporary). Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

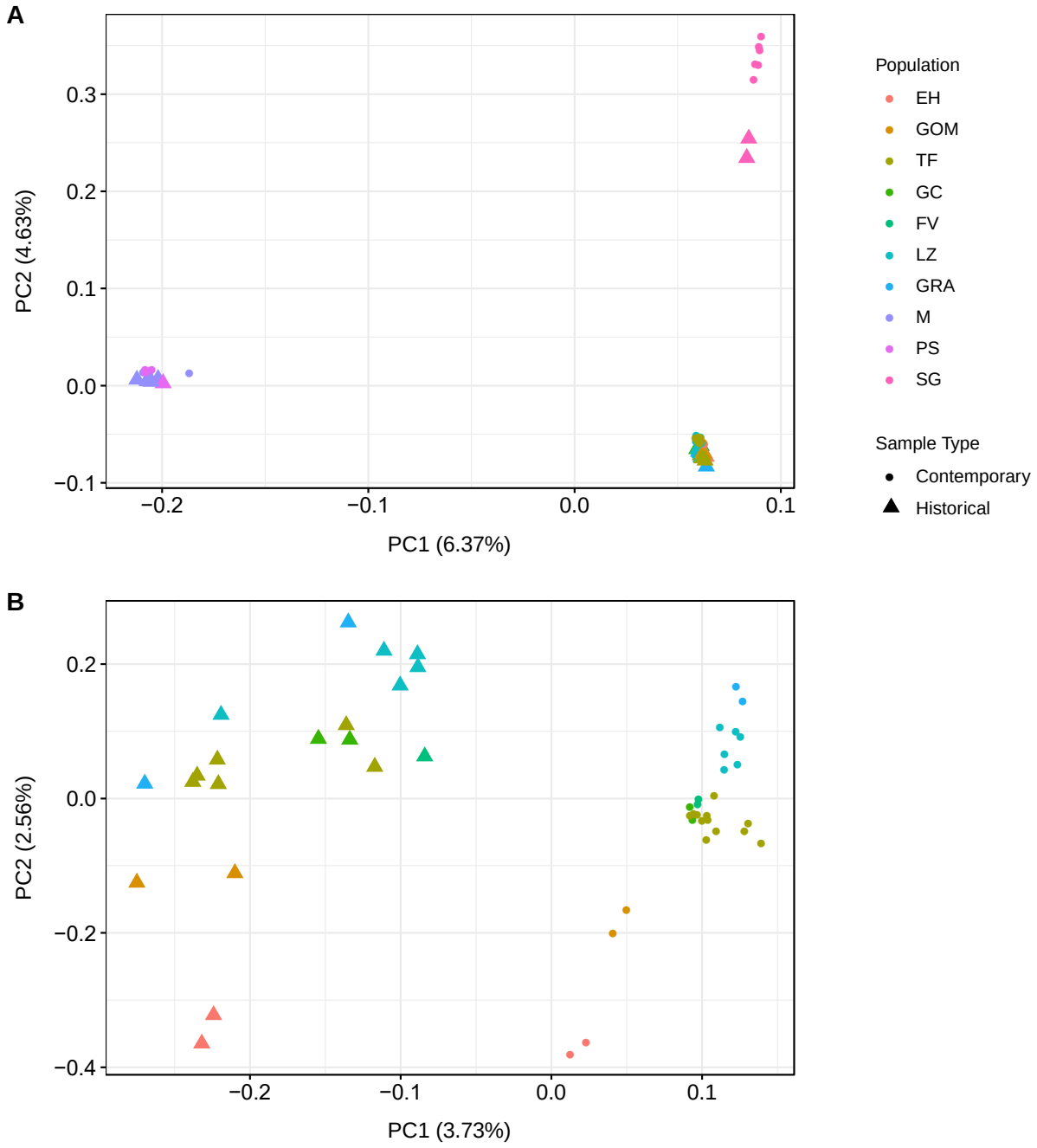

Figure S4.6. Principal component analyses of genomic variation in both historical and contemporary Berthelot's pipit samples, with sites potentially affected by reference bias/mis-mapping removed (**A**) among all populations, and (**B**) among populations within the largest archipelago only, the Canary Islands. The analysis is based on genotype likelihoods derived from 672,241 LD-pruned SNPs on the autosomes. The first two principal components are shown, each indicating the proportion of the observed variance explained (in brackets). Samples are colour-coded based on their respective island populations and shaped according to sample type (historical versus contemporary). Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa; Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande.

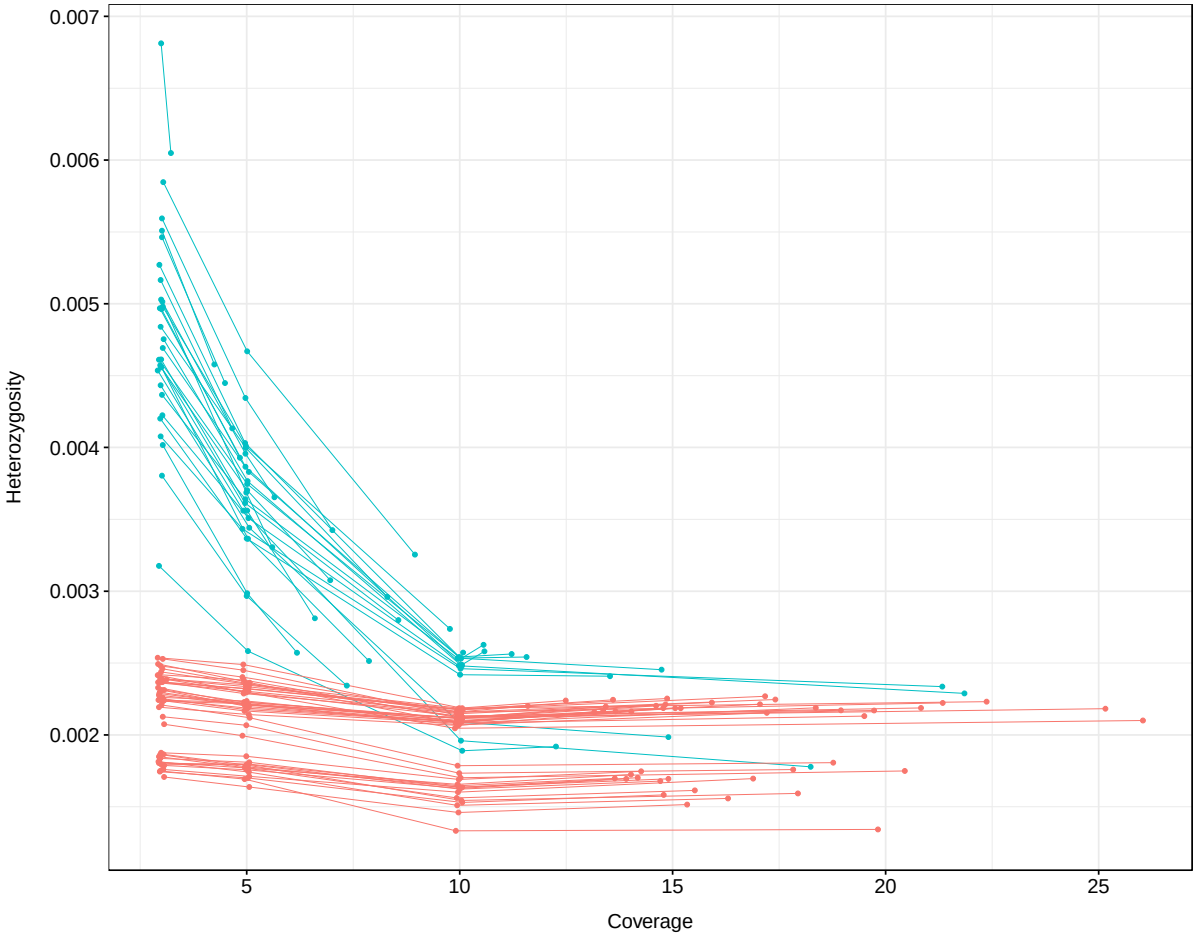

Figure S4.7. The effect of depth of coverage on autosomal heterozygosity estimates in Berthelot's pipits. Each line represents a sample, tracing its average coverage across different datasets. The datasets include 3X down-sampled, 5X down-sampled, 10X down-sampled, or 'full coverage', represented by the final point of each line. Samples are coloured according to their sampling time point: historical samples are depicted in blue, while contemporary samples are shown in red.

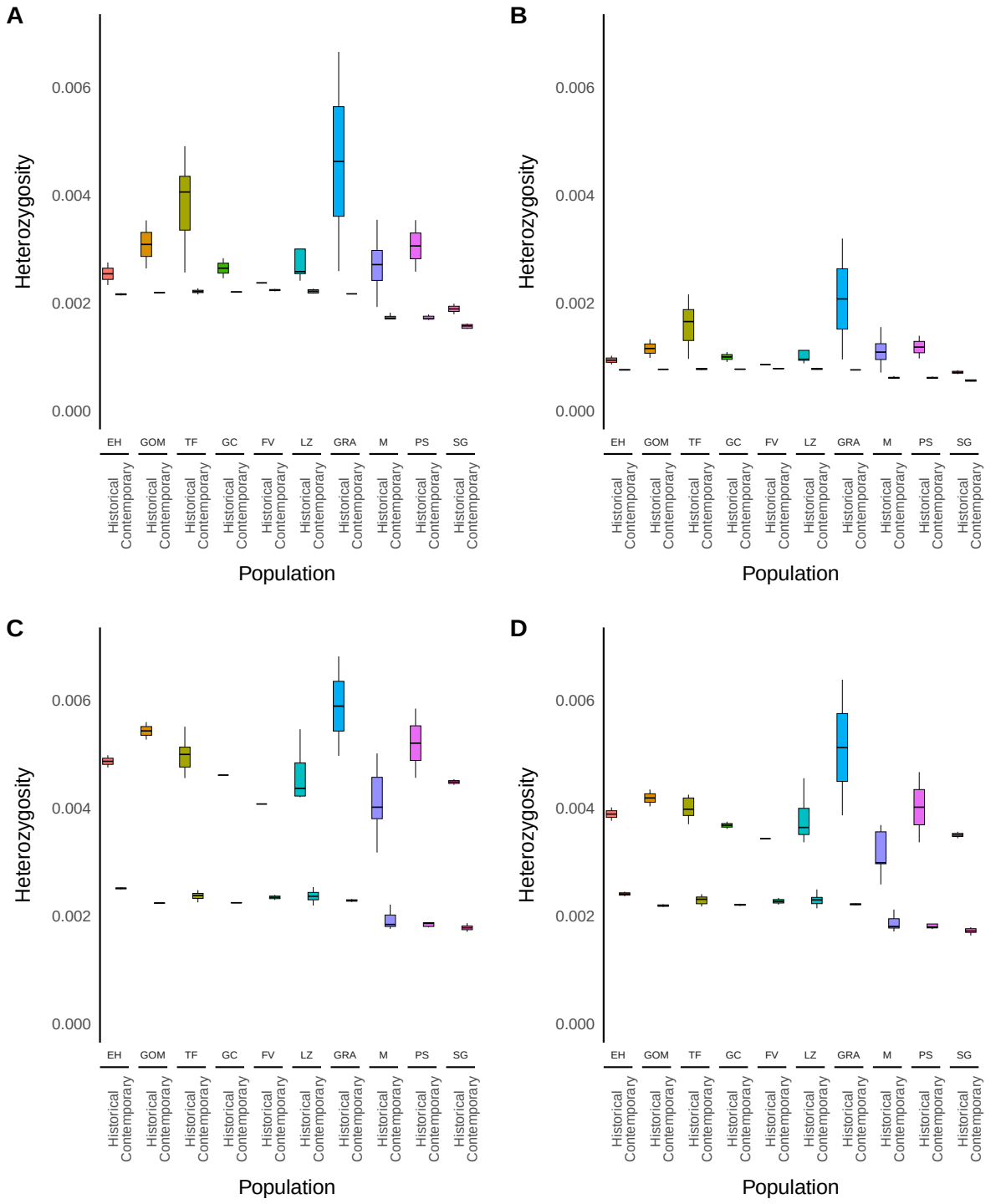

Figure S4.8. Spatio-temporal patterns of autosomal heterozygosity estimates of Berthelot's pipits across different datasets that attempt to identify any potential bias between historical and contemporary data. Individual autosomal heterozygosity estimates derived from (**A**) the 'full coverage' dataset, (**B**) the 'full coverage' dataset, with all transitions removed, (**C**) all samples with the same coverage after down-sampling to an average of 3X, (**D**) higher coverage individuals down-sampled to an average of 5X, and (**E**) the 'full coverage' dataset, excluding sites with high proportion of reads with lower mapping quality scores (as a proxy for reference bias). In all plots, estimates are grouped according to island population and time window. Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande; Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa.

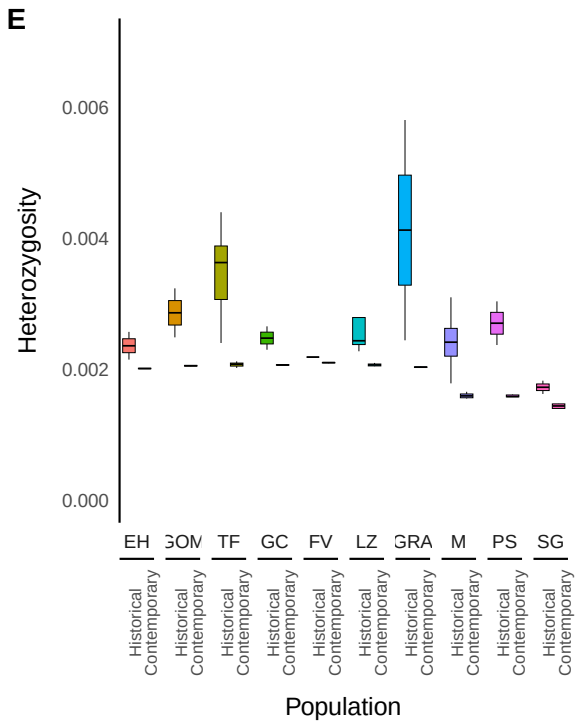

Figure S4.8. (Continued). Spatio-temporal patterns of autosomal heterozygosity estimates of Berthelot's pipits across different datasets that attempt to identify any potential bias between historical and contemporary data. Individual autosomal heterozygosity estimates derived from (**A**) the 'full coverage' dataset, (**B**) the 'full coverage' dataset, with all transitions removed, (**C**) all samples with the same coverage after down-sampling to an average of 3X, (**D**) higher coverage individuals down-sampled to an average of 5X, and (**E**) the 'full coverage' dataset, excluding sites with high proportion of reads with lower mapping quality scores (as a proxy for reference bias). In all plots, estimates are grouped according to island population and time window. Madeiran archipelago populations: M = Madeira, PS = Porto Santo; Selvagens archipelago populations: SG = Selvagem Grande; Canary Island populations: EH = El Hierro, GOM = La Gomera, TF = Tenerife, GC = Gran Canaria, FV = Fuerteventura, LZ = Lanzarote, GRA = La Graciosa.

Chapter 5

General discussion

A Berthelot's pipit ready for release after sampling

5.1. Thesis summary

This thesis highlights the importance and potential of examining patterns of genetic diversity and differentiation among populations across both spatial and temporal dimensions. The overall objective was to better understand the underlying evolutionary processes, and their interactions, that shape variation within relatively small and isolated populations. To do this, I employed the island endemic bird, Berthelot's pipit (*Anthus berthelotii*), as a model system, generating genomic datasets for populations across this species' range to disentangle the interplay between neutral and adaptive processes. This approach – a spatio-temporal genomic assessment – provides powerful insights into the evolutionary dynamics of such fragmented wild populations in variable environments, especially relevant given the contemporary environmental changes associated with anthropogenic pressures worldwide (Pelletier & Coltman, 2018).

In brief, the key findings of this thesis are as follows: Chapters 2 and 3 identified potential genetic adaptations to pathogen infection among Berthelot's pipit populations by focusing on the impact of spatially variable pathogen pressures (intensity and type) on contemporary genomic variation. Diverse genes, beyond those classically thought of as immune genes, were implicated in host protection and adaptation to both avian malaria and pox. Chapter 4 provided a whole-genome analysis across populations and over the 20th century, revealing shared patterns of genetic variation within the Canary Islands and an overall loss of genetic diversity over time across all populations. These patterns may reflect both historical bottlenecks and more recent population reductions, potentially linked with anthropogenic effects. While spatial analyses highlight the lasting effects of founder events, the temporal data captures the possibility of more recent reductions that could be due to newer environmental challenges. While Chapter 3 provided some discussion on the relationship between the different pathogens investigated in Chapters 2 and 3, and the combined implications of these chapters regarding the exploration of host-pathogen interactions, this final chapter synthesises my findings across the entire thesis. Emphasis is placed on the insights gained from both spatial and temporal dimensions, and the combined impact of neutral and adaptive forces, with suggestions for future research directions.

5.2. Contemporary selection on relevant immunogenetic variation

Divergence between populations results from the gradual accumulation of genetic changes over time, leaving a record of their impact in the genome. Consistent environmental variations across populations (i.e., in the spatial dimension) provide an excellent framework for investigating adaptation to these environmental variables (Nosil *et al.*, 2008; Funk *et al.*, 2011; Bradbury *et al.*, 2013; Ericson *et al.*, 2021). In Berthelot's pipit, the nested arrangement of islands within archipelagos, along with distinct pathogen communities (Spurgin *et al.*, 2012), imposes different selection pressures on different populations. This consistent variability in pathogen prevalence—particularly elevated levels of avian malaria and pox—creates a spatial pattern that drives adaptation. By employing genotype-environment association analyses, I examined genomic variation across these populations, revealing allele frequency variations linked to pathogen-mediated selection gradients and identifying genomic regions that may be responsible for adaptation to these pathogen pressures. While historical bottlenecks have restricted genomic variation and adaptive potential (as identified in previous work and Chapter 4), my results provide evidence of the species' capacity to adapt to local pathogen pressures.

Notably, several genes implicated in pathogen resistance/tolerance, including those regulating inflammation, metabolism, stress response, and cell death, emerged as strong candidates. Although not all are classic immune genes, many interact with immune pathways and play crucial roles in signalling, highlighting the interconnected nature of genetic adaptation. Some candidates were previously linked to these pathogen responses in humans and other bird species (e.g., Fernandes *et al*., 2015; Videvall *et al*., 2015; Cassin-Sackett *et al*., 2019). This agnostic approach, moving beyond well-established immune genes, can further help to unravel the complexities of evolutionary processes and also help identify genes important in wildlife health.

Given the complexities of assessing individual infection statuses in wild populations, pathogen prevalence was used as a proxy for selection pressure. This method also carries challenges, as population prevalence is derived from these individual assessments that can be biased by factors such as differential exposure, diagnostic errors or sampling limitations –

e.g., the underrepresentation of severely infected individuals who are less likely to be sampled – potentially distorting prevalence estimates. Nevertheless, by aggregating individual data into population-level measures, and given that similar biases would affect each population, relative differences across populations still yield valuable insights. This analytical approach leverages the consistent pattern of pathogen pressure across populations, corroborated by temporal sampling (see Chapters 2 and 3), to assess selection pressures.

The primary methodology used in this thesis, Bayenv, identifies genes under pathogen-mediated selection by correlating SNP frequency with pathogen prevalence across populations. This approach operates on an assumption of parallel selection: that consistent pathogen pressures, such as those from malaria or pox, will drive convergent genetic responses across populations. Strong correlations between SNPs and pathogen prevalence thus serve as indicators of selection. While parallel selection offers a practical framework for detecting broad-scale adaptive patterns, it risks oversimplifying the complexities of pathogen-mediated adaptation across diverse environments. Our findings both complement and contrast with those of Armstrong *et al*. (2018, 2019), who investigated genetic associations with malaria infection risk in Berthelot's pipits across the Canary Islands and other archipelagos. In the 2018 study, Armstrong *et al*. found that population-level allele frequencies at the strongest candidate SNPs were not uniformly associated with malaria prevalence across all populations, highlighting the role of both founder effects and localised selection pressures in shaping distinct genetic responses. In a follow-up study, Armstrong *et al*. (2019) reinforced these findings, demonstrating that the same allele could exhibit opposing associations with malaria infection in different populations. Such findings align with other candidate gene studies, which highlight how local ecological differences and varying genetic backgrounds across populations can lead to distinct, population-specific responses to pathogen pressures (Bonneaud *et al*. 2006; Loiseau *et al*. 2011), thereby challenging the assumption of parallelism.

Our study in Chapter 3 builds on these insights by detecting SNPs associated with malaria prevalence across all archipelagos, including a candidate SNP near the gene *ZNF598*, which Armstrong *et al*. (2018) identified as significantly divergent within the Canary Islands. This

broader association suggests that certain adaptive signals may persist across populations. Notably, we did not identify Armstrong's most significant SNPs, perhaps indicating that broader-scale studies capture more nuanced allele frequency shifts that may be overlooked in single-population analyses.

While localised adaptations are evident, it is important to recognise that parallel and population-specific adaptations may not be mutually exclusive. For example, Alves *et al*. (2019) demonstrated that intense selection from myxomatosis led to parallel genetic changes in rabbit populations across Australia, France, and the UK, despite diverse environmental backgrounds and population structure. Similarly, Gignoux-Wolfsohn *et al*. (2021), identified SNPs in bat populations affected by white-nose syndrome that showed consistent selection across populations, despite some site-specific variation. In both studies, many of these candidate alleles were already present across populations prior to outbreaks, suggesting that adaptation involved standing genetic variation, which can promote parallel selection under common pressures. Such cases demonstrate how intense, directional pathogen selection can foster both parallel and unique genetic adaptations, creating a mosaic of shared and specific responses. By leveraging Bayenv, we can capture these broad-scale selection pressures, identifying genes and pathways consistently shaped by pathogen exposure and capturing subtle allele frequency shifts characteristic of polygenic traits. This is crucial, as such shifts – often involving many small genetic changes – might not be evident in isolated population analyses. Although an emphasis on parallel selection could lead to potential oversights, such as the risk of misinterpreting outlier variation as general trends, Bayenv's framework accounts for demographic covariance, mitigating some of these challenges (Gunther & Coop, 2013).

Ultimately, while parallel selection can simplify the diverse forces shaping pathogen resistance, it remains a useful model for identifying general evolutionary patterns. While it is crucial to interpret results within the context of both shared and population-specific dynamics, the search for core adaptive variants remains a key component in understanding how pathogens shape the evolutionary trajectories of natural populations.

5.3. Spatio-temporal patterns of past and 'recent' demography

The inclusion of temporal genomic data, collected from the same populations at different points in their history, in conjunction with spatial data, facilitates a more direct examination of changes over time, shedding light on the factors and processes contributing to divergence between populations. The initial colonisation-related bottlenecks consistently emerge as significant factors shaping genetic diversity across Berthelot's pipit populations (Spurgin *et al.*, 2014; Martin *et al.*, 2021, 2023), similar to other island species (Jensen *et al*. 2013; Sendell-Price *et al.*, 2021). Spatial perspectives can only really attribute the low contemporary variation observed across populations to these strong historical founder effects, subsequent drift, and the lack of gene flow between archipelagos. However, a spatio-temporal approach revealed potential additional demographic shifts that may correspond with either recent or pre-modern environmental disturbances, potentially anthropogenic (Chapter 4). Interestingly, despite the variation in population sizes across the Canary Islands (Spurgin *et al*., 2014; Illera, 2020), all populations experienced similar magnitudes of diversity loss. Theory predicts populations of different sizes would experience genetic drift at varying rates, with smaller populations losing genetic diversity more rapidly (Wright, 1931). However, the observed similarity in diversity decline suggests that gene flow within each archipelago may be sufficient to homogenise genetic variation across populations, counteracting size-related differences in drift. Other archipelagic birds in Macaronesia, such as trumpeter finches (*Bucanetes githagineus*) and common kestrels (*Falco tinnunculus*), show similar patterns of genetic differentiation and gene flow (Barrientos *et al.*, 2009; Kangas *et al.*, 2018), suggesting these processes may be shared across species. Additionally, the consistent loss of diversity across all archipelagos – even with limited inter-archipelago gene flow – points to the potential role of broader pressures that could be uniformly impacting these populations. Together, gene flow and external pressures might thus contribute to the observed consistency in diversity loss. Given these findings, I considered three potential explanations for the observed diversity loss: (1) recent environmental or anthropogenic pressures, (2) historical declines due to pre-modern disturbances, and (3) persistent genetic drift in small, isolated populations (see full explanation in Chapter 4 discussion above). While the exact drivers remain uncertain, recent population reductions in Berthelot's pipit and subsequent accelerated drift may underlie the

loss of diversity and increased differentiation between the Selvagens and other archipelagos. These temporal declines in diversity and increased contemporary differentiation may echo the impacts of anthropogenic fragmentation in the black-capped vireo (*Vireo atricapilla*) (Athrey *et al.*, 2012).

5.4. Future directions

The ultimate aim of my research was to employ a spatio-temporal approach to thoroughly investigate both neutral and adaptive evolutionary processes and evaluate the genetic health of Berthelot's pipit populations. However, the full realisation of this goal was somewhat limited by challenges associated with museomics. Analysing comparative historical-contemporary datasets presents technical and bioinformatic challenges, primarily due to variations in average read length, depth of coverage, and DNA damage levels, compounded by potential interactions between these factors and sample genome versus reference genome sequence divergence (Sefc *et al.*, 2007; Burrell *et al.*, 2015; Billerman & Walsh, 2019; Raxworthy & Smith, 2021). While Chapter 4 discussed efforts to identify and mitigate biases in the Berthelot's pipit data, it's clear that further efforts are needed to address these challenges, both for this species and across the broader field. Despite the rapid progress in museomics, including the development of bioinformatic pipelines tailored for such data (e.g., Irestedt *et al.*, 2022; Kutschera *et al.*, 2022), one of which we utilised, there remains no one-size-fits-all solution. In general, more exploratory analyses are required to assess the impact of historical DNA characteristics on variant calling in empirical data (see Parks & Lambert, 2015). Hopefully these can lead to the development of new bioinformatic tools to address systematic differences between historical and contemporary genomes.

Notably, the choice of reference genome and the mapping strategy in any genomics project can influence bioinformatic processing and downstream inferences (Li *et al.*, 2008; Li & Durbin, 2009; Ekblom & Wolf, 2014). Employing a species-specific reference genome would remove another source of uncertainty regarding my data before I focus on identifying any changes at specific sites or regions in the genome. Therefore, the availability of a high-quality genome assembly for the pipit – currently in progress as part of the B10K project at

the Evolutionary and Organismal Research Center in Zhejiang, China – would substantially help take this work forward. Although not directly involved in the project, our group provided DNA samples for the pipit, granting us access to the genome once completed. With this genome it would be possible to expand my analyses of genomic erosion, including the investigation of runs of homozygosity. This would enable further insights into recent population history and the extent, and causes and consequences, of inbreeding within Berthelot's pipit populations.

The integration of my findings from all chapters of this thesis, encompassing the identification of spatial variation linked to pathogen-mediated selection and identification of candidate genes (Chapters 2 and 3), alongside a comprehensive understanding of population history, structure and genome-wide changes in diversity (Chapter 4), provides a robust framework for examining temporal patterns of intraspecific pathogen-related variation. Reconstructing allele frequency changes over time at these specific candidate sites among Berthelot's pipit populations varying in pathogen prevalence, would further help to elucidate evolutionary responses to pathogens. Different patterns of allele frequency change at these candidate sites, compared to the neutral background average, in populations experiencing different degrees of pathogen prevalence would provide compelling evidence for pathogen-mediated selection at these sites over the past century. Similarly, we can extend this approach to other pre-identified candidate genes in this study system. These could include those identified through genome-wide association studies in earlier work (Armstrong *et al.*, 2018) and those identified in selection scans with links to craniofacial development, lipid metabolism, and carbohydrate metabolism (Armstrong *et al.*, 2018; Martin *et al.*, 2021), albeit with less clarity regarding the ultimate drivers of selection. Similar temporal comparisons using historical DNA have revealed adaptations to climate change in chipmunk species (*Tamias alpinus*) (Bi *et al.*, 2019), resistance development to pathogens and parasites like myxoma virus (*Leporipoxvirus*) in wild rabbits (*Oryctolagus cuniculus*) (Alves *et al.*, 2019) and ectoparasitic mites (*Varroa destructor*) in honey bees (*Apis mellifera*) (Mikheyev *et al.*, 2015), as well as variations in avian β-defensin genes (key immune genes) in the Seychelles warbler (*Acrocephalus sechellensis*) (Gilroy *et al.*, 2016). Accurately calculating allele frequencies for specific sites would necessitate sequencing many more Berthelot's pipit samples (see below). However, with a priori hypotheses about which

regions of the genome may play a key role in certain processes or in response to certain pathogens derived from the evidence in my chapters, detecting such patterns would be extremely powerful and informative. Such methods could also be used to identify more new regions under selection; however, this would again require a substantially increased sample size (see below).

Limited historical sample sizes may reduce the power to accurately assess individual regions, making it challenging to capture unique patterns of variation at specific sites. An alternative approach could involve collating gene sets predicted to undergo rapid changes over time and comparing their rate of change, as well as other signatures of selection, as a collective entity against the genome-wide average. This method allows for increased power by averaging changes over a larger number of sites, thereby somewhat mitigating the impact of individual variations or biases from specific sites. Immune genes would undoubtedly represent an interesting and important group to investigate given the presumed strength of pathogen-mediated selection (Spurgin *et al.*, 2011, 2012; González-Quevedo, Phillips, *et al.*, 2015; González-Quevedo, Spurgin, *et al.*, 2015; Armstrong *et al.*, 2018, 2019). Other gene families, such as those involved in sexual conflict, may also be interesting to investigate in a temporal context. Such genes, like those encoding seminal fluid proteins, are characterised by antagonistic coevolution between males and females and thus are predicted to exhibit rapid evolutionary changes (Sirot *et al.*, 2015). Given that each archipelago and island population may (if isolated from the others) undergo a unique evolutionary trajectory and arms race, rapid divergence at these loci among populations may be expected. This is something which, to my knowledge, has rarely (if ever) been investigated in natural vertebrate populations.

The spatio-temporal analyses are currently limited to relatively small sample sizes (2–12 individuals per population per time point). With continual reductions in sequencing costs, the feasibility of resequencing genomes for larger numbers of individuals is continually improving. My efforts to locate historical Berthelot's pipit samples have identified > 200 specimens preserved in museums worldwide, covering all of the main island populations and the largest islets, dating back to the 1870–1920s. However, museums are understandably cautious about allowing the sampling of many specimens due to the

destructive nature of the process, even for small toepad samples typically taken from bird skins. The successful generation of relatively high-quality sequencing data from these initial specimens outlined in this thesis, and the groundwork and logic laid for further exciting research questions, may help persuade museums to release more samples. If so, there is potential for researchers to expand this sampling in the future, particularly in larger populations where considerably more specimens are available, and thus resolve additional questions. Additionally, there are numerous contemporary Berthelot's pipit samples available (as utilised in Chapters 2 and 3 for calculating pathogen prevalence), although these have yet to be sequenced. While there is some evidence that relatively small sample sizes can provide accurate population parameters (Nazareno *et al.*, 2017; Li *et al.*, 2020), larger sample sizes enable far greater accuracy, especially improving inferences from approaches that utilise allele frequencies in demographic parameter estimation and genome-wide selection scans or genome-wide association studies. The benefits of sampling more individuals become even more pronounced especially at lower coverage levels (Han *et al.*, 2014; Lou *et al.*, 2021).

More recent sampling of Berthelot's pipit populations has been undertaken between 2016–2024 (during my PhD, previous studies and new ongoing studies), covering the majority of the islands studied thus far. This sample base presents a valuable opportunity not only to augment analyses with larger sample sizes, thereby enhancing statistical power and accuracy, but also to introduce an additional time point to the investigation. With this addition, it would be possible to assess finer changes occurring over *ca.* 15 years, encompassing *ca.* 4–7 generations depending on estimates of generation time (Garcia-Del-Rey & Cresswell, 2007; Bird *et al.*, 2020). By incorporating three distinct time points – spanning from the 1880s, through 2009, to 2024 – the temporal resolution of the study could be enhanced. This increased resolution would enable a more nuanced exploration of how genetic variation evolves over time, potentially revealing subtle yet significant patterns of divergence and adaptation. This would be a particularly powerful approach if researchers can identify any environmental changes that have occurred within individual populations during that time frame (e.g., changes in pathogen, key anthropogenic, or invasive species pressures) (Snead & Clark, 2022; Clark *et al.*, 2023).

In Chapters 2 and 3, I presented evidence linking variation at specific genes to the prevalence of avian pox and malaria across populations, highlighting the significant role of selective processes in shaping gene frequency in response to pathogen pressures. However, the correlative nature of these findings introduces inherent uncertainties regarding causality. Additionally, a correlation between the pathogens at both individual and population levels was observed in Chapters 2 and 3. Consequently, it is challenging to definitively attribute genetic shifts to selection by one of these pathogens alone, or to exclude selection by an unobserved co-correlated variable, underscoring the necessity for further validation. While experimental verification would provide the most definitive evidence for selection by pox and/or malaria in this context, the constraints of working with wild populations, especially those of non-model organisms like Berthelot's pipit, probably necessitate alternative approaches. An alternative avenue involves the intense and accurate monitoring of marked individuals from birth to death to assess fitness implications associated with known genetic variation, similar to that undertaken in systems like the Seychelles warbler (Davies *et al.*, 2021) or the Soay sheep (*Ovis aries*) population (Sparks *et al.*, 2019). However, the feasibility of such a resource-costly approach, especially in large, human-inhabited islands, presents many significant challenges, rendering it highly unlikely in the context of our study on Berthelot's pipit populations.

Another approach to resolve some of the key issues discussed above would be to strategically leverage the spatial and temporal variation already observed in the pipit system. For example, focusing on outlier populations like La Gomera and Tenerife, where infection rates by the two pathogens discussed above do not correlate and/or where the observed pressure changes over time (e.g., the small decline in pox prevalence on Porto Santo between 2006–2016). Undertaking association studies in those populations may offer a means to corroborate the role of the candidate genes I identified in relation to pox/malaria in Berthelot's pipit populations. However, conducting within-population studies based on individual-level assessments poses challenges, as discussed previously in the relevant chapters. Detailed phenotyping is necessary, and accurate disease status classification in wild populations is complicated by factors like differential exposure to pathogen infection and the presence of exposed but uninfected individuals. Large sample sizes would therefore need to be phenotyped and genotyped to increase statistical power and minimise false-

negative outcomes. Targeted genotyping of specific loci could offer a cost-effective approach to screen these genetic variants across hundreds to thousands of individuals, thereby improving the feasibility of direct association studies in this context (e.g., Micheletti *et al.*, 2018; Davies *et al.*, 2021).

To better represent pathogen pressure in Berthelot's pipit populations, future research could consider incorporating vector-based assessments, such as detecting parasite prevalence in local mosquito populations, as an additional or alternative measure. This approach offers a more direct ecological perspective on pathogen transmission, capturing variations in vector presence and abundance that host-based assessments may overlook. By reducing biases inherent in bird sampling – where individual exposure, health, and behaviour can influence observed prevalence – vector-based data could refine assumptions about pathogen pressure, thereby providing valuable insights into local infection risk and pathogen-driven genetic adaptations (e.g., Cornetti *et al*., 2018). However, this vector-based approach presents several significant challenges, particularly for accurately assessing selective pressures on the host and in the study of Berthelot's pipit and its pathogens, such as avian malaria and pox. While many mosquito species can carry avian *Plasmodium*, not all are competent vectors capable of effectively transmitting the parasite to birds (Santiago-Alarcon *et al*., 2012). Molecular detection of *Plasmodium* DNA in mosquitoes can confirm parasite presence but does not establish transmission capability (Valkiūnas, 2011). This distinction is crucial, as non-competent species might carry parasite DNA after feeding on infected birds without contributing to further transmission. In the case of Berthelot's pipit, the specific vectors for avian malaria are not fully identified, and avian pox has additional transmission routes beyond vectors (van Riper *et al.*, 2002). The mosquito fauna (Culicidae) in Macaronesia includes diverse genera like *Culex*, *Aedes*, and *Culiseta*, (Baez & Fernandez, 1980) each with varying potential to transmit different lineages of *Plasmodium* (Santiago-Alarcon *et al*., 2012). Without precise identification and knowledge of vector competence, detecting *Plasmodium* in vectors provides an incomplete picture of selective pressures on pipit populations. Additionally, some detected *Plasmodium* strains may not infect or significantly impact Berthelot's pipits. Thus, vector-based prevalence might overestimate the selective pressure exerted by pathogen strains that are not genuinely relevant to the fitness of the host. To accurately assess the infection potential, sporozoites in mosquito salivary

glands must be examined, vector competence assays conducted, and detailed field observations performed. Without these steps, prevalence estimates in mosquitoes risk being misleading, as they do not differentiate between infective and non-infective parasites. The study of avian pox further complicates vector-based assessments. Although arthropods, including mosquitoes, particularly of the genus *Culex*, can act as mechanical vectors, avian pox is not exclusively vector-borne; transmission can also occur via direct contact, aerosols, or contaminated food (van Riper *et al.*, 2002). Therefore, relying solely on vector-based data could miss critical pathways of transmission, leading to an incomplete understanding of selective pressure on hosts. A comprehensive approach, combining both bird-based and vector-based assessments, along with detailed studies of vector competence and ecological context, is recommended to fully understand the selective pressures acting on Berthelot's pipit and related species.

While this thesis has primarily examined avian pox and avian malaria as selective pressures across Berthelot's pipit populations, there remains a significant knowledge gap regarding other potential pathogens within this system. Efforts are now underway as part of a new PhD project to characterise the viral communities that infect this species using high-throughput sequencing (known as viromics), to understand their diversity, composition, and dynamics. Additionally, efforts are underway to investigate microbiome variation and its links to pathogenic infection across the landscape. These endeavours may uncover previously unknown pathogens within this system and shed light on their interactions with their host. Although this will introduce additional complexity, it promises to provide a deeper understanding of this host-pathogen system and the evolution of adaptation to multiple pathogens, both within and across populations of Berthelot's pipit.

5.5. Concluding remarks

This thesis sheds light on several fundamental aspects of how neutral and adaptive processes shape genetic diversity and divergence in island endemic populations. The enduring impacts of archipelago founding, drift, and subsequent isolation have heavily influenced the genetic composition of the studied Berthelot's pipit populations. Despite historical bottlenecks that have led to reduced genetic variation, my findings highlight the

species' capacity to adapt to local pathogen pressures, demonstrating how adaptive processes influence genetic variation in response to pathogen pressures across populations. Furthermore, while historical bottlenecks have led to diminished genetic diversity, recent declines in diversity and increased differentiation, particularly in isolated archipelagos, hint at more recent population declines and possible anthropogenic influences. These results suggest, though do not definitively confirm, that recent human-induced changes may have exacerbated the loss of diversity, adding complexity to the genetic landscape of fragmented populations. With the groundwork laid in this thesis, there is now a solid foundation for delving deeper into the exploration of adaptive variation across combined spatio-temporal scales in this species.

This thesis clearly shows that the Berthelot's pipit system serves as a valuable model for exploring micro-evolutionary processes across diverse wild populations over space and time, with insights relevant for understanding resilience and susceptibility to genetic erosion in fragmented, isolated populations. Although the species is currently not threatened, the findings underscore the importance of studying seemingly stable populations to better understand their dynamics in changing environments. Given that anthropogenic impacts across the planet are leading many species into such small, fragmented populations, understanding this phenomenon becomes crucial. Advancing our knowledge in this area will not only enhance our understanding of evolutionary processes but could also inform conservation strategies aimed at preserving genetic diversity and improving population persistence in the face of ongoing environmental challenges.

The insights from Berthelot's pipit populations benefited greatly from the replicated nature of multi-population data and historical DNA analysis. As whole-genome sequencing technology advances, the potential of historical DNA in population genomics will continue to grow, enhancing studies across extensive spatio-temporal scales. This approach offers a path to unravel evolutionary processes in wild populations, providing a crucial lens through which to view the effects of environmental and anthropogenic change, both historical and recent, on the natural world.

5.6. References

Alves, J. M., Carneiro, M., Cheng, J. Y., de Matos, A. L., Rahman, M. M., Loog, L., Campos, P. F., *et al.* (2019). Parallel adaptation of rabbit populations to myxoma virus. *Science*, *363*, 1319–1326.

Armstrong, C., Davies, R. G., González-Quevedo, C., Dunne, M., Spurgin, L. G., & Richardson, D. S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution*, *9*, 12482–12502.

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

Athrey, G., Barr, K. R., Lance, R. F., & Leberg, P. L. (2012). Birds in space and time: genetic changes accompanying anthropogenic habitat fragmentation in the endangered black-capped vireo (*Vireo atricapilla*). *Evolutionary Applications, 5*, 540–552.

Baez, M., & Fernandez, J. (1980). Notes on the mosquito fauna of the Canary Islands (Diptera: Culicidae). *Mosquito Systematics*, *12*, 349–55.

Barrientos, R., Kvist, L., Barbosa, A., Valera, F., López-Iborra, G. M., & Moreno, E. (2009). Colonization patterns and genetic structure of peripheral populations of the trumpeter finch (*Bucanetes githagineus*) from Northwest Africa, the Canary Islands and the Iberian Peninsula. *Journal of Biogeography*, *36*, 210–219.

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J. L., Nielsen, R., Moritz, C., *et al.* (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLoS Genetics*, *15*, e1008119.

Billerman, S. M., & Walsh, J. (2019). Historical DNA as a tool to address key questions in avian biology and evolution: A review of methods, challenges, applications, and future directions. *Molecular Ecology Resources*, *19*, 1115–1130.

Bird, J. P., Martin, R., Akçakaya, H. R., Gilroy, J., Burfield, I. J., Garnett, S. T., Symes, A., *et al.* (2020). Generation lengths of the world's birds and their implications for extinction risk. *Conservation Biology*, *34*, 1252–1261.

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

Bradbury, I. R., Hubert, S., Higgins, B., Bowman, S., Borza, T., Paterson, I. G., Snelgrove, P. V. R., *et al.* (2013). Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, *6*, 450–461.

Burrell, A. S., Disotell, T. R., & Bergey, C. M. (2015). The use of museum specimens with high-throughput DNA sequencers. *Journal of Human Evolution*, *79*, 35–44.

Cassin-Sackett, L., Callicrate, T. E., & Fleischer, R. C. (2019). Parallel evolution of gene classes, but not genes: Evidence from Hawai'ian honeycreeper populations exposed to avian malaria. *Molecular Ecology*, *28*, 568–583.

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., Sawkins, A., *et al.* (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources* (in press). Available at: https://doi.org/10.1111/1755-0998.13789.

Cornetti, L., Hilfker, D., Lemoine, M., & Tschirren, B. (2018). Small-scale spatial variation in infection risk shapes the evolution of a *Borrelia* resistance gene in wild rodents. *Molecular Ecology*, *27*, 3515–3524.

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

Ekblom, R., & Wolf, J. B. W. (2014). A field guide to whole-genome sequencing, assembly and annotation. *Evolutionary Applications*, *7*, 1026–1042.

Ericson, P. G. P., Irestedt, M., She, H., & Qu, Y. (2021). Genomic signatures of rapid adaptive divergence in a tropical montane species. *Biology Letters*, *17*, 20210089.

Fernandes, R. C., Hasan, M., Gupta, H., Geetha, K., Rai, P. S., Hande, M. H., D'Souza, S. C., *et al.* (2015). Host genetic variations in glutathione-S-transferases, superoxide dismutases and catalase genes influence susceptibility to malaria infection in an Indian population. *Molecular Genetics and Genomics*, *290*, 1155–1168.

Funk, D. J., Egan, S. P., & Nosil, P. (2011). Isolation by adaptation in *Neochlamisus* leaf beetles: Host-related selection promotes neutral genomic divergence. *Molecular Ecology*, *20*, 4671–4682.

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

Gignoux-Wolfsohn, S. A., Pinsky, M. L., Kerwin, K., Herzog, C., Hall, M., Bennett, A. B., Fefferman, N. H., *et al*. (2021). Genomic signatures of selection in bats surviving white-nose syndrome. *Molecular Ecology*, *30*, 5643–5657.

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

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

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, *195*, 205–220.

Han, E., Sinsheimer, J. S., & Novembre, J. (2014). Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular Biology and Evolution*, *31*, 723–735.

Illera, J. C. (2020). '*Anthus berthelotii*, Berthelot's pipit', in Keller, V., Herrando, S., Vorisek, P. *et al.* (eds) *European Breeding Birds Atlas 2: distribution, abundance and change*. Barcelona: European Bird Census Council & Lynx Edicions, 797.

Irestedt, M., Thörn, F., Müller, I. A., Jønsson, K. A., Ericson, P. G. P., & Blom, M. P. K. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Molecular Ecology Resources*, *22*, 2672–2684.

Jensen H, Moe R, Hagen IJ *et al*. (2013) Genetic variation and structure of house sparrow populations: is there an island effect? *Molecular Ecology*, *22*, 1792–1805.

Kangas, V. M., Carrillo, J., Debray, P., & Kvist, L. (2018). Bottlenecks, remoteness and admixture shape genetic variation in island populations of Atlantic and Mediterranean common kestrels *Falco tinnunculus*. *Journal of Avian Biology*, *49*, e01768.

Kutschera, V. E., Kierczak, M., van der Valk, T., von Seth, J., Dussex, N., Lord, E., Dehasque, M., *et al.* (2022). GenErode: a bioinformatics pipeline to investigate genome erosion in endangered and extinct species. *BMC Bioinformatics*, *23*, 228.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*, 1754–1760.

Li, H., Qu, W., Obrycki, J. J., Meng, L., Zhou, X., Chu, D., & Li, B. (2020). Optimizing sample size for population genomic study in a global invasive lady beetle, *Harmonia axyridis*. *Insects*, *11*, 290.

Li, H., Ruan, J., & Durbin, R. (2008). Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Research*, *18*, 1851–1858.

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 of London B*, *278*, 1264–1272.

Lou, R. N., Jacobs, A., Wilder, A. P., & Therkildsen, N. O. (2021). A beginner's guide to low-coverage whole genome sequencing for population genomics. *Molecular Ecology*, *30*, 5966–5993.

Martin, C. A., Armstrong, C., Illera, J. C., Emerson, B. C., Richardson, D. S., & Spurgin, L. G. (2021). Genomic variation, population history and within-archipelago adaptation between island bird populations. *Royal Society Open Science*, *8*, 201146.

Martin, C. A., Sheppard, E. C., Illera, J. C., Suh, A., Nadachowska-Brzyska, K., Spurgin, L. G., & Richardson, D. S. (2023). Runs of homozygosity reveal past bottlenecks and contemporary inbreeding across diverging island populations of a bird. *Molecular Ecology*, *32*, 1972–1989.

Micheletti, S. J., Hess, J. E., Zendt, J. S., & Narum, S. R. (2018). Selection at a genomic region of major effect is responsible for evolution of complex life histories in anadromous steelhead. *BMC Evolutionary Biology*, *18*, 140.

Mikheyev, A. S., Tin, M. M. Y., Arora, J., & Seeley, T. D. (2015). Museum samples reveal rapid evolution by wild honey bees exposed to a novel parasite. *Nature Communications, 6*, 7991.

Nazareno, A. G., Bemmels, J. B., Dick, C. W., & Lohmann, L. G. (2017). Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. *Molecular Ecology Resources*, *17*, 1136–1147.

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

Parks, M., & Lambert, D. (2015). Impacts of low coverage depths and post-mortem DNA damage on variant calling: A simulation study. *BMC Genomics*, *16*, 19.

Pelletier, F., & Coltman, D. W. (2018). Will human influences on evolutionary dynamics in the wild pervade the Anthropocene? *BMC Biology*, *16*, 7.

Raxworthy, C. J., & Smith, B. T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends in Ecology & Evolution*, *36*, 1049–1060.

Santiago-Alarcon, D., Palinauskas, V., & Schaefer, H. M. (2012). Diptera vectors of avian Haemosporidian parasites: untangling parasite life cycles and their taxonomy. *Biological Reviews*, *87*, 928–964.

Sefc, K. M., Payne, R. B., & Sorenson, M. D. (2007). Single base errors in PCR products from avian museum specimens and their effect on estimates of historical genetic diversity. *Conservation Genetics*, *8*, 879–884.

Sendell-Price, A. T., Ruegg, K. C., Robertson, B. C., & Clegg, S. M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology*, *30*, 2495–2510.

Sirot, L. K., Wong, A., Chapman, T., & Wolfner, M. F. (2015). Sexual conflict and seminal fluid proteins: a dynamic landscape of sexual interactions. *Cold Spring Harbor Perspectives in Biology*, *7*, a017533.

Snead, A. A., & Clark, R. D. (2022). The biological hierarchy, time, and temporal 'omics in evolutionary biology: a perspective. *Integrative and Comparative Biology*, *62*, 1872–1886.

Sparks, A. M., Watt, K., Sinclair, R., Pilkington, J. G., Pemberton, J. M., McNeilly, T. N., Nussey, D. H., *et al.* (2019). The genetic architecture of helminth-specific immune responses in a wild population of Soay sheep (*Ovis aries*). *PLoS Genetics*, *15*, e1008461.

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

Valkiūnas, G. (2011). Haemosporidian vector research: marriage of molecular and microscopical approaches is essential. *Molecular Ecology*, *20*, 3084–3086.

van Riper, C., van Riper, S. G., & Hansen, W. R. (2002). Epizootiology and effect of avian pox on Hawaiian forest birds. *The Auk*, *119*, 929–942.

Videvall, E., Cornwallis, C. K., Palinauskas, V., Valkiunas, G., & Hellgren, O. (2015). The avian transcriptome response to malaria infection. *Molecular Biology and Evolution*, *32*, 1255–1267.

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