Genomic insights into population history, drift and adaptation in the island endemic Berthelot's pipit

Claudia Anne Martin

Thesis submitted for the degree of Doctor of Philosophy
School of Biological Sciences
University of East Anglia, UK
November 2021

Abstract

Genomic insights into population history, drift and adaptation in the island endemic Berthelot's pipit. *DPhil thesis by Claudia Martin, submitted November 2021*

This thesis aims to use genomic techniques to elucidate how drift and selection shape genetic diversity across spatial and temporal scales, and thus develop our understanding of their role in incipient speciation. To accomplish this, I focus on Berthelot's pipit (*Anthus berthelotii*), an island endemic species, to gain insight into population history, drift and adaptation across this species' fragmented range. First, using RAD-seq I show that genome-wide divergence across the species range is largely shaped by initial colonisation and resulting bottlenecks, with limited evidence of subsequent gene flow between populations. Then, using a genome scan approach with this RAD-seq dataset, I identify loci putatively under differential selection within archipelagos, including a locus potentially involved in craniofacial development. I then use whole genome sequences to understand how colonisation events, associated bottlenecks, gene flow and genetic drift shape contemporary patterns of genetic diversity across populations. I show that there was a substantial loss of genetic diversity across the genome as a result of the initial island colonisation event by the ancestor of the Berthelot's pipit and its sister species the tawny pipit (*Anthus campestris*) *ca* 2.1 million years ago. These results show that population history, especially founder effects, can have a long-term influence on genome-wide genetic diversity, and that small contemporary *Ne* can result in signatures of severe inbreeding. Lastly, I investigate genomic landscapes of divergence through speciation from the tawny pipit to Berthelot's pipit, and across the three archipelago populations of Berthelot's pipit. Genome-wide divergence correlated with estimated colonisation timescales, with a few strongly divergent 'genomic islands' identified in each comparison. I investigate putative drivers of divergence across archipelagos, and find that selection interacts with founder effects and inbreeding to shape adaptation across these populations. Taken together, these findings suggest that evolution at genes involved with bill/body size, immune response, eye development and metabolism acts repeatedly to drive local adaptation across spatial and temporal scales. Collectively, this thesis furthers understanding of how different evolutionary mechanisms shape patterns of genetic diversity and divergence following the establishment of new populations, and how this may lead to eventual speciation.

Funding

My studentship was funded by the UKRI National Environmental Research Council (NERC) with additional contributions from the Institute of Natural Products and Agrobiology (IPNA-CSIC) in Tenerife. I am grateful to NERC and EnvEast Doctoral Training Partnership for their considerate approach during the Covid-19 pandemic, affording me the best opportunity to conduct the aims of my project, under the circumstances. My fieldwork was made possible by a Heredity fieldwork grant from the Genetics Society, who I also thank for their generous funding allowing me to attend training courses and conferences to present my work.

Acknowledgements

Many of the highlights of my PhD have come through opportunities to collaborate and this thesis would not have been possible without support and guidance from many people. First, I would like to thank Lewis Spurgin. I could not have asked for a more supportive and inspiring supervisor. Thank you for always having an open door, for sharing your bioinformatics skills and knowledge of population genetics, and taking the time to allow me opportunities to develop myself. Second, I would like to thank David Richardson. You have helped me to rationalise my thoughts when in the depths of coding trouble and provided comfort during times of isolation through the pandemic. I have fond memories of us racing to get out of the car on Lanzarote to be the first to catch a pipit!

I would like to thank Brent Emerson, my CASE supervisor from IPNA in Tenerife, for engaging conversations about island biodiversity and evolution, and Alex Suh, for taking on a stressed final year PhD, and welcoming me into your lab group. Alex, I look forward to our next work together. Thank you to everyone in the Organisms and Environment Group at the University of East Anglia, especially members of the Richardson and Spurgin groups, and my fellow ‘Cabbages’. Thank you for lunchtime cycles, ringing sessions, tea and cake, dog walks, badminton and listening to my permit woes. In particular I would like to thank, Cat Morrison; I loved evenings swimming together in Whittingham lake, Claire Armstrong for providing patient bioinformatics guidance in the early days of my PhD, Lorna Blackmore for many evenings chatting away in the office, Sarah Barnsley, Kirsty Franklin and Nel (Eleanor) Sheppard for your friendship. My personal thanks go to Sarah Yeates for always making me feel welcome to come and chat and for helping to get the PGR committee off the ground, and Bill Sturges, who was the director of EnvEast during my PhD, for his enthusiasm and kindness, including support in organisation of E3i Enterprise & Innovation club events.

I would like to thank Juan Carlos Illera for persevering to get us permits to sample pipits across the island populations and for his interest in my work. Juan Carlos and Lewis, and Claire collected field samples (2006 & 2009, respectively), which are used throughout this thesis and Claire also conducted the RAD-seq lab work. Fieldwork would not have been nearly as rewarding without the company of Nel and Johnny DeCoriolis – thank you for sharing the adventures with me and being such persistent pipit herders! My thanks also go to Beth, Tim, Joe, Elana and Jordy for their assistance during sample collection across Lanzarote and La Graciosa. I would like to thank Maria-Elena Mannarelli for her guidance in the molecular lab as well as her company on wellie boot walks which helped to keep spirts

up during the Covid-19 Winter of 2020. For insights into PSMC analysis, I'd like to thank Krystyna Nadachowska-Brzyska, and Jo Howard-McCombe, for many enjoyable afternoons of cake-fuelled genomics chatter. Finally, my thanks also go to the high-performance computing team at UEA, especially Jacob Newman, and Camilla Ryan at Earlham Institute for their bioinformatics advice, as well as Cock Van Oosterhout for his support.

My time in Norwich certainly would not have been the same without the friendship of my housemates James Christie, Bridie Davies, Greg Rix, Sam Prudence and Claire Buchan, who have shared this journey with me from the beginning, as well as Emma Langan and Callum Rollo who joined the Speedwell gang later. Going through this journey together has been of great comfort and joy, and I can't wait to see what awesome things you all end up doing.

Next I would like to thank my friends (not yet mentioned) and family, who's support and company has brought me immense joy, and grounded me when I have needed it the most. Especially to Miranda, Beth, Andy, Alex K, Alex P, Liz, Vic, Sarah G, Katie and Aimee. Mum and Dad, I am so grateful for all the years you sat by my side reading through homework or the 300 page dyslexia book "*Toe by Toe*". It was worth it. Thank you both for teaching me one of the best lessons, that **trying your hardest is good enough**. I wish my primary school headmistress could see me now, she would not believe I achieved GCSE's! Astrid, thank you for all the years chasing each other around the garden.

I never imagined that I would be undertaking my PhD through the pandemic events that have unfolded over the last 18 months, with a hectic return from fieldwork to my parents spare room. While this has provided huge mental and academic challenges, there are so many things I am grateful for; not least being able to spend lockdowns (1, 2 and 3!) surrounded by close family, Tim and making a special friend in Matt Royale (thanks for all the yoga, badminton, runs, baking, chats over a tea and a 9 month loan of your library card!). I would like to thank the Armitage's and Matt for providing a safe haven during times of uncertainty.

Last, I would like to thank my partner, Tim Armitage, whose companionship and support throughout this journey has bought me great strength; you have listened to my PhD woes, encouraged me to follow my dreams and shared in the excitements. Sharing our PhD journeys (albeit from the extremes of 200 miles away or the same lockdown desk!) has been a sense of comfort and added adventure to the last 4 years. Who would have thought you would be such a good pipit herder? Hears to many more years of adventures in the mountains, cycling through hidden valleys and jumping in lakes, wherever it might take us!

Author contributions

The following manuscripts have arisen from this thesis, and are presented in Chapters 2, 3 and 4. These are collaborative efforts, although in each case the majority of work is my own. Below I include a full citation for each chapter, and highlight my contribution.

Chapter 2: Genomic variation, population history and within-archipelago adaptation between island bird populations. Published in *Royal Society Open Science.* **8**: 201146.

> Claudia A. Martin led and performed majority of the data analysis and drafted the manuscript.
>
> L.G.S. and D.S.R. designed the research and obtained the funding, and with B.C.E., supervised the project; L.G.S. and J.C.I. (and to a lesser extent, D.S.R.) undertook the field sampling; C.A. performed the ddRAD sequencing and bioinformatics while affiliated with NBAF Sheffield; L.G.S. assisted with data analysis, with input from C.A., and assisted with writing the manuscript.

Chapter 3: Genomic signatures of historic bottlenecks and inbreeding in an island colonising bird. *In preparation.*

> Claudia A. Martin conceived the project, performed DNA extractions, bioinformatics, genomic analyses and drafted the manuscript, all with input from David S. Richardson and Lewis G. Spurgin.
>
> L.G.S. and D.S.R. obtained the funding, and with B.C.E. and A.S., supervised the project; L.G.S. and J.C.I. undertook the field sampling; L.G.S. assisted with data analysis, with PSMC guidance from K. N.B.. Input on the manuscript was obtained from D.S.R., L.G.S. and A.S..

Chapter 4: The genomic landscape of divergence across bottlenecked populations of an island colonising bird. *In preparation.*

> Claudia A. Martin conceived the project, performed DNA extractions, bioinformatics, genomic analyses and drafted the manuscript, all with input from David S. Richardson and Lewis G. Spurgin.
>
> L.G.S. and D.S.R. obtained the funding, and with B.C.E. and A.S., supervised the project; L.G.S. and J.C.I. undertook the field sampling; L.G.S. assisted with data analysis. Input on the manuscript was obtained from L.G.S. and D.S.R..

Contents

List of tables and figures

In order of appearance.

Figure 2. Evolutionary relationships between island populatio4ns of the Berthelot's pipit. **A)** Maximum-likelihood bifurcating tree of population history - without subsequent gene flow - across the pipit colonisation range as inferred by *TreeMix*. **B)** The relationship between linkage disequilibrium and base-pair distance for SNPs across each Madeiran Island, three Canary Island populations and the Selvagens.

Figure 3. Population structure based on genome-wide ddRAD SNPs among Berthelot's pipit populations separately across the Canary Islands and Madeiran archipelago. **A)** Principal component analysis (PCA) across the Canary Island populations. **B)** PCA of Madeiran archipelago populations.

Figure 4. Manhattan plots of EigenGWAS analyses based on genome-wide ddRAD SNPs among Berthelot's pipit populations within archipelagos. **A)** Canary Islands PC1. **B)** Madeiran archipelago PC1.

Figure 5. Within-archipelago genetic differentiation of genome-wide ddRAD SNPs among Berthelot's pipit populations.

Chapter 3. Genomic inference of colonisation timescales, associated bottlenecks and contemporary inbreeding across island bird archipelagos

Figure 1. Berthelot's pipit range across three archipelagos and the sampling location of its sister species, the tawny pipit.

Table 1. Berthelot's pipit and tawny pipit sampling information and genome sequencing for variants mapped to the Zebra finch genome.

Table 2. Genome-wide genetic diversity per individual across Berthelot's pipit populations.

Figure 2. A) Principal component analysis (PCA) based on the "All Pipits" genomic autosomal SNPs among six Berthelot's pipit populations and one tawny pipit; and **B)** "Berthelot's" PCA between Berthelot's pipit island populations across the three colonised north Atlantic archipelgos.

Table 3. Pairwise F_{ST} between island populations of Berthelot's pipit with differing levels of founding events. Divergence times are estimated from shared ancestry based on PSMC effective population sizes, N_e.

Figure 3. PSMC modelled divergence between the mainland tawny pipit and the Berthelot's pipit. Estimates of effective population size, N_e, through time for six Berthelot's pipit populations and the tawny pipit genome.

Figure 4. PSMC estimates of changes in effective population size, N_e, over time for six contemporary Berthelot's pipit populations across three archipelagos in the north Atlantic, from 2 Mya until 10 kya.

Table 4. Long ROH (>1 Mb) across populations of Berthelot's pipit. Archipelago populations are separated by grey dotted lines.

Figure 5. Genome-wide patterns of nucleotide diversity (π), inbreeding and ROH across island populations of Berthelot's pipit and its mainland sister species, the tawny pipit. **A)** Individual inbreeding coefficient, F_{IS}, across 6 island populations of Berthelot's pipit and the tawny pipit. **B)** Nucleotide diversity across the genome of the tawny pipit, and four Berthelot's pipit individuals with low, moderate and high levels of inbreeding using 2 Mb windows with 20% overlap.

Figure 6. Size classes of genome-wide runs of homozygosity (ROH) for the tawny pipit and populations of Berthelot's pipit. **A)** The relationship between the number of ROH detected covering a chomosome segement at least 250 kb in size and the cumulative length of the genome in ROH. **B)** Number of ROH detected in individual genomes between 250 kb and 1 Mb relative to the ROH of at least 1 Mb in length.

Figure 7. Nucleotide diversity across Chromosome 1 in four Berthelot's pipits. Nucleotide diversity was measured in 250 kb windows with 20% step across the entire chromosome.

Chapter 4. Genomic landscapes of divergence in an island archipelago bird: identifying the role of drift and selection

Figure 1. Berthelot's pipit range across three Macaronesian archipelagos and the sampling location for the tawny pipit.

Table 1. Genomic filtering for divergence analyses across populations of the Berthelot's pipit and mainland tawny pipit.

Table 2. Sampling information of divergence comparisons across populations of the Berthelot's pipit and mainland tawny pipit.

Figure 2. Distribution of pairwise genomic differentiation, F_{ST}, across Berthelot's and tawny pipit comparisons calculated in 50 kb autosomal windows.

Figure 3. Correlated patterns of F_{ST} between all Berthelot's pipit population comparisons, and between the Berthelot's pipit and tawny pipit.

Table 3. Genomic differentiation, F_{ST}, between the Berthelot's pipit and tawny pipit, and among Berthelot's pipit archipelagos.

Figure 4. Pairwise genomic differentiation, F_{ST}, across the genome for the Berthelot's and tawny pipit comparisons calculated in non-overlapping 50 kb windows.

Figure 5. Patterns of divergence, genetic diversity and loss of rare alleles surrounding genomic islands of divergence between the tawny pipit and Canary Islands Berthelot's pipit comparison.

Figure 6. Consistent patterns of divergence between the Berthelot's pipit and tawny pipit, and across the Berthelot's pipit range for chromosome 1A ~2 Mb genomic island of divergence.

Table 4. Candidate genes within genomic islands of divergence between the tawny pipit and the first colonised Berthelot's pipit archipelago (the Canary Islands) ~ 2.1 Mya. Genes under putative positive selection in the Canary Islands are identified, due to corresponding dips in nucleotide diversity and Tajima's D when compared to the tawny pipit.

Table 5. Genomic islands of divergence through archipelago colonisation in the Berthelot's pipit.

Table 6. Genomic valleys of similarity through speciation between the Berthelot's and tawny pipit.

1

General introduction

Sampling pipits on foot in La Graciosa, Canary Islands. Photograph by Beth Fiducia-Brookes

1 Evolutionary drivers of differentiation and speciation

Divergence and speciation are dynamic processes, shaped by the interactions of many evolutionary mechanisms (Lande 1980; Kirkpatrick and Ravigné 2002; Ravinet *et al.* 2017). The evolutionary dynamics of natural populations involve complex interactions among genes and the alleles that exist at those genes, and are influenced by forces that change these allele frequencies: mutation, gene flow, genetic drift and selection (Wright 1932). These evolutionary forces are the basis for genetic and phenotypic evolution, which results in divergence among populations. Speciation may then follow either in the presence, or absence, of geographical isolation and gene flow between diverging populations (Slatkin 1987; Papadopulos *et al.* 2014; Tigano and Friesen 2016; Cowles and Uy 2019). Studying divergence across diverse taxa, between closely-related species and below the species level, allows us to address some fundamental evolutionary questions: what processes shape genetic diversity; how does this variation allow species to adapt to different environments and selection pressures; and what evolutionary processes promote the formation of new species? Such questions are of particular importance in a time when we are observing unprecedented loss of global biodiversity and rapidly changing climate.

Selection acts on genetic variation that ultimately arises via mutation (Muller 1950; Nei 1983; Burke, 2012). Positive divergent selection may occur between geographically separated populations experiencing different selective pressures, and where alternative alleles are favoured in each population, this results in genetic diversification (Peláez *et al.* 2020). The classic model of positive selection assumes newly arisen mutations rapidly approach fixation in one lineage, and are *swept* to fixation. This occurs through two mechanisms: ‘hard selective sweeps’ occur when a single haplotype harbouring a selectively advantageous allele rises in frequency, whereas during a ‘soft sweep’, multiple haplotypes harbouring advantageous mutations can rise in frequency simultaneously (Burke 2012; Hermisson and Pennings 2017). In many natural systems this process is more gradual, with moderate changes in genetic variation at many loci, each with smaller but cumulative effects on trait expression and evolution (Tennessen and Akey 2011). Indeed, detecting and characterising this type of selection requires extensive genomic data across populations with diverse phenotypes for the trait. In a similar way, strongly deleterious variation is rapidly removed by purifying selection. Where geographically separated populations are adapting to the same selective pressure (i.e., ‘parallel selection’), divergence may occur by mutation-order speciation, in which different, incompatible mutations are fixed between populations (Mani and Clarke 1990). Such divergent selection can result in reproductive barriers due to genetic

and potentially morphological incompatibilities that can form between populations (termed 'barrier loci') and have been assumed to play an important role in some speciation events (Rice 1987; Elmer 2019). It is also possible for loci of importance to be conserved between diverging populations over long evolutionary timescales, termed 'balancing selection'. Balancing selection may be associated with increased genetic diversity within populations but localised reduced differentiation between populations, due to active maintenance of variation at a locus (see for example Hohenlohe *et al.* 2010).

Divergence between populations, and even speciation, does not need to involve selection. The 'nearly neutral' theory of molecular evolution holds that the majority of evolutionary change at the molecular level is due to mutation and random genetic drift (stochastic allele fixation or loss), influenced by gene flow and population size (Wright 1931; Kimura 1968; Kimura 1991). Genetic drift alone can be sufficient to create large variation in allele frequencies between populations, which may - especially where gene flow is limited - result in genetic structure between populations (Slatkin 1987).

Normally genetic drift is gradual, however it can be exacerbated by strong demographic fluctuations, such as through founder effects and population bottlenecks (see for example Hoeck *et al.* 2010). Such effects have been shown to affect both putatively neutral and adaptive genetic variation (Leberg 1992; Bollmer *et al.* 2011; Sutton *et al.* 2011). The genetic variation within populations founded by few individuals may not be representative of the variation within the source population, and can result in new, genetically distinct populations. Similarly, population bottlenecks can reduce the genetic diversity of a population (Mayr 1965; Frankham 1996). It has long been thought that founder effects accelerate the speciation process (Lande 1976), especially where sequential colonisation events occur (Charlesworth and Smith 1982) and populations are small in geographic range and number (Barton and Charlesworth 1984). Where population size recovery is slow, the impact of drift will increase (Nei *et al.* 1975; Maruyama and Fuerst 1985) and in extreme cases this may lead to inbreeding and inbreeding depression (Kirkpatrick and Jarne 2000). On the other hand, adaptive potential and genetic diversity may be regenerated following population founding through immigration accompanied by gene flow or mutation (Dlugosch and Parker 2008). Thus, while the genetic diversity of a population is initially shaped by the founder effect, this signature may be enhanced or eroded over time.

As neutral genetic variation is not directly influenced by natural selection, it is valuable for inferring past and present population and individual level evolutionary processes. At the population level, analysing neutral or genome-wide variation can reveal patterns of

population structure (e.g., Shannon *et al.* 2015; du Plessis *et al.* 2019), changes in population size such as bottlenecks (e.g., Nadachowska-Brzyska *et al.* 2016; Patton *et al.* 2019), relatedness and inbreeding (e.g., Brzeski *et al.* 2014; Küpper *et al.* 2016; Hooper *et al.* 2020), hybridisation and admixture (e.g., Nadeau *et al.* 2014; Macleod *et al.* 2015) and speciation (e.g., Marques *et al.* 2016; Turbek *et al.* 2021). Gaining an understanding of population dynamics in this way can provide an evolutionary 'baseline' and help to disentangle the relative roles of neutral and selective processes shaping genetic diversity.

Selection and drift interact to promote divergence at a range of geographical and temporal scales within and across populations (Losos and Ricklefs 2009). Very fine landscape-scale adaptation within populations may act very rapidly due to strong ecological selection pressures (Delahaie *et al.* 2017). Comparisons of natural populations exposed to a selective gradient have demonstrated genetic and morphological divergence across a very small spatial scale including the Mascarene grey white-eye (*Zosterops borbonicus;* Milá *et al.* 2010; Bertrand *et al.* 2016) and 'ohi'a lehua, a Hawaiian endemic tree (*Metrosideros polymorpha;* Izuno *et al.* 2017). Fine-scale divergence and population structure may also result from the effects of drift due to limited dispersal, population size or range, and population demography. This has been observed across a wide-range of natural populations including fragmented and restricted ranges of the black toad (*Bufo exsul*; Wang 2009), bottlenose dolphin (*Tursiops truncates*; Mirimin *et al.* 2011), paper wasp (genus *Polistes*; Bluher *et al.* 2020) and eastern redbud (*Cercis canadenis L.*; Ony *et al.* 2020).

Some species are known to exhibit exceptionally rapid evolution with repeated trait divergence (Nevado *et al.* 2019), including African cichlid fish (family Cichlidae; Seehausen 2006), new world lupins (genus *Lupinus;* Nevado *et al.* 2016) and white-eye birds (genus *Zosterops*; Cornetti *et al.* 2015), which have become model systems for studying adaptive radiations and diversification. Exposure to strong selection pressures for short time periods may drive genetic and morphological divergence, as is observed through fluctuating climate extremes in Darwin's finches (subfamily Geospizinae; Boag and Grant 1981). As well as selection, rapid evolution may be driven by dramatic changes in population demography, and is very common as a result of dispersal, for example in European grey wolves (*Canis lupus*; Pilot *et al.* 2014), African buffalo (*Syncerus caffer*; de Jager *et al.* 2021) and Leach's storm-petrel (*Oceanodroma leucorhoa*; Bicknell *et al.* 2012).

Evolution can also be studied at broad spatial scales, among populations between which there has been long-term isolation and limited gene flow (Ellegren *et al.* 2012; Manel *et al.* 2012). Typically divergence between such populations accumulates rapidly (for review see

Ravinet *et al.* 2017). However, this is not always the case: for example, limited population structure or differentiation was observed over vast geographic scales across populations of Lake Erie watersnake (*Nerodia sipedon insularum*), and morphological change was only observed over exceptionally long temporal scales (Ray and King 2006). Balancing selection can also lead to exceptionally constant allele frequencies over time and space (Charlesworth *et al.* 1997; Novembre and Di Rienzo 2009), although these same processes can lead to genomic divergence between populations (Cruickshank and Hahn 2014). Studies across a range of spatial and temporal scales, with known demographic and selective history, may be especially powerful to gain an in-depth understanding of how evolutionary forces shape variation within and among populations.

2 Genomic landscapes of diversity and divergence

The genomes of individuals from diverging populations do not change uniformly – instead, some regions diverge rapidly, while others diverge more slowly or are conserved over long time periods (Nosil *et al.* 2009; Nosil and Feder 2012; Fig. 1). Loci under divergent selection - and those tightly linked to them (Maynard-Smith and Haigh 1974) - form 'genomic islands of divergence' due to selection operating over a small fraction of the genome (Feder *et al.* 2013; Seehausen *et al.* 2014). 'Genomic valleys of divergence' may also form where there is little or no differentiation between populations, which may result from simultaneous balancing selection within both populations. Recombination shapes genomic architecture, acting to break apart chromosomal segments and reduce linkage disequilibrium between loci. However, genomes are not homogenous entities and genomic features (e.g., centromeres) can contribute to reduced recombination rates. As a result, patterns of diversity and divergence may be shaped differently across the genome (Ravinet *et al.* 2017). Variation in divergence across the genome can be quantified using Wright's fixation index, F_{ST} (0 = identical allele frequencies between populations, 1 = populations fixed for different alleles) (Wright 1931) estimated across loci. Genomic regions of high F_{ST} may represent loci that are divergent as a result of selection, since drift is expected to affect the entire genome whereas selection acts locally (Schneider *et al.* 2021). Using this and similar approaches, highly heterogeneous genomic landscapes, including genomic islands and valleys of divergence, have been identified in a range of taxa, including the island endemic tree 'ohi'a lehua (*Metrosideros polymorpha*; Choi *et al.* 2020), Swainson's thrush (*Catharus ustulatus*; Ruegg *et al.* 2014), *Heliconius* butterflies (Nadeau *et al.* 2012) and the threespine stickleback (*Gasterosteus aculeatus*; Jones *et al.* 2012).

Evolutionary forces other than natural selection may also result in heterogeneous divergence across the genome (Feder *et al.* 2013). For example, strong founder bottlenecks and persistence of small population size may result in large regions of low genetic diversity which are strongly divergent between the two populations (Pilot *et al.* 2014; Sendell-Price *et al.* 2021). Alternatively, such patterns may be caused by strong, recent purifying selection, resulting in persistence of only one haplotype within the population. Since the genomic landscape is also shaped by recombination, demographic processes that alter how recombination acts across the genome also affect patterns of diversity and divergence (Ravinet *et al.* 2017). For example, small population size and low genetic diversity, including inbreeding, will result in high linkage disequilibrium due to inefficient recombination. Similarly, limited gene flow between or within populations can reduce effective recombination rates. As a result, neutral loci linked to a selected locus may be favoured through background selection (Cvijovic *et al.* 2018), which can result in broad peaks of divergence surrounding a selected locus. Thus, our ability to detect genomic islands and valleys of divergence is dependent on the number of loci involved in the adaptive trait under selection, the rate and timing of selection, and demographic processes, as drift, mutation and recombination erode these signals over time (Via and West 2008). Studying patterns of genetic diversity may reveal evolutionary processes driving the divergence landscape.

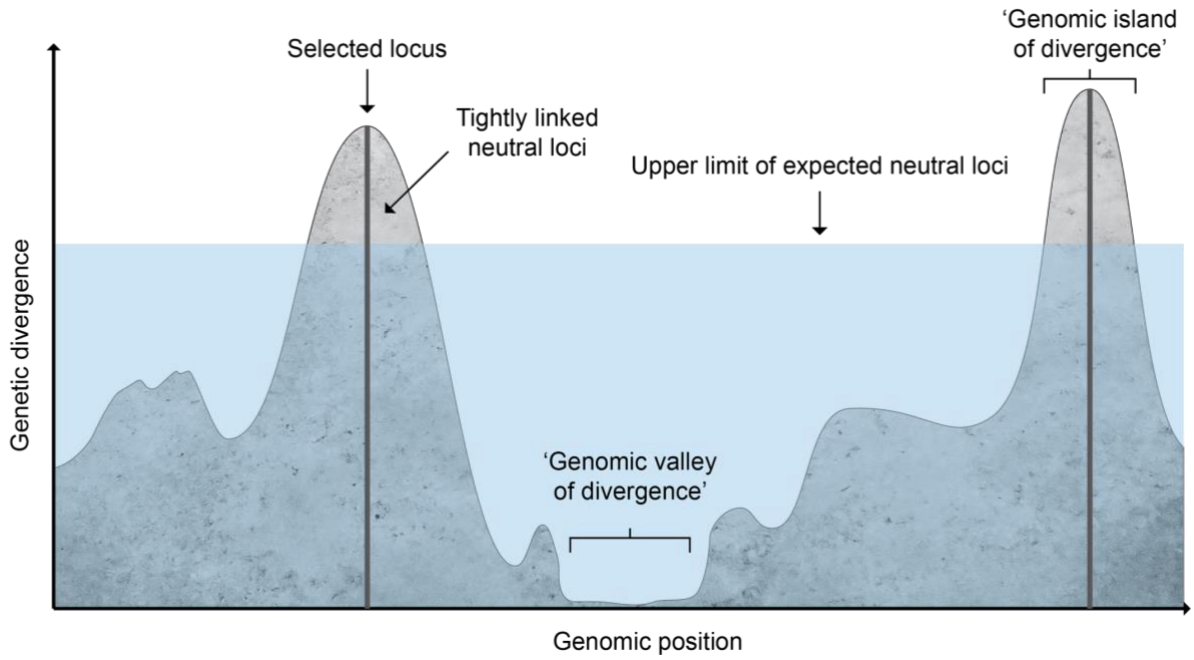


Figure 1. Schematic representation of a heterogenous genomic landscape of divergence showing two divergence peaks and a strongly conserved valley. Adapted from Nosil and Feder (2012).

Prior to the introduction of high-throughput sequencing, approaches to studying genetic variation in natural populations included using allozymes (enzyme polymorphisms) (Garten, 1976; Nevo, 1978; Oostermeijer *et al.*, 1995), microsatellite markers (simple sequence repeats, SSRs) (Bruford and Wayne, 1993; Maroof *et al.* 1994), small-scale SNP genotyping (Brumfield *et al.,* 2003) and amplified fragment length polymorphisms (AFLPs) (Vos *et al.,* 1995; Bensch *et al.,* 2002; Baxter *et al.,* 2008; Papa *et al.,* 2013). These techniques are still valid for studying population history and dynamics (Brumfield *et al.,* 2003; Kardos *et al.,* 2017), population structure and migration (Gaudeul *et al.* 2004; Lander *et al.* 2021), and individual patterns of inbreeding and heterozygosity (Goossens *et al.* 2016). Furthermore, these methods can go some way to identifying potential signatures of selection, or at least sites linked to those loci under selection (Ross, 1997; Baxter *et al.,* 2008). However, a fundamental issue with all of these approaches is that sampling is limited to a tiny fraction of the genome. Consequently, estimates of genome-wide variation, and thus inferences about neutral processes such as inbreeding or demography, may not be accurate (Luikart *et al.,* 2003; Reed and Frankham, 2003; Slate *et al.,* 2004). Furthermore, with low genome coverage, the power to identify genomic markers linked to genes under selection is very low (Britten, 1996; Hansson and Westerberg, 2002; Chapman *et al.* 2009; Nunes *et al.,* 2012).

High-throughput or ‘next-generation sequencing’ (NGS), has dramatically improved the power with which we can study the evolutionary processes shaping diversity and speciation in populations, making it possible to quantify diversity across high-density mapped genetic markers. Targeted regions, subsets of the genome, or whole-genome sequences are amplified and sequenced, reads can then be mapped to a reference genome or *de novo* assembled, and variants within and between individuals called (Fig. 2A; Nielsen *et al.,* 2011). With many more markers we have more reliable estimates of the impact of neutral demographic processes on the genome and genetic linkage (Luikart *et al.,* 2003; Burri, 2017). Using this understanding of neutral genome-wide patterns, studies are able to identify outlier loci which may form due to selection. Furthermore, the ability to screen the genome of hundreds, or even thousands, of individuals improves our ability to identify genetic variants existing in a region under natural selection (Stapley *et al.,* 2010; Barrett and Hoekstra, 2011). Coupled with ecological data and a high quality reference genome and/or genetic maps, studies can identify the function and phenotypic consequences of genes and alleles within these regions under selection, and potentially identify the ecological factors driving selection at these loci (Pardo-Diaz *et al.,* 2015).

There are many high-throughput NGS platforms and choosing which is best to use is dependent on the biological question being asked, existing genetic resources and target

genome characteristics, as each platform has different properties. Ultimately there is a trade-off between accuracy of base-calling, sequence read length, sequence coverage/depth and cost (Table 1). Illumina is currently the most widely used platform for population genetics, generating data by reduced representation sequencing, targeted sequence capture of loci or whole-genome resequencing, producing sequence reads ~150 - 300bp (insert fragment sizes up to 20 kb mate-pairs) with high coverage and low base-calling error. For studies with access to high quality DNA with minimal shearing, long-read platforms such as Pacific Biosystems (PacBio) Sequel (>10,000 bp) provides the ability to study structural genomic variation, such as indels, and DNA methylation with ease (Flusberg *et al.,* 2010; Rhoads and Au, 2015). Most mapping and assembly approaches require high quality base calling, long read length and coverage >10x for accurate sequence mapping and variant calling. To obtain this, many studies have applied a hybrid/multiplatform sequence approach, for example, utilizing high copy number and accurate Illumina reads with long sequence reads (>10,000 bp) from PacBio/Oxford Nanopore to improve alignment which can be a cost-effective way to produce high quality genome assemblies (Goodwin *et al.,* 2015; Küpper *et al.,* 2016; Korlach *et al.,* 2017; Peona *et al.* 2021; Fig. 2A). More recently, Hi-C technologies have allowed short-read sequencing to generate chromosome-level genome scaffolds by quantifying interactions between distant fragment pairs (Dudchenk *et al.* 2017). Furthermore, other technologies, such as BioNano Genomics for sequence motif staining and 10x Genomics Linked-Read sequencing, also provide accurate near chromosome-length scaffolds for haplotype phasing and identifying structural variants, reducing the need for other scaffolding methods (Mostovoy *et al.* 2016; Paajanen *et al.* 2019). Long-read sequencing is likely the future of genomics, providing greater structural information in fewer fragments, and in recent years base call errors have decreased dramatically (see PacBio HiFi >99.5% base calling accuracy; Wenger *et al.* 2019; Hon *et al.* 2020) (Table 1).

Until relatively recently, it was often only feasible for most studies to produce genome resequencing data for 10s of individuals. The cost (and feasibility) of generating genome resequencing data is dependent on genome size and complexity (e.g., ploidy, synteny, repetitive elements) which affect the amount and quality of sequence required to produce high-quality assemblies. Reduced-representation sequencing methods, such as restriction-site associated DNA sequencing (RAD-seq; Miller *et al.* 2007), somewhat alleviated this by providing genome-wide markers - often where no pre-existing sequence knowledge is required (Willing *et al.* 2011) - and has been utilised to study genetic variation comparatively cheaply (Fig. 2B). Such strategies identify many thousands of markers, typically covering 0.1-10% of the genome, allowing many individuals to be sequenced, but with the loss of genomic coverage and hence information (Lowry *et al.* 2017). Whole genome resequencing

with reads mapped to a reference genome, or assembled *de novo,* is becoming increasingly more accessible as sequencing costs rapidly decrease and highly contiguous reference genomes become available for many non-model species (Fig. 2A).

Table 1. Comparison of common sequencing technologies for genomic marker discovery applied to population genetics, conservation biology and evolutionary biology.

Sequencing technology	Pros	Cons	Data generated and biological applications
Sanger (1977 -)	Long sequence reads (<1 kb)[1], high precision base-calling (error ~0.0001%).	Very costly and slow, only one read therefore no sequence validation (unless using the reverse primer).	Single-call per base sequence reads: Targeted sequencing (e.g., candidate genes). Largely redundant due to cheaper, faster, more accurate NGS technologies. First human genome sequence (Lander *et al.* 2001).
Illumina (2007 -)	Low cost per base, low base-calling error (0.1-1%), high sequence depth, ability to generate large (<750 bp) fragment sizes from paired-end reads or Hi-C libraries for distant fragment pairs.	Short read length (< 600 bp or 2* 150 bp paired-end).	High-throughput sequence reads → Whole-genome re-sequencing or *de novo* assembly, reduced representation or sequence capture: Widely used: e.g., demography, selection, population history, divergence, RNA-seq differential gene expression (transcriptomics) (Parchman *et al.,* 2012; Finseth and Harrison, 2014; Zhang *et al.,* 2016; Kardos *et al.,* 2017; Li *et al.* 2020).
Pacific Biosciences (2011 -)	Long read length (~20 kb) – identify indels and genetic rearrangements, study DNA methylation, high consensus accuracy when coverage is high.	Costly, lots of high quality DNA required, older platforms have high base-calling error (1-10% per bp, partially corrected through high depth and newer HiFi).	Long sequence reads → Whole-genome re-sequencing or *de novo* assembly: Study genomic evolution (genetic architecture), selection, epigenetics (Küpper *et al.,* 2016; Korlach *et al.,* 2017). Improve pre-generated genome assemblies (Korlach *et al.* 2017).
10x Genomics (2012 -)	Linked sequence reads – chromosome-scale scaffolds, phase, chromosomal rearrangements, high accuracy (phase switch error rate 0.05%).	Unable to connect some distant heterozygous sites (e.g., at centromeres or regions with low heterozygosity).	Long- range information on a genome-wide scale: Genome-scale phasing (Srikanth *et al.* 2020) and genomic architecture such as inversions (Bedoya and Leaché 2021).
Oxford Nanopore (2015 -)	Potential for ultra-long read length (<1,000 kb), hand-held sequencers (MinION), affordable instrument cost, rapid real-time sequencing, simple preparation, RNA read directly.	High base-calling error (90% consensus accuracy, order of magnitude less than Illumina), high per-base cost.	Ultra-long sequence read(s) → Improve short-read genome assemblies, *de novo* assembly or in field species identification: Species/ population identification in the field (conservation, migration, demography, phylogeny etc.), gene mapping/ linkage, genetic architecture (Goodwin *et al.,* 2015; Loman *et al.* 2015).

[1] Despite generating relatively long sequence reads, where coverage is low and variant lengths exceed read lengths, many structural variants and indels will still be missed.

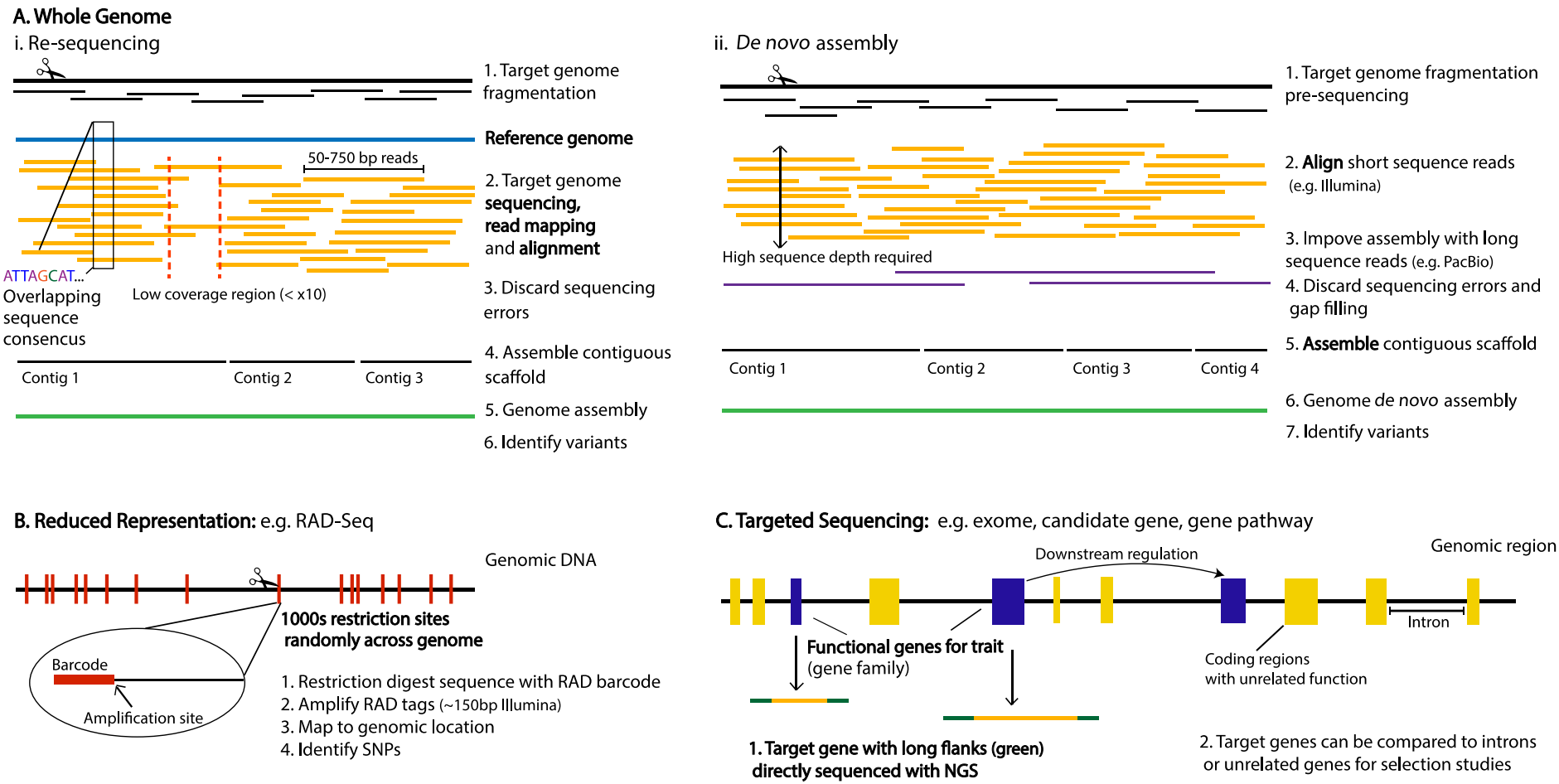


Figure 2. Sequencing strategies commonly used in population genetics. **Ai)** Whole-genome re-sequence assembly using thousands of short sequence reads and mapping to a reference genome. **Aii)** Whole-genome *de novo* assembly using short and long sequence reads. **B)** Restriction-site associated DNA Sequencing (RAD-seq). **C)** Targeted sequencing of genomic regions.

3 Linking genotype to phenotype

Combining genetic data with phenotypic measures and/or knowledge of gene function can allow hypotheses to be generated about drivers of adaptive evolution. 'Forward genetics' approaches aim to identify loci that may underlie adaptive traits. One such approach involves quantitative trait loci (QTL) mapping through pedigrees to identify loci associated with trait variation. The traditional QTL approach generates variation for a particular trait through artificial selection, driving progeny lines to express alternative phenotypes through many generations of crosses. While many studies have moved away from relying on traditional QTL mapping to determine the genetic basis for adaptation, the approach is still commonly used to narrow down loci linked to selection (Protas *et al.,* 2006; Gross, Borowsky and Tabin, 2009; Jeffery, 2009; Brachi *et al.,* 2010; Johnston *et al.,* 2010; Hendrick *et al.,* 2016), followed by loci confirmation. Second, the genome-wide association study (GWAS) approach takes advantage of historical recombination in populations to statistically detect non-random associations between genome-wide markers or candidate genes, and the trait of interest (Stinchcombe and Hoekstra, 2008; Stapley *et al.,* 2010; Pardo-Diaz *et al.,* 2015; Santure and Garant 2018). As such, this investigates variation that has been shaped over the evolutionary history of the populations, instead of few artificially crossed generations. Association studies were initially restricted to use of few, low coverage genomic markers (e.g., ALFPs, microsatellites) (Holliday, Ritland and Aitken, 2010; Johnston *et al.,* 2010), but now commonly use many thousands, or even millions, of makers via reduced representation sequencing (RAD-seq) (Parchman *et al.,* 2012; Chaves *et al.,* 2016; Hahn *et al.,* 2017), large-scale targeted sequencing (Nadeau *et al.* 2012) or whole-genome sequencing (Brawand *et al.* 2014; Nosil *et al.* 2018; Yoshida and Yáñez 2021).

In contrast to forward genetic approaches, reverse genetics attempt to identify which regions of the genome are under selection within/between populations, with the ultimate aim of identifying the biologically important phenotypes that these regions encode (Jensen *et al.* 2016). This approach applies genome-wide screening to identify regions with patterns of variation that are divergent from the rest of the genome (i.e., outlier loci reflecting signatures of selection), then attempts to link variation at genes within these regions to phenotypes which are potentially important for adaptation (Luikart *et al.,* 2003). A key benefit of this 'genome scan' approach is that the phenotypes studied are not predetermined, and instead are identified as an output of the analyses, allowing the technique to identify any actual targets of selection. This approach is commonly applied to identify genomic regions under selection between populations that vary ecologically, i.e., in habitat or geography (e.g., Bonin *et al.,* 2006; Hohenlohe *et al.,* 2010; Fabian *et al.,* 2012; Nadeau *et al.,* 2012).

Signatures of selection may be detected and mapped using a range of statistical analyses (Beaumont and Balding, 2004; Nunes *et al.,* 2012; De Villemereuil *et al.,* 2014; Tuttle *et al.,* 2016). Selection for a specific allele leaves a pattern of reduced genetic variation around the selected locus, elevated linkage disequilibrium or a skewed distribution of genetic polymorphisms (allele frequency spectrum) (Luikart *et al.* 2003). Across populations, genetic differentiation will be higher for loci under divergent selection compared to the rest of the genome (discussed above). Within a single population, patterns of reduced heterozygosity (nucleotide diversity, π) and allelic fixation (Tajima's D), are commonly used to map the position and strength of selection on loci (e.g., Huynh *et al.,* 2011; Riesch *et al.,* 2017). Between populations, genetic differentiation as a result of allele frequency differences, measured through F_{ST}, is most often used to identify footprints of selection (Beaumont and Nichols, 1996; e.g., Westram *et al.,* 2014; Hahn *et al.,* 2017), although d_{xy}, a measure of absolute divergence, is also commonly used (Huynh *et al.,* 2011; Küpper *et al.,* 2016).

In the last decade, the focus of studies has largely shifted from forward genetics approaches using candidate genes with known function for a trait of interest, to studying genome-wide markers either to confirm the genetic basis of a trait or to identify which traits are being selected for. This is largely because of the reduction in costs of genome screening, which makes it possible to identify strong divergent loci with greater power and link these loci to genomic locations. Genome scan approaches are enabling the identification of new loci and traits involved in adaptation, which have not previously been identified through forward genetics approaches (e.g., Jones *et al.,* 2012; Bosse *et al.* 2017; Walsh *et al.* 2019; McCulloch *et al.* 2021).

4 Island archipelagos to study evolution

Populations on island archipelagos provide excellent systems for evolutionary research, providing an opportunity to study numerous evolutionary processes including founder effects, gene flow, genetic drift and differential selection, across spatial and temporal scales in nature (Warren *et al.* 2015). There are many reasons for this (see review Emerson 2002). Firstly, each island forms a geographically distinct landscape harbouring a unique combination of selection pressures, which may be nested within an archipelago (Aguilée *et al.* 2021). Within archipelagos there is often variation in the ecology of individual islands as a result of varied geology, topography, climate and inhabiting species. Second, the colonisation of each island may have occurred over different divergence timeframes,

resulting in different population history including founder effects. The ocean provides a barrier to gene flow, moderating the level of migration between populations and may result in inter-island geographical isolation. Gene flow also moderates the dispersal of pathogens or communities of predators across archipelagos resulting in selection gradients which may exert substantial and differing pressure on populations within and across islands. Third, islands are typically small in size compared to mainland population ranges, and as such are relatively simplistic in terms of species diversity and ecosystem dynamics. Likely owing to the combination of these evolutionary dynamics across island systems, the rates of speciation, endemism and ecological diversity in such systems are high (Emerson 2008).

Island biogeography has been keenly studied since Darwin (1831-1836) and Wallace's (1854-1862) voyages to - among other places - the Galapagos Islands and Malay archipelago, the fauna of which shaped their theories about natural selection and adaptation. Many famous studies of evolutionary dynamics across island archipelagos now exist which are developing our understanding of how genetic and morphological diversity is shaped across natural populations (for examples see San Nicolas island foxes (*Urocyon littoralis*; Robinson *et al.* 2018), Bojer's skink (*Gongylomorphus bojerii*; du Plessis *et al.* 2019), silvereye (*Zosterops lateralis*; Sendell-Price *et al.* 2021) and Hawaiian spiders (genus *Tetragnatha*; Cotoras *et al.* 2018). Islands will continue to be important model systems for evolutionary biology in the era of genomics.

5 Berthelot's pipit (*Anthus berthelotii*) – a recently evolved island endemic

Berthelot's pipit (*Anthus berthelotii*) is a small, sedentary passerine endemic to three Macaronesian archipelagos in the north Atlantic ocean; the Canary Islands (nine populations on eight islands), Selvagens (one island population), and the Madeiran archipelago (three island populations; Fig. 3). On Tenerife in the Canary Islands, two populations exist separated by dense forest: one across the coastal lowlands and a second on the alpine plateau of El Teide >2000m asl. These islands are volcanic in origin, formed between ~0.8 (El Hierro, Canary Islands) and 30 million years ago (Selvagens), and vary in size, isolation, altitude, climate and habitat (Hoernle and Carracedo 2020). Berthelot's pipits inhabit a diversity of habitats across the range, including semi-arid coastal scrub, dry subalpine scrub, temperate meadows and urban landscapes (Fig. 4 C-D). The sister species and closest relative of the Berthelot's pipit is the tawny pipit (*Anthus campestris*), which is distributed across mainland Africa and the Iberian peninsula. The two species are estimated (using

cytochrome *b* mitochondrial sequences) to have been separated for approximately two million years (Voelker 1999). Previous research shows that the ancestor of Berthelot's pipit initially colonised the Canary Islands, likely from mainland Africa (Illera *et al.* 2007). More recent northward range expansion then occurred to the Madeiran archipelago and Selvagens through independent founder events from the Canaries (Spurgin *et al.* 2014). The effects of genetic bottlenecks have been detected in both the Madeiran and Selvagens populations of the Berthelot's pipit, and genetic and morphological variation supports a model of isolation-by-colonisation, shaped by founder events (Spurgin *et al.* 2014). Contemporary estimates of population size also reflect island size across the range (Spurgin *et al.* 2014). Berthelot's pipit are classified into two subspecies: *Anthus berthelotii berthelotii* inhabits the Canary Islands and Selvagens, while *Anthus berthelotii madeirensis*, which inhabits the Madeiran islands, is characterised by longer bill lengths (Arctander *et al.* 1996; Illera *et al.* 2007) and larger body size (Spurgin *et al.* 2014). The relative role of selection and founder effects in shaping divergence for this trait is currently not clear (Armstrong *et al.* 2018).

As well as variation in habitat, climate and altitude, pathogen prevalence varies substantially between Berthelot's pipit populations (Illera *et al.* 2008). Pathogen screening across the islands has detected the presence of *Plasmodium* and *Leucocytozoon* parasites, and *Avipoxvirus* (avian pox virus) (see Fig. 4B). Importantly, the prevalence of these diseases is consistent over the years the populations have been studied and follows patterns expected as a result of island biogeography, i.e., smaller, more isolated islands harboured fewer pathogens, with no disease detected within some islands across the range (Spurgin *et al.* 2012). Studies of malaria and pox infection using RAD-seq data have identified associations between these pathogens and SNPs near or within genes involved in immune response, among archipelagos (malaria, Armstrong *et al.* 2018; pox, Sheppard *et al.*, 2022) and within individual island populations (Gonzalez-Quevedo, Davies, *et al.* 2014; Armstrong *et al.* 2019). Genome scans identified further loci in genes associated with immunity and metabolism with strongly divergent allele frequencies across populations. Further studies have identified evolution within key immune gene families including major histocompatibility complex (MHC) (Gonzalez-Quevedo, Phillips, *et al.* 2014; Spurgin *et al.* 2011) and toll-like receptors (TLRs) (Gonzalez-Quevedo *et al.* 2015), as well as near to genes within the Bone Morphogenetic Protein (BMP) family (Armstrong *et al.* 2018) potentially associated with bill morphology.

Previous studies of the Berthelot's pipit evidence the importance of pathogens, among other forces, in exerting selective pressure across the range, although it is unclear if genomic

regions, and hence traits, of importance for adaptive evolution have gone undetected. It is likely that i) reduced representation sequencing and candidate gene approaches have missed many loci under selection, and ii) genetic signatures are eroded by subsequent evolutionary forces across spatial scales. Furthermore, with few genetic markers, inferences of population history including colonisation, gene flow and bottlenecks have less power and inferences may be incorrect or missing details. Currently, little is known about mainland-island divergence of the tawny and Berthelot's pipit ancestor, or the genomic characteristics of this speciation event. Thus, many questions remain about how neutral and adaptive pressures shape genetic diversity across spatial scales in the Berthelot's pipit system.

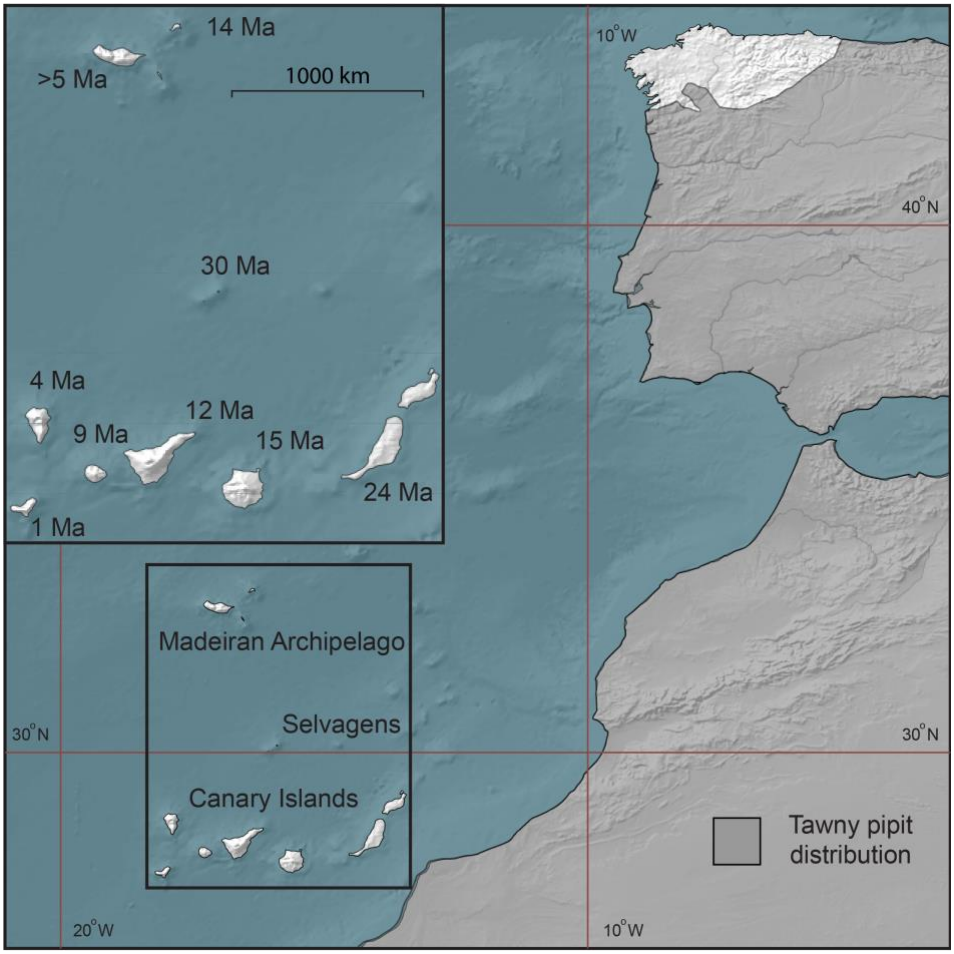


Figure 3. The distribution of Berthelot's pipit (*Anthus berthelotii*) populations across three north Atlantic archipelagos (see box) and its sister species, the tawny pipit (*Anthus campestris*) across mainland Africa and the Iberian peninsula (grey). Geological age estimates across the volcanic provinces from which the Canary Islands, Selvagens and Madeiran archipelago have arisen are provided (top box) based on data from Geldmacher *et al.* (2005) and Guillou *et al.* (1996).

Figure 4. A) Clap traps baited with mealworm (*Tenebrio molitor*) larvae are used to catch Berthelot's pipit individuals for subsequent sampling. **B)** An adult Berthelot's pipit infected with avian pox showing characteristic lesions on the foot and bent nail and bill growth. **C)** typical Berthelot's pipit semi-arid scrub habitat and **D)** subalpine scrub habitat.

6 Aims of thesis

In this thesis, I examine how population history, genetic drift and adapation shape patterns of genomic diversity across the natural range of an island endemic bird. Specifically, I ask: how do neutral and selective forces shape genomic variation across spatial scales in nature?

In **Chapter 2** I use RAD-seq, across 13 populations of Berthelot's pipit to explore evolutionary processes acting between populations within archipelgos separated relatively recently, by assessing genetic diversity, population history, gene flow and selection. I test for population structure, determine if signatures of post-colonisation gene flow exist across this species range, and assess the strength and drivers of selection at such spatial scales. This chapter builds on previous analyses of population history and selection at broad spatial scales, to uncover the evolutionary processes acting at finer spatial scales across the Berthelot's pipit range, that may be eroded by subsequent drift and mutation.

In **Chapter 3**, I use data from whole genome resequencing of Berthelot's pipits to explore colonisation timescales, associated bottlenecks and contemporary inbreeding for the initial mainlaind-island colonisation and speciation from the tawny pipit, and subsequent dispersal across the three north Atlantic archipelgos. In particular, this chapter investigates patterns of genetic diversity across the genomes of contemporary individuals - including identfying runs of homozygosity (ROH) - to infer the evolutionary mechanisms driving changes in genetic diversity across the Berthelot's pipit range. By modelling population history through initial island colonisation, speciation and dispersal across archipelgos, this chapter assesses how sequential population founding results in cumulative loss of genetic diversity where there is an absence of post-colonisation gene flow. This chapter also investigates how whole genomes can be used to build a dynamic picture of population history across different evolutionary timescales.

In **Chapter 4**, I use the genomic information derived in Chapter 3 to explore the landscape of genomic divergence between Berthelot's pipit populations, and between Berthelot's pipits and tawny pipits. A particular focus is on identifying 'genomic islands of divergence' to identify the genes that may be important for adaptive evolution and incipient speciation across the Berthelot's pipit's range.

Finally, in the **general disscussion**, I discuss the findings of Chapters 2-4 and their combined significance, and suggest possible directions for future research.

7 References

Aguilée, R., Pellerin, F., Soubeyrand, M., Choin, J. and Thébaud, C. (2021). Biogeographic drivers of community assembly on oceanic islands: The importance of archipelago structure and history. *Journal of Biogeography*:1–13.

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

Armstrong, C., Davies, R.G., González-Quevedo, C., Dunne, M., Spurgin, L.G. and Richardson, D.S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution* **0**:1–21.

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.

Barrett, R.D.H. and Hoekstra, H.E. (2011). Molecular spandrels: tests of adaptation at the genetic level. *Nature Reviews Genetics* **12**:767–780.

Barton, N.H. and Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**:133–164.

Baxter, S.W., Papa, R., Chamberlain, N., Humphray, S.J., Joron, M., Morrison, C., Ffrench-Constant, R.H., *et al.* (2008). Convergent evolution in the genetic basis of Müllerian mimicry in *Heliconius* butterflies. *Genetics* **180**:1567–1577.

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

Beaumont, M. and Nichols, R. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B-Biological Sciences* **263**:1619–1626.

Bedoya, A.M. and Leaché, A.D. (2021). Characterization of a pericentric inversion in plateau fence lizards (*Sceloporus tristichus*): Evidence from chromosome-scale genomes. *G3: Genes, Genomes, Genetics* **11**:2–6.

Bertrand, J.A.M., Delahaie, B., Bourgeois, Y.X.C., Duval, T., Garcia-Jimenez, R., Cornuault, J., Thebaud, C., *et al.* (2016). The role of selection and historical factors in driving population differentiation along an elevational gradient in an island bird. *Journal of Evolutionary Biology* **29**:824–836.

Bicknell, A.W.J., Knight, M.E., Bilton, D., Reid, J.B., Burke, T. and Votier, S.C. (2012). Population genetic structure and long-distance dispersal among seabird populations: Implications for colony persistence. *Molecular Ecology* **21**:2863–2876.

Bluher, S.E., Miller, S.E. and Sheehan, M.J. (2020). Fine-scale population structure but limited genetic differentiation in a cooperatively breeding paper wasp. *Genome Biology and Evolution* **12**:701–714.

Boag, P.T. and Grant, P.R. (1981). Intense natural selection in a population of Darwin's Finches (*Geospizinae*) in the Galápagos. *Science* **214**:82–85.

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

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

Brachi, B., Faure, N., Horton, M., Flahauw, E., Vazquez, A., Nordborg, M., Bergelson, J., *et al.* (2010). Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genetics* **6**:1–17.

Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., *et al.* (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**:375–381.

Britten, H.B. (1996). Meta-analyses of the association between multilocus heterozygosity and fitness. *Evolution* **50**:2158–2164.

Bruford, M.W. and Wayne, R.K. (1993). Microsatellites and their application to population genetic studies. *Current Opinion in Genetics and Development* **3**:939–943.

Brumfield, R.T., Beerli, P., Nickerson, D.A. and Edwards, S. V. (2003). The utility of single nucleotide polymorphisms in inferences of population history. *Trends in Ecology and Evolution* **18**:249–256.

Brzeski, K.E., Rabon, D.R., Chamberlain, M.J., Waits, L.P. and Taylor, S.S. (2014). Inbreeding and inbreeding depression in endangered red wolves (*Canis rufus*). *Molecular Ecology* **23**:4241–4255.

Burke, M.K. (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proceedings of the Royal Society B: Biological Sciences* **279**:5029–5038.

Chapman, J.R., Nakagawa, S., Coltman, D.W., Slate, J. and Sheldon, B.C. (2009). A quantitative review of heterozygosity-fitness correlations in animal populations. *Molecular Ecology* **18**:2746–2765.

Charlesworth, B., Nordborg, M. and Charlesworth, D. (1997). The effects of background and interference selection on patterns of genetic variation in subdivided populations. *Genetics Research* **70**:155–174.

Charlesworth, B. and Smith, D.B. (1982). A computer model of speciation by founder effects. *Genetical Research* **39**:227–236.

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

Choi, J.Y., Purugganan, M. and Stacy, E.A. (2020). Divergent selection and primary gene flow shape incipient speciation of a riparian tree on Hawaii Island. *Molecular Biology and Evolution* **37**:695–710.

Cornetti, L., Valente, L.M., Dunning, L.T., Quan, X., Black, R.A., Hébert, O.H. and Savolainen, V. (2015). The genome of the 'great speciator' provides insights into bird diversification. *Genome Biology and Evolution* **7**:2680–2691.

Cotoras, D.D., Bi, K., Brewer, M.S., Lindberg, D.R., Prost, S. and Gillespie, R.G. (2018). Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. *BMC Evolutionary Biology* **18**:1–13.

Cowles, S.A. and Uy, J.A.C. (2019). Rapid, complete reproductive isolation in two closely related *Zosterops* White-eye bird species despite broadly overlapping ranges. *Evolution* **73**:1647–1662.

Cruickshank, T.E. and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology* **23**:3133–3157.

Delahaie, B., Cornuault, J., Masson, C., Bertrand, J.A.M. and Eb, C.T.H. (2017). Narrow hybrid zones in spite of very low population differentiation in neutral markers in an island bird species complex. *Journal of Evolutionary Biology* **30**:2132–2145.

Dlugosch, K.M. and Parker, I.M. (2008). Founding events in species invasions: Genetic variation, adaptive evolution, and the role of multiple introductions. *Molecular Ecology* **17**:431–449.

Dudchenk, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., *et al.* (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**:92–95.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., *et al.* (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**:756–760.

Elmer, K.R. (2019). Barrier Loci and Evolution. *eLS*:1–7.

Emerson, B.C. (2008). A Century of Evolution: Ernst Mayr (1904 – 2005); Speciation on islands: what are we learning? *Biological Journal of the Linnean Society* **95**:17–25.

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. and Flatt, T. (2012). Genome-wide patterns of latitudinal differentiation among populations of *Drosophila melanogaster* from North America. *Molecular Ecology* **21**:4748–4769.

Feder, J.L., Flaxman, S.M., Egan, S.P., Comeault, A.A. and Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**:73–97.

Flusberg, B.A., Webster, D., Lee, J., Travers, K., Olivares, E., Clark, A., Korlach, J., *et al.* (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**:461–465.

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

Garten, C.T. (1976). Relationships between aggressive behavior and genic heterozygosity in the Oldfield Mouse, *Peromyscus polionotus*. *Evolution* **30**:59–72.

Gaudeul, M., Till-Bottraud, I., Barjon, F. and Manel, S. (2004). Genetic diversity and differentiation in *Eryngium alpinum L.* (*Apiaceae*): Comparison of AFLP and microsatellite markers. *Heredity* **92**:508–518.

Geldmacher, J., Hoernle, K., Bogaard, P. V.D., Duggen, S. and Werner, R. (2005). New 40Ar39Ar age and geochemical data from seamounts in the Canary and Madeira volcanic provinces: Support for the mantle plume hypothesis. *Earth and Planetary Science Letters* **237**:85–101.

Gonzalez-Quevedo, C., Davies, R.G. and 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.

Gonzalez-Quevedo, C., Phillips, K.P., Spurgin, L.G. and Richardson, D.S. (2014). 454 screening of individual MHC variation in an endemic island passerine. *Immunogenetics* **67**:149–162.

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

Goodwin, S., Gurtowski, J., Ethe-Sayers, S., Deshpande, P., Schatz, M. and McCombie, W.R. (2015). Oxford Nanopore Sequencing and de novo Assembly of a Eukaryotic Genome. *bioRxiv*:13490.

Goossens, B., Sharma, R., Othman, N., Kun-Rodrigues, C., Sakong, 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.

Gross, J.B., Borowsky, R. and Tabin, C.J. (2009). A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics* **5**.

Guillou, H., Carracedo, J.C., Pérez Torrado, F. and Rodriguez Badiola, E. (1996). K-Ar ages and magnetic stratigraphy of a hotspot-induced, fast grown oceanic island: El Hierro, Canary Islands. *Journal of Volcanology and Geothermal Research* **73**:141–155.

Hahn, C., Genner, M.J., Turner, G.F. and Joyce, D.A. (2017). The genomic basis of cichlid fish adaptation within the deepwater 'twilight zone' of Lake Malawi. *Evolution Letters*:184–198.

Hansson, B. and Westerberg, L. (2002). On the correlation between heterozygocity and fitness in natural populations. *Molecular Ecology* **11**:2467–2474.

Hendrick, M.F., Finseth, F.R., Mathiasson, M.E., Palmer, K.A., Broder, E.M., Breigenzer, P. and Fishman, L. (2016). The genetics of extreme microgeographic adaptation: an integrated approach identifies a major gene underlying leaf trichome divergence in Yellowstone *Mimulus guttatus*. *Molecular Ecology* **25**:5647–5662.

Hermisson, J. and Pennings, P.S. (2017). Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution* **8**:700–716.

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

Hoernle, K. and Carracedo, J.-C. (2020). Canary Islands, Geology. *Encyclopedia of Islands*:133–143.

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

Holliday, J.A., Ritland, K. and Aitken, S.N. (2010). Widespread, ecologically relevant genetic markers developed from association mapping of climate-related traits in Sitka spruce (*Picea sitchensis*). *New Phytologist* **188**:501–514.

Hon, T., Mars, K., Young, G., Tsai, Y.C., Karalius, J.W., Landolin, J.M., Maurer, N., *et al.* (2020). Highly accurate long-read HiFi sequencing data for five complex genomes. *Scientific Data* **7**:1–11.

Hooper, R., Excoffier, L., Forney, K.A., Gilbert, M.T.P., Michael, D., Morin, P.A., Wolf, J.B.W., *et al.* (2020). Runs of homozygosity in killer whale genomes provide a global record of demographic histories. *bioRxiv*:1–22.

Huynh, L.Y., Maney, D.L. and Thomas, J.W. (2011). Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). *Heredity* **106**:537–546.

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

Izuno, A., Kitayama, K., Onoda, Y., Tsujii, Y., Hatakeyama, M., Nagano, A.J. and Honjo, M.N. (2017). The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha (Myrtaceae)*. *Molecular Ecology* **26**:1515–1532.

de Jager, D., Glanzmann, B., Möller, M., Hoal, E., van Helden, P., Harper, C. and Bloomer, P. (2021). High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Scientific Reports* **11**:1–15.

Jeffery, W.R. (2009). Regressive Evolution in *Astyanax* Cavefish. *Annual Review of Genetics* **43**:25–47.

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

Johnston, S.E., Beraldi, D., McRae, A.F., Pemberton, J.M. and Slate, J. (2010). Horn type and horn length genes map to the same chromosomal region in Soay sheep. *Heredity* **104**:196–205.

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

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., *et al.* (2017). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution* **2**:124–131.

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

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

Kirkpatrick, M. and Jarne, P. (2000). The effects of a bottleneck on inbreeding depression and the genetic load. *American Naturalist* **155**:154–167.

Kirkpatrick, M. and Ravigné, V. (2002). Speciation by natural and sexual selection: Models and experiments. *American Naturalist* **159**.

Korlach, J., Gedman, G., Kingan, S.B., Chin, C.S., Howard, J.T., Audet, J.N., Cantin, L., *et al.* (2017). De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience* **6**:1–16.

Küpper, C., Stocks, M., Risse, J.E., Dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., *et al.* (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics* **48**:79–83.

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

Lande, R. (1980). Genetic variation and phenotypic evolution during allopatric speciation. *The American Naturalist* **116.**

Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., *et al.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**:860–921.

Lander, T.A., Klein, E.K., Roig, A. and Oddou-Muratorio, S. (2021). Weak founder effects but significant spatial genetic imprint of recent contraction and expansion of European beech populations. *Heredity* **126**:491–504.

Leberg, P.L. (1992). Effects of population bottlenecks on genetic diversity as measured by allozyme electrophoresis. *Evolution* **46**:477–494.

Li, F., Bian, L., Ge, J., Han, F., Liu, Z., Li, X., Liu, Y., *et al.* (2020). Chromosome-level genome assembly of the East Asian common octopus (*Octopus sinensis*) using PacBio sequencing and Hi-C technology. *Molecular Ecology Resources* **20**:1572–1582.

Loman, N.J., Quick, J. and Simpson, J.T. (2015). A complete bacterial genome assembled de novo using only nanopore sequencing data. *Nature Methods* **12**:733–735.

Losos, J.B. and Ricklefs, R.E. (2009). Adaptation and diversification on islands. *Nature Review Insight* **457**.

Lowry, D.B., Hoban, S., Kelley, J.L., Lotterhos, K.E., Reed, L.K., Antolin, M.F. and 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. and Taberlet, P. (2003). The power and promise of population genomics: From genotyping to genome typing. *Nature Reviews Genetics* **4**:981–994.

Macleod, A., Rodrıguez, A., Vences, M., Orozco-terWengel, P., Garcia, C., Trillmich, F., Gentile, G., *et al.* (2015). Hybridization masks speciation in the evolutionary history of the Galapagos marine iguana. *Proceedings of the Royal Society B*.

Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., Gielly, L., *et al.* (2012). Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology* **21**:3729–3738.

Marques, D.A., Lucek, K., Meier, J.I. and Mwaiko, S. (2016). Genomics of rapid incipient speciation in sympatric Threespine Stickleback. *PLoS Genetics* **12**:1–34.

Maruyama, T. and Fuerst, P.A. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics* **111**:675–689.

Maynard-Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**:23–35.

Mayr, E. (1965). *The Nature of Colonizations in Birds*. Academic Press, London.

McCulloch, G.A., Guhlin, J., Dutoit, L., Harrop, T.W.R., Dearden, P.K. and Waters, J.M. (2021). Genomic signatures of parallel alpine adaptation in recently evolved flightless insects. *Molecular Ecology*:1–10.

Milá, B., Warren, B.H., Heeb, P. and Thébaud, C. (2010). The geographic scale of diversification on islands: Genetic and morphological divergence at a very small spatial scale in the Mascarene grey white-eye (*Aves : Zosterops borbonicus*). *BMC Evolutionary Biology* **10**:1–13.

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

Mirimin, L., Miller, R., Dillane, E., Berrow, S.D., Ingram, S., Cross, T.F. and Rogan, E. (2011). Fine-scale population genetic structuring of bottlenose dolphins in Irish coastal waters. *Animal Conservation* **14**:342–353.

Mostovoy, Y., Levy-sakin, M., Lam, J., Lam, E.T., Hastie, A.R., Lee, J., Chu, C., *et al.* (2016). A Hybrid Approach for de novo Human Genome Sequence Assembly and Phasing. *Nature Methods* **13**:587–590.

Muller, H.J. (1950). *Our Load of Mutations*.

Nadachowska-Brzyska, K., Burri, R., Smeds, L. and Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white *Ficedula* flycatchers. *Molecular Ecology* **25**:1058–1072.

Nadeau, N.J., Ruiz, M., Salazar, P., Counterman, B., Medina, J.A., Ortiz-Zuazaga, H., Morrison, A., *et al.* (2014). Population genomics of parallel hybrid zones in the mimetic butterflies, *H. melpomene* and *H. erato*. *Genome Research* **24**:1316–1333.

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

Nei, M. (1983). *Genetic polymorphism and the role of mutation in evolution*.

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

Nevado, B., Atchison, G.W., Hughes, C.E. and Filatov, D.A. (2016). Widespread adaptive evolution during repeated evolutionary radiations in New World lupins. *Nature Communications* **7**.

Nevado, B., Wong, E.L.Y., Osborne, O.G. and Filatov, D.A. (2019). Adaptive evolution is common in rapid evolutionary radiations. *Current Biology* **29**:3081–3086.e5.

Nevo, E. (1978). Genetic variation in natural populations: Patterns and theory. *Theoretical Population Biology* **13**:121–177.

Nielsen, R., Paul, J.S., Albrechtsen, A. and Song, Y.S. (2011). Genotype and SNP calling from next-generation sequencing data. *Nature Reviews Genetics* **12**:443–451.

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

Nosil, P., Villoutreix, R., De Carvalho, C.F., Farkas, T.E., Soria-Carrasco, V., Feder, J.L., Crespi, B.J., *et al.* (2018). Natural selection and the predictability of evolution in *timema* stick insects. *Science* **359**:765–770.

Novembre, J. and Di Rienzo, A. (2009). Spatial patterns of variation due to natural selection in humans. *Nature Reviews Genetics* **10**:745–755.

Nunes, V.L., Beaumont, M.A., Butlin, R.K. and Paulo, O.S. (2012). Challenges and pitfalls in the characterization of anonymous outlier AFLP markers in non-model species: Lessons from an ocellated lizard genome scan. *Heredity* **109**:340–348.

Ony, M.A., Nowicki, M., Boggess, S.L., Klingeman, W.E., Zobel, J.M., Trigiano, R.N. and Hadziabdic, D. (2020). Habitat fragmentation influences genetic diversity and differentiation: Fine-scale population structure of *Cercis canadensis* (eastern redbud). *Ecology and Evolution* **10**:3655–3670.

Oostermeijer, J.G.B., Van Eijck, M.W., Van Leeuwen, N.C. and Den Nijs, J.C.M. (1995). Analysis of the relationship between allozyme heterozygosity and fitness in the rare *Gentiana pneumonanthe L. Journal of Evolutionary Biology* **8**:739–759.

Paajanen, P., Kettleborough, G., López-Girona, E., Giolai, M., Heavens, D., Baker, D., Lister, A., *et al.* (2019). A critical comparison of technologies for a plant genome sequencing project. *GigaScience* **8**:1–12.

Parchman, T.L., Gompert, Z., Mudge, J., Schilkey, F.D., Benkman, C.W. and Buerkle, C.A. (2012). Genome-wide association genetics of an adaptive trait in lodgepole pine. *Molecular Ecology* **21**:2991–3005.

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

Patton, A.H., Margres, M.J., Stahlke, A.R., Hendricks, S., Lewallen, K., Hamede, R.K., Ruiz-Aravena, M., *et al.* (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian Devils. *Molecular Biology and Evolution* **36**:2906–2921.

Peláez, P., Ortiz-Martínez, A., Figueroa-Corona, L., Montes, J.R. and Gernandt, D.S. (2020). Population structure, diversifying selection, and local adaptation in *Pinus patula*. *American Journal of Botany* **107**:1555–1566.

Peona, V., Blom, M.P.K., Xu, L., Burri, R., Sullivan, S., Bunikis, I., Liachko, I., *et al.* (2021). Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Molecular Ecology Resources* **21**:263–286.

Pilot, M., Greco, C., Vonholdt, B.M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., Sidorovich, V.E., *et al.* (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity* **112**:428–442.

du Plessis, S.J., Howard-McCombe, J., Melvin, Z.E., Sheppard, E.C., Russo, I.R.M., Mootoocurpen, R., Goetz, M., *et al.* (2019). Genetic diversity and cryptic population re-establishment: management implications for the Bojer's skink (*Gongylomorphus bojerii*). *Conservation Genetics* **20**:137–152.

Protas, M.E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W.R., Zon, L.I., *et al.* (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* **38**:107–111.

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.

Ray, J.M. and King, R.B. (2006). The temporal and spatial scale of microevolution : Fine-scale colour pattern variation in the Lake Erie watersnake, *Nerodia sipedon insularum*. *Evolutionary Ecology Research* **8**:915–928.

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

Rhoads, A. and Au, K.F. (2015). PacBio sequencing and its applications. *Genomics, Proteomics and Bioinformatics* **13**:278–289.

Rice, W.R. (1987). Speciation via habitat specialization: the evolution of reproductive isolation as a correlated character. *Evolutionary Ecology* **1**:301–314.

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A.A., Farkas, T.E., Lucek, K., *et al.* (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology and Evolution* **1**:1–13.

Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E. and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Current Biology* **28**:3487–3494.e4.

Ross, K.G. (1997). Multilocus evolution in fire ants: Effects of selection, gene flow and recombination. *Genetics* **145**:961–974.

Ruegg, K., Anderson, E.C., Boone, J., Pouls, J. and Smith, T.B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**:4757–4769.

Saghai Maroof, M.A., Biyashev, R.M., Yang, G.P., Zhang, Q. and Allard, R.W. (1994). Extraordinarily polymorphic microsatellite DNA in barley: species diversity, chromosomal locations, and population dynamics. *Proceedings of the National Academy of Sciences* **91**:5466–5470.

Santure, A.W. and Garant, D. (2018). Wild GWAS—association mapping in natural populations. *Molecular Ecology Resources* **18**:729–738.

Schneider, K., White, T.J., Mitchell, S., Adams, C.E., Reeve, R. and Elmer, K.R. (2021). The pitfalls and virtues of population genetic summary statistics: Detecting selective sweeps in recent divergences. *Journal of Evolutionary Biology* **34**:893–909.

Seehausen, O. (2006). African cichlid fish: A model system in adaptive radiation research. *Proceedings of the Royal Society B: Biological Sciences* **273**:1987–1998.

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. and Clegg, S.M. (2021). An island-hopping bird reveals how founder events shape genome-wide divergence. *Molecular Ecology* **30**:2495–2510.

Shannon, L.M., Boyko, R.H., Castelhano, M., Corey, E., Hayward, J.J., McLean, C., White, M.E., *et al.* (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences of the United States of America* **112**:13639–13644.

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

Slate, J., David, P., Dodds, K.G., Veenvliet, B.A., Glass, B.C., Broad, T.E. and McEwan, J.C. (2004). Understanding the relationship between the inbreeding coefficient and multilocus heterozygosity: Theoretical expectations and empirical data. *Heredity* **93**:255–265.

Slatkin, M. (1987). Gene Flow and the Geographic Structure of Natural Populations. *The American Journal of Tropical Medicine and Hygiene* **236**:787–792.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. and 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. and 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. and Richardson, D.S. (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**:5213–5225.

Srikanth, K., Park, J.E., Lim, D., Cha, J., Cho, S.R., Cho, I.C. and Park, W. (2020). A comparison between HI-C and 10x genomics linked read sequencing for whole genome phasing in hanwoo cattle. *Genes* **11**.

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

Sutton, J.T., Nakagawa, S., Robertson, B.C. and Jamieson, I.G. (2011). Disentangling the roles of natural selection and genetic drift in shaping variation at MHC immunity genes. *Molecular Ecology* **20**:4408–4420.

Tennessen, J.A. and Akey, J.M. (2011). Parallel adaptive divergence among geographically diverse human populations. *PLoS Genetics* **7**.

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

Turbek, S.P., Browne, M., Giacomo, A.S. Di, Kopuchian, C., Hochachka, W.M., Estalles, C., Lijtmaer, D.A., *et al.* (2021). Rapid speciation via the evolution of pre-mating isolation in the Iberá Seedeater. *Science* **371**.

Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., *et al.* (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology* **26**:344–350.

Via, S. and West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**:4334–4345.

De Villemereuil, P., Frichot, É., Bazin, É., François, O. and Gaggiotti, O.E. (2014). Genome scan methods against more complex models: When and how much should we trust them? *Molecular Ecology* **23**:2006–2019.

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

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., *et al.* (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**:4407–4414.

Walsh, J., Benham, P.M., Deane-Coe, P.E., Arcese, P., Butcher, B.G., Chan, Y.L., Cheviron, Z.A., *et al.* (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters* **3**:324–338.

Wang, I.J. (2009). Fine-scale population structure in a desert amphibian: Landscape genetics of the black toad (*Bufo exsul*). *Molecular Ecology* **18**:3847–3856.

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

Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.C., Hall, R.J., Concepcion, G.T., Ebler, J., *et al.* (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature Biotechnology* **37**:1155–1162.

Westram, A.M., Galindo, J., Alm Rosenblad, M., Grahame, J.W., Panova, M. and Butlin, R.K. (2014). Do the same genes underlie parallel phenotypic divergence in different Littorina saxatilis populations? *Molecular Ecology* **23**:4603–4616.

Willing, E.M., Hoffmann, M., Klein, J.D., Weigel, D. and Dreyer, C. (2011). Paired-end RAD-seq for de novo assembly and marker design without available reference. *Bioinformatics* **27**:2187–2193.

Wright, S. (1931). Evolution in mendelian populations. *Bulletin of Mathematical Biology* **52**:241–295.

Wright, S. (1932). The roles of mutation, inbreeding, crossbreeding and selection in evolution. *Proceedings of the Sixth International Congress of Genetics* **1**:356–366.

Yoshida, G.M. and Yáñez, J.M. (2021). Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* **22**:1–13.

Zhang, W., Dasmahapatra, K.K., Mallet, J., Moreira, G.R.P. and Kronforst, M.R. (2016). Genome-wide introgression among distantly related *Heliconius* butterfly species. *Genome Biology* **17**:25

2

Genomic variation, population history and within-archipelago adaptation between island bird populations

A version of this chapter is published in *Royal Society Open Science* **8**: 201146

Claudia A. Martin, Claire Armstrong, Juan Carlos Illera, Brent C. Emerson, David S. Richardson, Lewis G. Spurgin

Catching pipits using clap traps on La Graciosa. Photograph by Beth Fiducia-Brookes

1 Abstract

Oceanic island archipelagos provide excellent models to understand evolutionary processes. Colonisation events and gene flow can interact with selection to shape genetic variation at different spatial scales. Landscape-scale variation in biotic and abiotic factors may drive fine-scale selection within islands, while long-term evolutionary processes may drive divergence between distantly related populations. Here, we examine patterns of population history and selection between recently diverged populations of the Berthelot's pipit (*Anthus berthelotii*), a passerine endemic to three north Atlantic archipelagos. First we use demographic trees and f_3 statistics to show that genome-wide divergence across the species range is largely shaped by colonisation and bottlenecks, with evidence of very weak gene flow between populations. Then, using a genome scan approach, we identify signatures of divergent selection within archipelagos at SNPs in genes potentially associated with craniofacial development and DNA repair. We did not detect within-archipelago selection at the same SNPs as were detected previously at broader spatial scales between archipelagos, but did identify signatures of selection at loci associated with similar biological functions. These findings suggest that similar ecological factors may repeatedly drive selection between recently separated populations, as well as at broad spatial scales across varied landscapes.

Keywords: Adaptation, gene flow, colonisation history, genome scan, spatial scales.

2 Introduction

Characterising evolution at the genetic level is fundamental to our understanding of how populations adapt in response to changing ecological pressures (Pardo-Diaz *et al.* 2015). The ability of species to adapt depends upon the amount of genetic diversity within populations, which in turn depends upon mutational processes, past and present demography, and selection. For a comprehensive understanding of how natural selection shapes genetic variation, studies are required on a variety of species with differing (and known) demographic histories, and across populations which have faced a wide range of selection pressures (Luikart *et al.* 2003; Burri 2017). Studies on humans and laboratory model species have been important for our understanding of natural selection (e.g., Exposito-Alonso *et al.* 2018; Guo *et al.* 2018), but large-scale studies can now be carried out

in most non-model organisms, providing opportunities for novel insights into evolutionary dynamics in the wild (see Jones *et al.* 2012; Brawand *et al.* 2014; Huber *et al.* 2015).

Island archipelagos provide replicated, ecologically variable and simplified landscapes, that can greatly facilitate the study of adaptation in the wild (Losos and Ricklefs 2009; Warren *et al.* 2015). Varying abiotic environments, combined with independent evolutionary histories of inhabiting organisms, result in islands harbouring unique ecological communities (Illera *et al.* 2012; Rominger *et al.* 2016; Lawson and Petren, 2017). The distinct geographical and ecological structure of individual islands, combined with the barrier to gene flow provided by the ocean, enables hierarchical population structure to develop over time, and for local adaptation to occur (Clegg *et al.* 2002). When combined with the large-scale genomic marker sets that can now be generated (e.g., Huber *et al.* 2015; Enciso-Romero *et al.* 2017), island systems provide an excellent opportunity to tease apart the roles of selection, drift and gene flow in shaping patterns of genetic diversity in nature (Lamichhaney *et al.* 2015; Salces-Castellano *et al.* 2020).

Selection operates at a range of geographic scales within and across island archipelagos (Losos and Ricklefs 2009). Studying very fine landscape-scale adaptation within island populations may reveal ecologically relevant and rapid adaptation, but may be limited to detecting very strong signatures of selection (Ray and King 2006; Milá *et al.* 2010; Bertrand *et al.* 2016; Izuno *et al.* 2017). Consequently, studies of fine-scale local adaptation may be biased towards detecting phenotypes determined by genes of large effect, while smaller effect loci or highly polygenic phenotypes are likely to go undetected (Delahaie *et al.* 2017). Furthermore, local adaptation may be transient, as a temporary response to fluctuating selection pressures, and therefore patterns of adaptation at one timepoint may not be relevant to longer-term evolutionary processes. In contrast, selection can also be studied at broad spatial scales, among island archipelagos between which there has been long-term isolation and limited gene flow (Ellegren *et al.* 2012; Manel *et al.* 2012). Such studies may reveal patterns of strong adaptive evolution (Pilot *et al.* 2014), but signatures of selection may be eroded by subsequent evolutionary forces including mutation, drift and gene flow that accumulate over time (Lenormand 2002; Tigano and Friesen 2016). It has long been recognised that consideration of spatial scales is important when identifying patterns and drivers of adaptation among populations, but few studies have quantified adaptation across a range of scales. In particular, it is important not to neglect intermediate scales (e.g., between populations on closely located islands with recent divergence histories and/or potential for gene flow) when studying adaptation, as these may provide powerful systems with which to detect ecologically relevant adaptations.

Berthelot's pipit (*Anthus berthelotii*) is a Macaronesian endemic passerine distributed across three north Atlantic archipelagos (Fig. 1). Previous research suggests that this species initially colonised the Canary Islands from mainland Africa approximately 2.5 million years ago (Voelker 1999), before dispersing independently from the Canary Islands to both the Selvagens and the Madeiran archipelago, approximately 8,500 years ago (Spurgin *et al.* 2014). These founder events resulted in population bottlenecks and reduced population size across the northward colonised archipelagos, with a subsequent absence of gene flow between archipelagos (Illera *et al.* 2007; Spurgin *et al.* 2014; Armstrong *et al.* 2018). Founder effects across archipelagos appear to shape genetic and morphological divergence of populations at broad scales (Spurgin *et al.* 2014; Armstrong *et al.* 2018). Little is known about divergence, or levels of migration, between populations within archipelagos. Selective pressures including diseases and climatic factors vary greatly across these populations, both at broad geographic scales between archipelagos and at finer geographic scales between and within islands (Illera *et al.* 2016). For example, pathogen prevalence (i.e., *Avipoxvirus, Leucocytozoon* and *Plasmodium*) varies greatly among islands within archipelagos. Both the Canary Islands and Madeiran archipelago have populations with both high and low pathogen loads, and population level patterns of pathogen prevalence are consistent over time (Spurgin *et al.* 2012). Broad-scale balancing selection appears to have maintained variation at an important immune gene family, the major histocompatibility complex (MHC) across archipelagos (Spurgin *et al.* 2011). Selection also operates over very fine spatial scales within this system, with previous work having identified landscape-level environmental drivers of pathogen distribution and immunogenetic variation within specific islands (Gonzalez-Quevedo *et al.* 2014; Gonzalez-Quevedo *et al.* 2016; Armstrong *et al.* 2019). Climatic conditions, and rainfall in particular, also vary strongly between western and eastern Canary Islands, and between Madeiran islands (Cropper and Hanna 2014). Thus, the Berthelot's pipit system provides an excellent framework with which to investigate how natural selection operates across different spatial scales in nature.

A recent study by Armstrong *et al.* (2018) used a genome-wide set of markers to investigate genetic variation and selection at broad spatial scales across the Berthelot's pipit system - specifically between the three archipelagos. Analysis showed strong genetic structure among, but not within, archipelagos, while a genome scan to identify loci under selection between archipelagos identified candidate genes associated with bill morphology, immunity and adaptation to climate (metabolism). However, we do not yet understand i) patterns of colonisation, gene flow and drift within archipelagos, and ii) whether the same loci and/or traits showing divergent selection between archipelagos are also under selection between

recently separated populations within archipelagos. Such information will provide useful insight into how selection operates across different spatial scales in this and other systems.

Here, we use genomic approaches to investigate population history and genetic diversity across island populations of Berthelot's pipit, and test for signatures of selection between recently separated populations within archipelagos. We use genome-wide restriction-site associated DNA sequenced (RAD-seq) markers from across the Berthelot's pipit range to address the following questions: (1) What new insights do analyses of genomic variation provide for population history, including colonisation, bottlenecks and gene flow across the species range? (2) Can we detect signatures of selection across recently diverged populations within archipelagos? (3) What are the loci, and traits, under selection within archipelagos? and (4) Are the same loci under selection within and across archipelagos? To address these questions, we first use population genetic analyses to examine colonisation history and gene flow across the species range. We also quantify population structure and genetic diversity within the Canarian and Madeiran archipelagos independently, providing a finer-scale assessment of genetic structure compared to previous studies in this system. We then use genome scan approaches to identify loci under divergent selection within archipelagos, and where appropriate link patterns of genetic variation to variation in phenotypic traits. Finally, we compare our results to previous research on this species, to help understand how population history, selection and drift interact to shape patterns of diversity at different scales across island populations.

3 Methods

3.1 Population sampling and sequencing

Berthelot's pipits were sampled on 12 islands across their geographical range (Fig. 1), as reported in detail by Illera *et al.* (2007) and Spurgin *et al.* (2012). As in Armstrong *et al.* (2018), we consider the pipits inhabiting El Teide mountain plateau of Tenerife (>2,000m above sea level) as a separate population to that inhabiting the island's lowlands due to their separation by a wide strip of forest vegetation on the mountain side which the pipits do not inhabit. Individuals were sampled widely across the populations, reducing the probability of sampling closely related individuals, and caught using spring traps baited with mealworm larvae (*Tenebrio molitor*). A blood sample (ca. 25ul) was taken from each bird by brachial venipuncture and stored in 800μl absolute ethanol at room temperature. DNA was extracted

using the salt extraction protocol described by Richardson *et al.* (2001), and birds were molecularly sexed (Griffiths *et al.* 1998). Seven morphometric measurements were taken; weight, wing, head and tarsus length, and bill height, length and width. Each individual was fitted with a colour or metal ring to prevent resampling of the same individuals. Birds were released unharmed at the point they were captured. Twenty putatively unrelated individuals were selected from each population (22 from the lowland Tenerife population) for ddRAD-seq, with efforts made to equalise the sex ratio within each population sample (Armstrong *et al.* 2018).

The initial ddRAD library was generated using the protocol by DaCosta and Sorenson (2014) which assigns RAD reads to samples based on an 8 bp barcode sequence and retains the read with the highest quality score. Loci that could not be confidently genotyped in more than four samples and those where 10% or more calls were missing or ambiguous were treated as missing data in the "Berthelot's" library. The "All Pipits" dataset containing all Berthelot's pipit and tawny pipit samples was filtered to contain SNPs from RAD loci that were successfully genotyped in 100% of individuals, removing loci that contained SNPs with >2 alleles. RAD loci were mapped to the Zebra finch genome (*Taeniopygia guttata*; v. 3.2.4; Warren *et al.* 2010). The data included multiple SNPs originating from the same RAD loci (throughout, marker names refer to distance in bp from the start of the RAD tag: "Locus number – s – bp from start").

The Berthelot's marker sets were first grouped by archipelago and then trimmed using Plink 1.9 (Chang *et al.* 2015) to remove sex-linked loci and SNPs with low minor allele frequency (MAF <0.03) with the aim of removing exceptionally rare variants within archipelagos while retaining a large marker set (MAF threshold reviewed by Linck and Battey 2019). We used Plink and GCTA v1.91.7 (Yang *et al.* 2011) to calculate genetic relatedness between each pair of individuals (dyad) for each of the populations. GCTA relatedness values were strongly correlated with those calculated by Plink (Pearson correlation; r = 0.92, 0.98 and 0.96 for the Canary Islands, Madeiran archipelago and Selvagens, respectively), so we only report the Plink calculated values. Using these (Fig. S1), one individual from any pair identified as having a relatedness value >0.2 was randomly removed to avoid first and second order relatives being included in the population genetic and selection analyses.

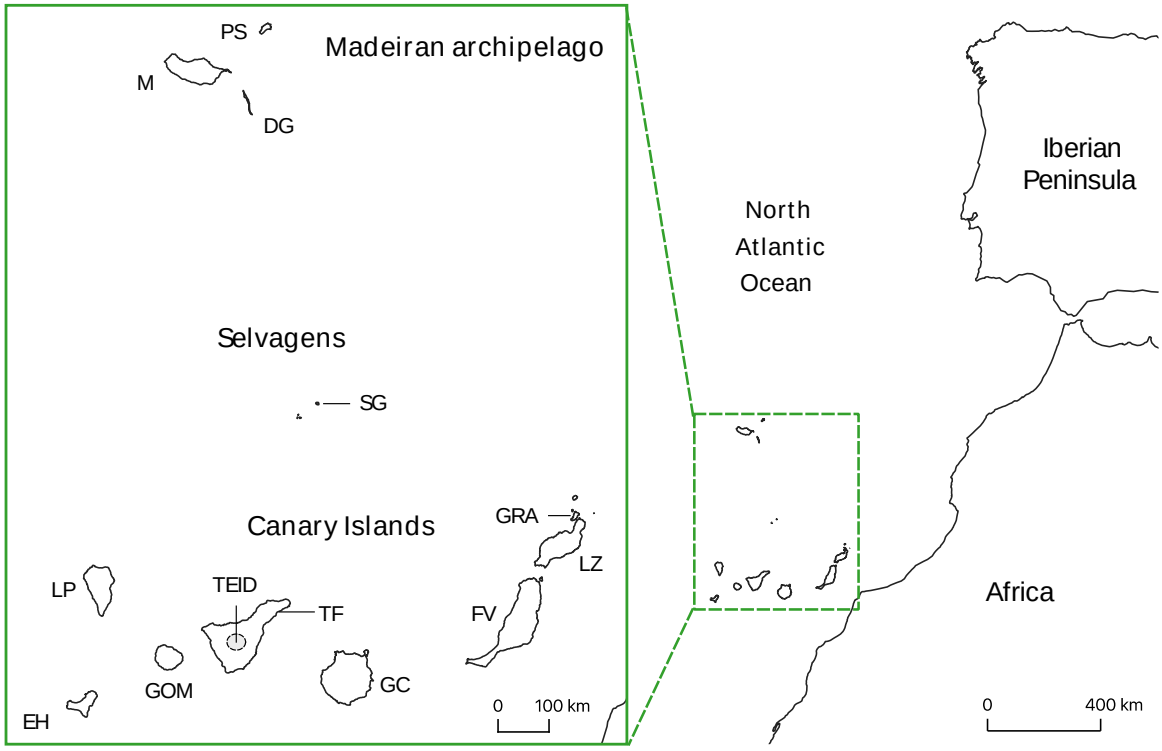


Figure 1. Locations of Berthelot's pipit populations used in the current study. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID) mountain population of Tenerife, Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ) and La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagem Grande (SG), Selvagens archipelago.

3.2 Inferring population divergence, admixture and genetic diversity

Population genetic analyses were carried out using two datasets to determine patterns of colonisation and gene flow among populations across the species range (Fig. 1). Strong population structuring exists between archipelagos of the Berthelot's pipit supporting our previous inference of absence of contemporary gene flow at broad scales in this system (Armstrong *et al.* 2018). We have reported weak east-west population structure between populations within the Canary Islands (Armstrong *et al.* 2018), but it is unknown whether this is due to contemporary gene flow between closely located islands or weak population divergence since colonisation. We implemented *TreeMix* at these different population scales with the aim of further understanding the evolutionary processes behind the patterns of

population structuring we see. First, we used an "All Pipits" dataset which, in addition to the Berthelot's pipits includes 16 tawny pipits (*Anthus campestris*), the Berthelot's pipit sister species (Voelker 1999), sampled from north east African and Spanish populations. The tawny pipit was used to root divergence from the mainland across the three Macaronesian archipelagos, to gain insight into the earliest colonised islands. The "All Pipits" dataset was processed bioinformatically as in Armstrong *et al.* (2018) providing 8927 polymorphic loci across the 262 Berthelot's pipits and 16 tawny pipits; these samples were collected in Spain (n = 11), Mauritania (n = 4) and Morocco (n = 1). These data were trimmed to remove any ambiguously genotyped loci, excluding loci with multi-allelic SNPs. For *TreeMix* analyses, individuals with high pairwise relatedness were removed as in the Berthelot's tree, while sex-linked loci were removed and a reduced MAF threshold of 0.01 was applied to retain a large enough SNP set, while removing many variants unique to the tawny pipits only (Table S1). These trimming steps retained 1650 SNPs across all Berthelot's and tawny pipit populations. *TreeMix* was run using allele frequencies averaged across windows of *k* 100 SNPs to account for LD. We also investigated the tree without MAF filtering, which retained many more SNPs, but was dominated by variation within the tawny pipit and hence we do not report these results further. Second, we used the "Berthelot's" data set, as described above, which includes only Berthelot's pipits. This data set provides a greater number of polymorphic loci within the Berthelot's pipit populations due to lower within-species divergence which may enhance the ability to detect population splits, migration and drift among populations. For this analysis, we trimmed the marker sets across all 13 populations using Plink, to remove closely related individuals (as above) and loci with MAF <0.03, loci in strong linkage disequilibrium (LD) (>0.4 r^2 threshold, for a sliding window 50 kb with 10 marker step) and sex-linked loci (Table S1).

Using *TreeMix* v 1.13 (Pickrell and Pritchard 2012), we inferred a tree in which populations (i.e., one population per island except in Tenerife with two populations, one in the lowlands and one in the highlands) may maintain gene flow after they split from a common ancestor. This method first infers a maximum-likelihood tree from genome-wide allele frequencies and then identifies populations with poor fit to this model (populations with residuals deviating strongly from zero); migration events involving these populations are added in order to improve the likelihood of the model. Allele frequencies for *TreeMix* analysis were calculated within populations using Plink, after marker pruning (Table S1). We modelled several scenarios allowing zero to eight migration events (Table S1), discounting migration events when the relative increase in model likelihood was <1%. For each analyses 10,000 bootstrap replicates were generated, resampling blocks of 20, 50 and 80 SNPs to evaluate the robustness of the tree topology; this corresponds to a window size of approximately 10-30

Mb as used by Pickrell and Pritchard (2012). The total fraction of the variance explained by each model was estimated with the 'get_f()' R function, in *TreeMix*. Residual plots were assessed to display model fit and identify poorly fitted population pairs. F_{ST} was calculated between pairs of populations in Plink (Weir and Cockerham 1984) using the genome-wide RAD dataset as trimmed for the "All Pipits" and "Berthelot's" tree to support *TreeMix* tree topology. To test for admixture among Berthelot's pipit populations, we computed the three-population statistic (f_3 statistic; Patterson *et al.* 2012) for all population triplets through software threepop (Reich *et al.* 2009) implemented in *TreeMix*, jackknifing over blocks of 50 SNPs. An observed negative value of the f_3 statistic and Z-score <-2 are indicative of historical admixture (Reich *et al.* 2009).

We next investigated fine-scale population genetic structure between recently separated populations within archipelagos. From the initial Berthelot's RAD library, we generated separate Canary Islands and Madeiran archipelago marker sets prior to trimming to maximise the number of loci at each level of clustering within archipelago datasets (Table S1). As we only sampled one population in the Selvagens, no within-archipelago analysis was conducted for this archipelago. These data were also trimmed to remove SNPs with MAF <0.03 (PCA was filtered according to MAF (SNPs with MAF < 0.03 excluded) within archipelagos, LD analysis was MAF-filtered (SNPs with MAF < 0.03 excluded) within populations), and closely related individuals were removed (as above). LD summarises both mutational and recombination history, whereby larger, more outbred populations show rapid decay of LD between genetic markers compared to small inbred populations (Flint-Garcia *et al.* 2003). Patterns of LD have been used extensively to detect historic fluctuations in population size (N_e) and founder events in humans (Reich *et al.* 2001; Reich *et al.* 2009), selectively bred species such as Chinese Merino sheep, Xinjiang type (Liu *et al.* 2017) and wild species including European grey wolves, (*Canis lupus;* Pilot *et al.* 2014) and village dogs (*Canis lupus familiaris;* Shannon *et al.* 2015). The relationship between proximate SNPs reflects historic N_e, and LD at distant SNPs reflects N_e in more recent time. To further understand patterns of genetic diversity and population size in the Berthelot's pipit, we estimated LD for each island population using Plink. The r^2 values were compared to physical distance between loci for all pairs of SNPs situated on the same chromosome. We fitted a locally weighted linear regression (loess) curve to the r^2 data using the R function 'loess' using the default span parameter (0.75), with 95% confidence intervals calculated. Population structure was examined within the Canary Islands and Madeiran archipelago independently using a principal component analysis (PCA), implemented using Plink, based on the trimmed and filtered marker sets.

3.3 Genome scan for signatures of selection within archipelagos

For genome scan analyses, using the archipelago level marker sets, close relatives and SNPs with a MAF <0.03 were removed (as above), but we did not filter based on LD, which enabled us to identify and visualise genomic regions under selection (Table S1). We used EigenGWAS (Chen *et al.* 2016), implemented in the program GEAR (www.github.com/gc5k/GEAR/wiki), to identify loci consistent with selection within archipelagos. EigenGWAS performs a PCA to generate gradients of population structure, then assesses each genetic marker individually for an association with these axes. EigenGWAS provides genomic inflation factor corrected *P* values (λ_{GC}) – with the significance threshold determined by Bonferroni- correction - to control for genome-wide population stratification and drift. Loci above this significance threshold are putatively under selection across the gradient of population structure (see PCAs, Fig. 3). We also calculated F_{ST} for each SNP, using all SNPs that had passed the trimming stages (Table S1). All SNPs in the Madeiran subset were also in the Canary Island dataset as a result of reduced genetic diversity in the Madeiran archipelago, and hence, direct comparisons of SNP variation are made. To identify genes located near outlier SNPs, we viewed regions of interest using the Zebra finch genome (v. Taeniopygia_guttata-3.2.4) in NCBI Genome Data Viewer v. 4.8. (www.ncbi.nlm.nih.gov/genome/gdv/browser).

After having identified candidate SNPs which may be associated with skeletal development (see Results/Discussion), we determined the genotype-phenotype associations for these loci across all Berthelot's pipit populations. We used Linear Mixed Models (LMMs) implemented in R using the lme4 package v1.1.15, to test the hypothesis that SNP variation within the candidate gene is associated with phenotypic variation in morphological traits. To check whether skeletal development is associated with genotypic variation at these candidate SNPs, we fitted LMMs with wing, tarsus and head length, bill length, width and height and weight as dependant variables. Population was modelled as a random effect nested within-archipelago, and genotype (number of copies of the minor allele), sex and age as fixed effects in the model. Separate models were fitted for each of the SNPs within the morphology associated gene. All estimates are reported with associated 95% confidence intervals.

4 Results

The initial RAD library provided 9960 genome-wide polymorphic SNPs across the entire geographical population range of Berthelot's pipit. After MAF filtering separately in each archipelago, we retained 4470 SNPs in the Canary Islands subset and 2938 in the Madeiran archipelago subset. For the *TreeMix* Berthelot's tree, we trimmed all 13 Berthelot's pipit populations together from the initial RAD library, and after MAF and LD filtering retained 2850 loci across the Canary Islands, Madeiran archipelago and Selvagens.

Relatedness varied within and among populations: whilst most pairs of individuals showed low relatedness (r <0.05), there were some pairs with high relatedness sampled in the smallest isolated populations of the Canary Islands (El Hierro and La Graciosa) and the islands of Madeira and Deserta Grande in the Madeiran archipelago (Fig. S1). No pairs of individuals had relatedness of r >0.2 in the Selvagens. To avoid including closely related individuals in the population genetic and selection analyses, one individual was removed from each dyad with a relatedness score r >0.2, resulting in three individuals from La Graciosa, one from Lanzarote, two from El Hierro, three from Madeira and seven from Deserta Grande being removed.

4.1 Population genetic analyses

We used *TreeMix* to produce maximum likelihood trees of divergence and gene flow. The tawny pipit rooted tree showed strong divergence between mainland tawny pipits and the contemporary Berthelot's pipit populations (F_{ST}; Canary Islands 0.36– 0.38, Madeiran archipelago 0.42–0.44 and Selvagens 0.46), with shortest branch lengths to the central Canary Islands (Fig. S2A). Populations in the Madeiran archipelago had poor residual fit with the tawny pipit (Fig. S2), suggesting that these populations may be more closely related than is presented by the best-fit tree. However, adding migration events did not improve the residual fit of these models. Despite this, the tawny pipit rooted tree, obtained without adding migration events, explained the majority of allele frequency variation (variance = 99.55%) between populations.

Maximum likelihood trees limited to the 13 Berthelot's pipit populations placed the Madeiran and Selvagens archipelagos on long, independent branches, grouping with the central/eastern and western Canary Islands, respectively (Fig. 2A). Tree topology within the

Canary Islands roughly reflects geographic distance between islands, with branches east and west of Tenerife, and Mount Teide as a separate branch point from lowland Tenerife (Fig. 2A). Generally weak drift is observed across the Canary Islands, with the longest within-archipelago branch lengths in El Hierro and La Graciosa. These same patterns are reflected in pairwise F_{ST} values, with moderate divergence between El Hierro and La Graciosa (F_{ST} = 0.04) the most geographically distant pair of populations, and increasingly weaker divergence between more closely located populations in the Canary Islands, especially in the central islands in the archipelago (F_{ST} range = 0.01–0.03). In the Madeiran archipelago, longer branch lengths were observed, with the highest divergence between Deserta Grande and the other Madeiran islands (F_{ST}; Porto Santo = 0.06, Madeira = 0.05). Tree topology was broadly robust to window size, but there were minor differences within the Canary Islands including the source populations for the Madeiran archipelago (Fig. S3); here we present specific model results calculated using windows of 50 SNPs. The majority of allele frequency variation (variance = 99.86%) is described solely by the tree topology, with good residual support for most populations (Fig. S4). Sequentially adding migration events did not substantially improve model support (Fig. S5) or the degree of variance explained by trees (increase in variance = 0.11% after five migration events added; Fig. S6). Weakly negative f_3 statistics (>- 5.3 x10^{-4}) and Z-scores >-1.3 were found for Tenerife (including El Teide), Gran Canaria, Fuerteventura and Lanzarote in the Canary Islands and the island of Madeira when in a three-way population comparison (Table S2). F_3 results suggest few admixture events subsequent to branch divergence may have occurred between Madeira and Porto Santo and Fuerteventura/Lanzarote and La Graciosa. This is consistent with geographic distance between islands, suggesting no admixture between geographically distant populations.

Patterns of linkage disequilibrium (LD) within populations are shown in Figures 2B and S7. LD was highest in the smallest and most isolated populations across the Berthelot's pipit range. Thus, across all populations, the Selvagens had the highest LD with a long-range decay pattern; LD was lower across the three Madeiran islands, and was lowest in the Canary Islands, especially in large central islands (Fig. 2B).

To investigate potential-fine-scale population structure within archipelagos, we conducted a PCA of individuals using the archipelago datasets. Within the Canary Islands, we found that the first principal component roughly reflects an east-west gradient of population structure, with El Hierro and La Graciosa separating most distinctly from the other islands (Fig. 3A). The second component separates El Hierro, Lanzarote and La Graciosa from the other islands. PCA of the Madeiran archipelago separated Deserta Grande from Madeira and

Porto Santo along the first axis, and separated Porto Santo from Madeira on the second axis, with a weak gradient from Madeira to Deserta Grande to Porto Santo (Fig. 3B).

4.2 Genome scan to identify loci under selection within archipelagos

We used EigenGWAS analyses to identify loci under divergent selection across the gradients of population structure seen in the Canary Islands and Madeiran archipelago, separately (see Fig. 3). Across the Canary Islands, genomic inflation factors (λ_{GC}) were substantial for both PCs (PC1 = 2.4 and PC2 = 2.1, where a value >1 indicates population structuring; Hinrichs *et al.,* 2009). Correcting for this and using a Bonferroni-corrected significance of $P < 1.12 \times 10^{-5}$ (n SNPs = 4470) for the Canary Islands, we detected one outlier SNP for PC1 ($P = 3.56 \times 10^{-9}$, Fig. 4A & Fig. S8A). No outlier SNPs were detected for PC2 (Fig. S9A & B). For the Madeiran archipelago, after correcting for the λ_{GC} of 3.4 for PC1 and 1.4 for PC2, three outlier SNPs exceeded our Bonferroni-corrected significance threshold of $P < 1.70 \times 10^{-5}$ (n SNPs = 2938). Two of these were on the same RAD locus, within 20 bases of each other (Fig. 4B, S8B & Table 1). No outlier loci were detected for PC2 across Madeira (Fig. S9C & D). Allele frequencies for all loci detected in the EigenGWAS analyses are reported in Table S3.

Locus F_{ST} values were not correlated between archipelagos (Spearman's-rank Correlation, $r = 0.031$, $P = 0.184$), but significant SNPs from the EigenGWAS analyses had the highest F_{ST} values (Fig. 5). The SNP detected by EigenGWAS as being under selection across the Canary Islands (219s24), had a high MAF in the western islands of El Hierro and La Palma, and a low MAF across the central and eastern islands within that archipelago (Table S3). All SNPs under selection across the Madeiran archipelago had near 50% prevalence of the "minor allele" in the Deserta Grande population while being absent from the two other islands, with a low frequency of the minor allele observed across the Canary Islands and Selvagens.

We were able to map all significant EigenGWAS SNPs to the Zebra finch genome and determine their likely genomic location (Table 1 & Fig. S10). The two closest genes to the Canary Island SNP, 219s24, are *WDHD1* and *GCH1*, which are involved in DNA binding/repair and enzyme synthesis, respectively (Table 1). This SNP maps to chromosome 5, with *WDHD1* 2,071 bases upstream and *GCH1* 6,252 bases downstream. In the Madeiran archipelago, the two significant outlier SNPs in the same RAD locus (1585s94

& 1585s112) mapped to intronic regions of a candidate gene for morphology, *ADAM12*, on chromosome 6 (see Discussion; Table 1). The third SNP, 790s54, was not close to a gene (closest gene 70,799 bp downstream).

Table 1. Outlier SNPs identified by EigenGWAS analyses across the Canary Islands and/or Madeiran archipelago populations of Berthelot's pipit (see Fig. 4). Genes within 10,000 bp of the SNP are identified. Relative positions of the candidate genes are stated in bp Upstream (US) or Downstream (DS) from the SNP site.

SNP	*P* Canary Islands/ *P* Madeira	F_{ST} Canary Islands/ F_{ST} Madeira	Genomic location (bp)	Candidate Gene(s)	Gene product	Trait
219s24	3.56×10^{-9} / -	0.23 / 0	Chr 5: 58990367	*WDHD1* (2,071 US)	WD repeat and HMG-box DNA binding protein 1	DNA binding and repair (Hsieh *et al.,* 2011).
				GCH1 (6,252 DS)	GTP cyclohydrolase 1	Rate-limiting enzyme for tetrahydrobiopterin (BH4), a vital cofactor and modulator of peripheral neuropathic and inflammatory pain (Tegeder *et al.,* 2006).
1585s94	0.61 / 1.59×10^{-6}	0.02 / 0.56	Chr 6: 33504910	*ADAM12* (In gene)	Disintegrin and metalloprotease domain 12	Body size by affecting bone/cartilage development (Tokumasu *et al.,* 2016).
1585s112	0.65 / 1.59×10^{-6}	0.02 / 0.56	Chr 6: 33504928	*ADAM12* (In gene)	Disintegrin and metalloprotease domain 12	Body size by affecting bone/cartilage development (Tokumasu *et al.,* 2016).
790s54	0.31 / 1.16×10^{-5}	0.01 / 0.47	Chr 24: 718487	–	–	–

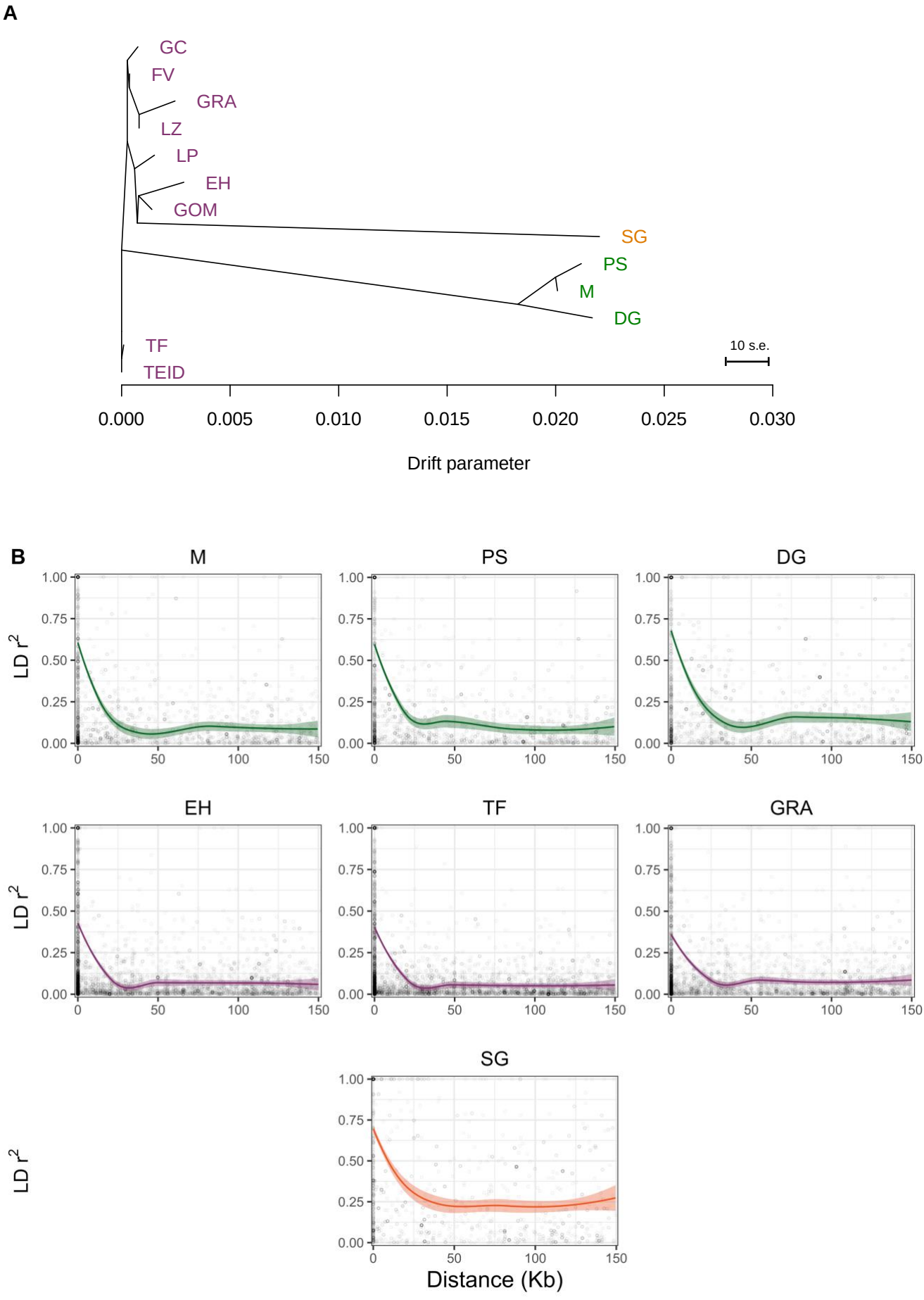

A
GC
FV
GRA
LZ
LP
EH
GOM
SG
PS
M
DG
TF
TEID
10 s.e.
0.000
0.005
0.010
0.015
0.020
0.025
0.030
Drift parameter
B
M
PS
DG
EH
TF
GRA
SG
LD r2
1.00
0.75
0.50
0.25
0.00
0
50
100
150
Distance (Kb)

Figure 2. Evolutionary relationships between island populations of the Berthelot's pipit. **A)** Maximum-likelihood bifurcating tree of population history - without subsequent gene flow - across the pipit colonisation range as inferred by *TreeMix*. The branch length scale bar shows ten times the average standard error in the covariance matrix of ancestry. **B)** The relationship between linkage disequilibrium and base-pair distance for SNPs across each Madeiran Island (green), three Canary Island populations (purple) and the Selvagens (orange); Tenerife, a central island assumed to be a large outbred population with low within-archipelago divergence (see Fig. 3A), and El Hierro and La Graciosa populations which have long branch lengths and strongest within-archipelago genome-wide divergence. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID), Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ) and La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). The fit lines show a local regression model, with a shaded band indicating 95% confidence intervals.

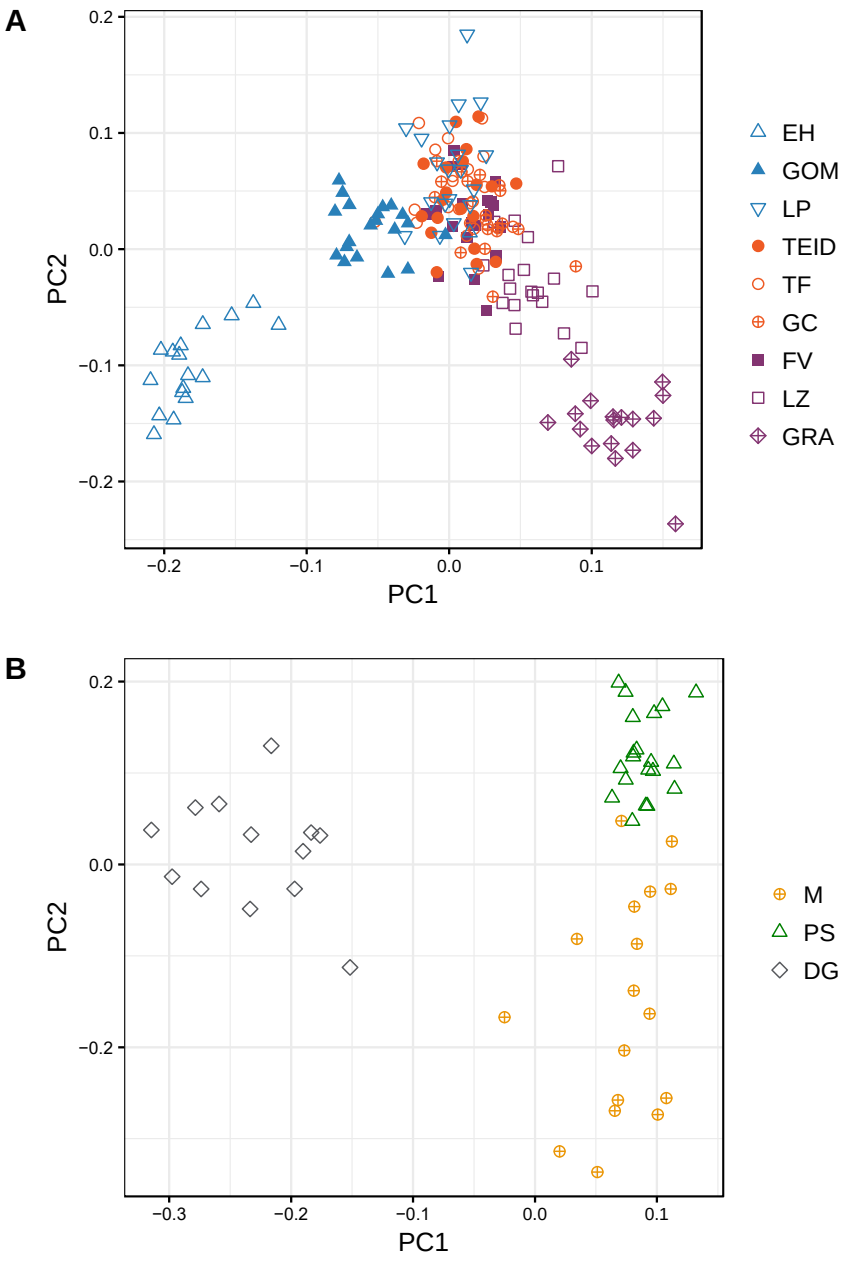


Figure 3. Population structure based on genome-wide ddRAD SNPs among Berthelot's pipit populations separately across the Canary Islands and Madeiran archipelago. **A)** Principal component analysis (PCA) across the Canary Island populations. PC1 and PC2 explained 2.7% and 2.3% of genomic variation, respectively. **B)** PCA of Madeiran archipelago populations; PC1 = 3.5%, PC2 = 1.6% of genomic variation explained. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID), Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ) and La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG).

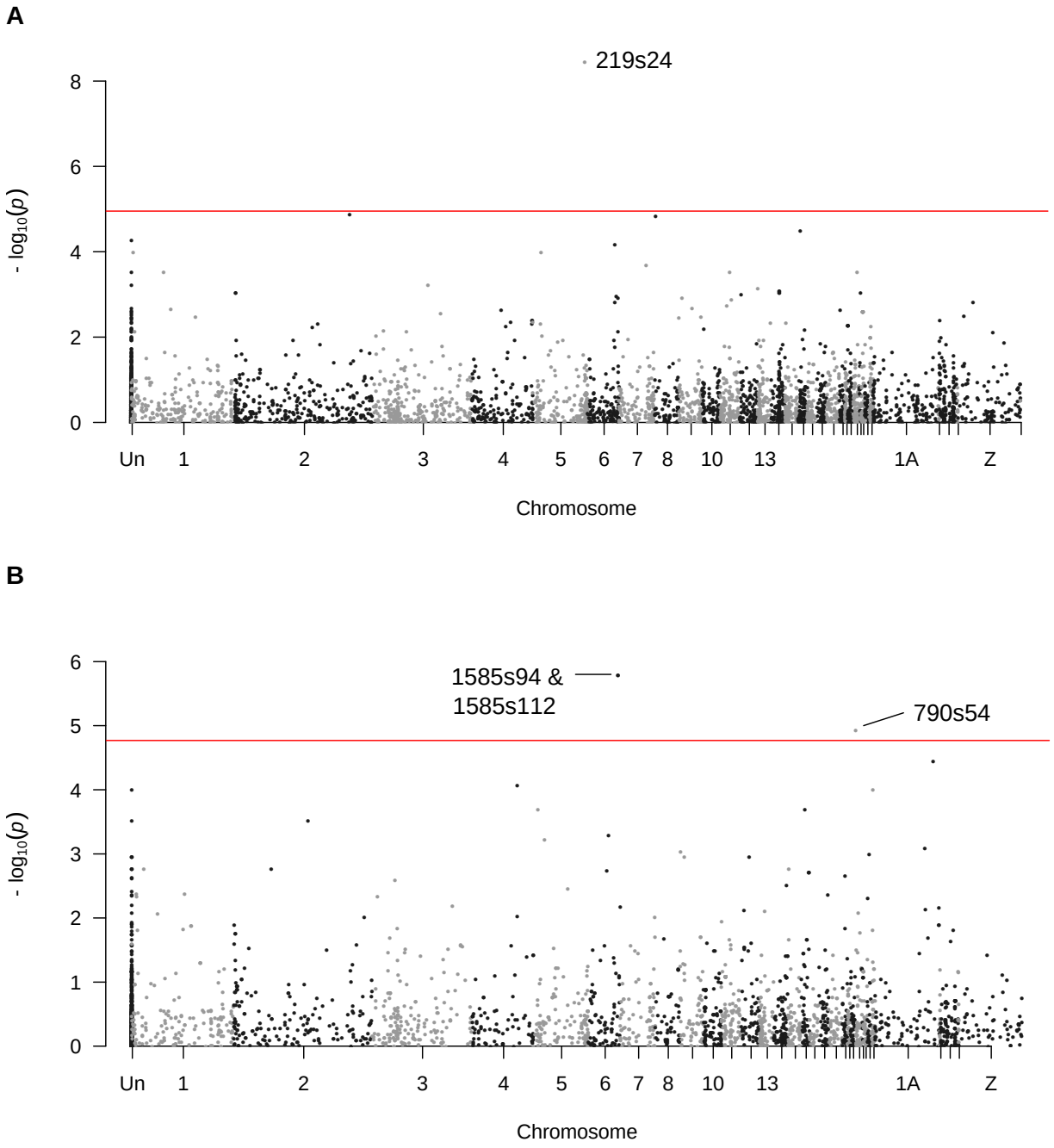


Figure 4. Manhattan plots of EigenGWAS analyses based on genome-wide ddRAD SNPs among Berthelot's pipit populations within archipelagos. **A)** Canary Islands PC1, clustering east-west geographic gradient, as seen in Fig. 3A. Horizontal red line indicates Bonferroni-corrected significance of P <1.12 x10^{-5} based on 4470 genome-wide ddRAD SNPs. **B)** Madeiran archipelago PC1, separating Deserta Grande from Madeira and Porto Santo islands, as seen in Fig. 3B. Horizontal red line indicates Bonferroni-corrected significance of P <1.70 x10^{-5} based on 2938 genome-wide ddRAD SNPs. Unmapped SNPs are recorded as "Un", and alternate black-grey colouring indicates chromosomal limits.

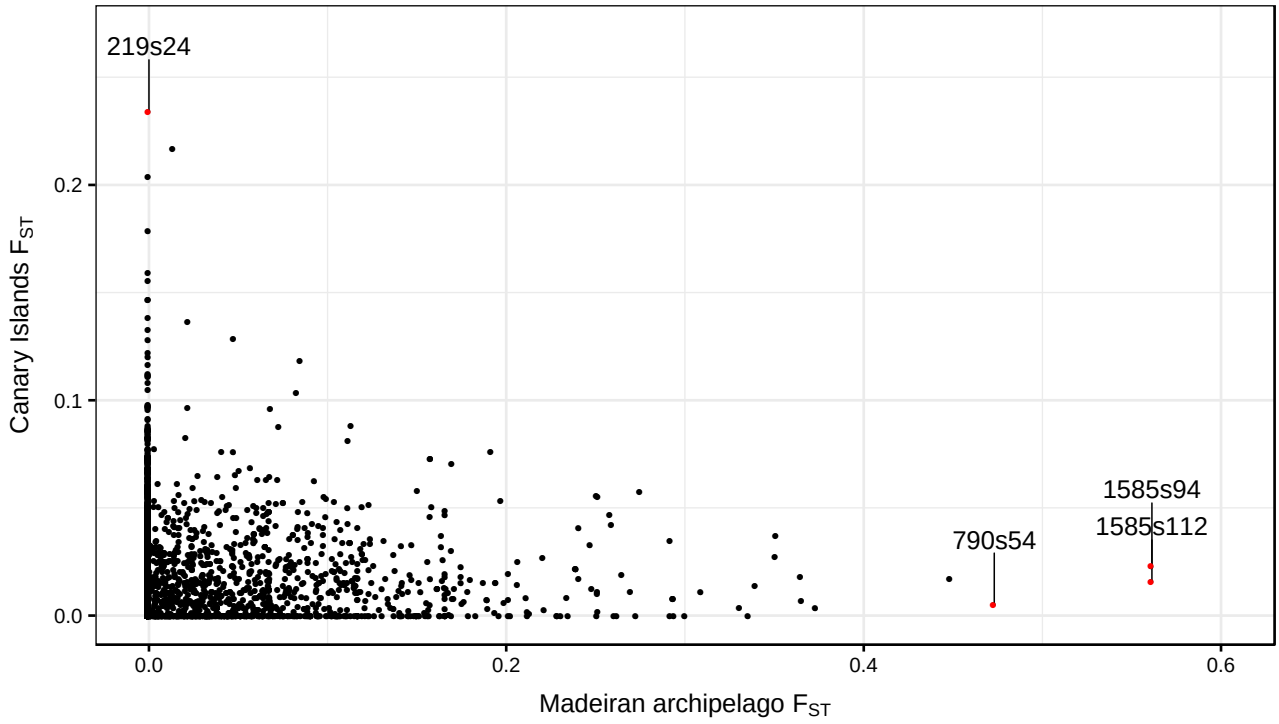


Figure 5. Within-archipelago genetic differentiation of genome-wide ddRAD SNPs among Berthelot's pipit populations. Points represent F_{ST} of 3531 mapped genome-wide ddRAD SNPs between populations, across the Canary Islands and the Madeiran archipelago. SNPs identified by the EigenGWAS analysis for the Canary Islands archipelago (y axis) and Madeiran archipelago (x axis), are highlighted in red and labelled with their SNP code. *No unmapped SNPs had F_{ST} > 0.16 or > 0.40 for the Canary Islands and Madeiran archipelago, respectively.

4.3 Genotype-phenotype association across populations

ADAM12 has been shown to play a role in skeletal development and is therefore a potential candidate for being associated with morphology (Table 1, see Discussion). Using LMMs, we tested how variation at this locus was related to candidate morphology traits across all pipit populations. To determine what effect candidate SNP variation may have on morphology, we tested for genotype associations with wing, tarsus and head length, weight and bill length, width and height. Genotypes for SNP 1585s94 and 1585s112 within the *ADAM12* gene were strongly colinear (R^2 >0.982). We found a similar effect of genotype on head length at both SNPs putatively under selection within the *ADAM12* gene (Gaussian LMM, SNP 1585s112 estimate ± s.e. = -0.39 ± 0.13, P = 0.003; R^2 = 0.75; Fig. S11 & SNP 1585s94 estimate ± s.e. = -0.36 ± 0.13, P = 0.006; R^2 = 0.75) as well as strong differences between the sexes (P <3.3 x10^{-6}). Homozygous individuals for the minor allele were only

detected in the Deserta Grande population for both of the SNPs, while heterozygous individuals were present at low frequency for eight of the 13 pipit populations (Fig. S11 & Table S3). Genotype was not significantly associated with beak morphology variables (bill length, height or width), weight, wing length or tarsus length, although there were differences between sexes for tarsus length (P <0.003), wing length (P <0.002) and bill length (P <0.010) as we expect for a sexually dimorphic species.

5 Discussion

We examined genetic divergence and selection between recently founded island populations in an attempt to understand population history and uncover traits of adaptive importance across selective environments in the wild. Using RAD sequenced markers generated for 13 populations of Berthelot's pipit, we first analysed genome-wide variation to uncover patterns of colonisation, admixture and population demography. Our analyses support the establishment of Berthelot's pipit across the three archipelagos via independent colonisation events, with evidence of weak subsequent gene flow between populations. Patterns of genetic diversity are consistent with signatures of founder events and geographic isolation. We applied a genome scan approach to identify signatures of selection, and inferred traits of ecological importance between recently separated populations within archipelagos. We detected SNPs putatively under selection within the Canarian and Madeiran archipelagos, but found no overlap between candidate SNPs identified from previous analyses at a broader spatial scale i.e., between archipelagos (Armstrong *et al.* 2018). We found evidence for selection at SNPs associated with head length across the Madeiran islands, and for a SNP located between candidate genes involved in the regulation of DNA repair and enzyme pathways across the Canary Islands.

Previous studies have used microsatellites to examine modes and patterns of population divergence across the three archipelagos colonised by the Berthelot's pipit (Illera *et al.* 2007; Spurgin *et al.* 2014), while more recent studies have used genome-scale analyses to examine broad-scale population structure and bottlenecks between archipelagos (Armstrong *et al.* 2018). Here, we complement these findings by inferring colonisation and gene flow at different geographic scales using *TreeMix*, and examine population-level patterns of genetic diversity using LD decay. *TreeMix* shows strong divergence between the tawny and Berthelot's pipit (Fig. S2A), consistent with phylogeny-based estimates (Voelker 1999). Tree topology suggests initial divergence of Berthelot's pipit from the tawny pipit may have been

to the central or eastern Canary Islands (Fig. S2A), with long archipelago branch lengths consistent with independent colonisation events to the Madeiran and the Selvagens archipelagos with associated bottlenecks (Fig. 2A). There are weak signatures of within-archipelago divergence and structure (F_{ST} 0.01–0.06), with the longest population branch lengths (Fig. 2A) and individual PCA clustering (Fig. 3) across the small isolated populations of El Hierro, La Graciosa and Deserta Grande relative to other within-archipelago populations. Past colonisation history and associated bottlenecks are reflected in patterns of population level LD; rapid LD decay at proximate SNPs and low long range LD indicates larger and more outbred populations across the Canary Islands, while high long range LD indicate bottlenecks and/ or inbreeding and reduced genetic diversity (Pilot *et al.* 2014; Shannon *et al.* 2015). Our patterns of LD are consistent with reduced genetic diversity in the Madeira and Selvagens archipelagos (Fig. 2B) and support previously reported patterns based on microsatellite and RAD data (see Spurgin *et al.* 2014; Armstrong *et al.* 2018). Simulation-based approaches, using a greater density of SNPs, may be useful to further confirm if our LD patterns are as a result of population bottlenecks or to add detail to the population size estimates at different historical time points. One common feature of population level LD decay is a dip in the regression between 25 and 50 kb followed by a rise. This pattern has been found in previous studies of LD in this system using the loess line fitting method (Armstrong *et al.* 2018). We are unable to determine a biological explanation for this, but alternative line fitting methods such as those used by Hill and Weir (1988) reflect our archipelago level conclusions.

Limited gene flow between island populations of the Berthelot's pipit has previously been suggested based on strong genetic structure between archipelagos (Illera *et al.* 2007; Spurgin *et al.* 2014; Armstrong *et al.* 2018), while stable host-pathogen communities within populations suggest limited movement between closely located islands (Spurgin *et al.* 2012). Here, we explicitly test for gene flow between islands and find that adding migration events between populations did not significantly improve our model of population history. Further, high LD and reduced genetic diversity in the Madeiran archipelago and Selvagens are consistent with absence of contemporary gene flow between archipelagos. These findings are reflected in other studies that investigate admixture between populations diverging at different levels of geographic separation and across differing timescales (Martin *et al.* 2013; Sendell-Price *et al.* 2020). Given that levels of divergence differ between pairs of closely located Berthelot's pipit populations, we cannot discount the possibility of weak gene flow slowing the accumulation of genetic divergence between some pairs of recently separated populations which we have been unable to detect using our marker set.

We aimed to identify signatures of selection within archipelagos to uncover ecologically relevant adaptation between recently separated island populations of Berthelot's pipit, which may be eroded by other evolutionary processes at broader scales. Applying a reverse genetics approach identified loci with patterns of variation consistent with natural selection across ecological gradients within archipelagos, while controlling for neutral genome-wide divergence due to structure. The EigenGWAS analyses detected loci under divergent selection within both the Canary Islands and Madeiran archipelago, with different SNPs detected within each archipelago (Table 1). Further, we found no evidence for a correlation between locus F_{ST} values within the two sets of archipelagos, and no markers with high levels of structure within both archipelagos (Fig. 5). This suggests that either, i) different selective pressures act across the archipelagos resulting in outliers associated with different genes or ii) selection acts on similar traits, but these traits vary in genetic architecture among archipelagos. We identified one significant candidate SNP (219s24) as putatively under divergent selection across the Canary Islands. This SNP was in a gene-dense region of chromosome 5, with *WDHD1* upstream and *GCH1* downstream. These genes act in DNA repair (Hsieh *et al.* 2011) and enzyme pathways (Tegeder *et al.* 2006), respectively, but their function in Berthelot's pipits is unknown. It is worth pointing out that while we do not detect selection for the same SNPs within archipelagos as between archipelagos, our outlier SNP in the Canary Islands, 219s24, was in the same broad genomic region on chromosome 5 as a set of SNPs previously identified as being putatively under selection and associated with bill length across the range of Berthelot's pipit (Armstrong *et al.* 2018). Further research, with a higher density of SNPs, is needed to identify the importance of this genomic region for adaptation in the pipit system. Using our marker set, we detected only four loci under selection in our analyses. We see low levels of genetic variation, and hence few SNPs, within archipelagos of the Berthelot's pipit as a result of colonisation history and bottlenecks but it is also likely that we have no marker coverage in some regions of the genome that may be under selection. Therefore, future studies are needed to uncover greater detail on the loci and hence traits that are under selection in this system.

The most significant SNPs (1585s94 & 1585s112) in the EigenGWAS analysis of Madeiran populations mapped to the *ADAM12* gene, which has been linked to body growth through skeletal development in both Zebrafish (*Danio rerio*) and Nile tilapia (*Oreochromis niloticus*), and forms part of a family of proteins involved in development, homeostasis and disease (Tokumasu *et al.* 2016; Yoshida and Yáñez 2021). Mixed model analysis, based on all samples across the Berthelot's pipit's range, revealed that the genotype at these loci was significantly associated with head length in this species. There was a low MAF (<15%) for both SNPs across the Canary Island populations and Selvagem Grande. In the Madeiran

archipelago the minor allele was absent from both Madeira and Porto Santo, but at 50% prevalence in Deserta Grande (Table S3). As discussed, colonisation of the Madeiran archipelago was from the Canary Islands, and involved a bottleneck, with a further bottleneck during the subsequent colonisation of Deserta Grande. Based on this history, there are two potential explanations as to why we see these genotype patterns across the Madeiran populations. Firstly, positive selection may have increased the frequency of the minor allele on Deserta Grande. Alternatively, the minor allele may have been lost due to purifying/negative selection, or random genetic drift, on the other Madeiran islands whilst being maintained on Deserta Grande. This second scenario is unlikely as we see greater diversity at these SNPs in the most bottlenecked population where we would expect the lowest levels of diversity. The genotype-phenotype relationship for the SNP within the *ADAM12* gene was no longer significant when two individuals (out of 19) with particularly large beaks were removed from Deserta Grande, and larger sample sizes are needed to determine the robustness of this result. Nonetheless, this research adds to the growing body of evidence that genes associated with craniofacial development may be excellent candidates for the study of natural selection in wild birds (Lamichhaney *et al.* 2015; Bosse *et al.* 2017; Lundregan *et al.* 2018).

In this study, we identified SNPs putatively under selection in recently diverged island populations of Berthelot's pipit within archipelagos. Using the same EigenGWAS approach, we previously observed a larger number of divergent selection signatures *between* archipelagos, identifying dozens of SNPs putatively under selection (see Armstrong *et al.* 2018). Our findings are similar to those seen in other studies that have used genome scans to investigate adaptation at different spatial scales in the wild. The strength of genetic differentiation and selection increases with geographic distance between Mascarene grey white-eye (*Zosterops borbonicus*; Gabrielli *et al.* 2020) and barn swallow populations, (*Hirundo rustica*; Safran *et al.* 2016) and between lake and ocean populations in brown trout (*Salmo trutta*; Meier *et al.* 2011) and Atlantic salmon (*Salmo salar*; Vincent *et al.* 2013). Adaptive divergence between geographically close populations is expected to be eroded if high gene flow between populations counteracts selection, however, an increasing number of studies show that local adaptation can persist despite gene flow (reviewed in Tigano and Friesen 2016; Moody *et al.* 2015; Tusso *et al.* 2020). Given that strong differences in pathogen prevalence, habitat and climatic conditions exist between closely related Berthelot's pipit populations - and we provide evidence of only very weak gene flow between populations - the low number of outlier SNPs within compared to across archipelagos more likely reflects weak selection between recently separated populations instead of gene flow counteracting selection in this system.

6 Conclusions

Combining the study of population history, drift and selection between island populations at different spatial scales provides an opportunity to understand how evolution shapes variation in nature. We assessed contemporary patterns of variation across the range of Berthelot's pipit, revealing that genetic diversity is largely shaped by colonisation events, with very weak evidence of gene flow between islands. We uncover outlier loci putatively under divergent selection between recently separated populations within archipelagos. Patterns of diversity at these loci, and the ecological adaptation they may be involved in, may be masked by other evolutionary processes when assessing genetic variation at broader scales. Our findings suggest natural selection may act repeatedly on traits, particularly bill morphology, at different spatial scales, and that signals of selection appear to be weaker between recently separated populations. Moving forward, studying demography and selection at a range of spatial scales is likely to prove a powerful approach for determining the strength and nature of adaptation in the wild.

7 References

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*, ***0***(August), 1–21.

Armstrong, C., Richardson, D. S., Hipperson, H., Horsburgh, G. J., Küpper, C., Percival-Alwyn, L., … Spurgin, L. G. (2018). Genomic associations with bill length and disease reveal drift and selection across island bird populations. *Evolution Letters*, ***2***(1), 22–36.

Bertrand, J. A. M., Delahaie, B., Bourgeois, Y. X. C., Duval, T., Garcia-Jimenez, R., Cornuault, J., … Mila, B. (2016). The role of selection and historical factors in driving population differentiation along an elevational gradient in an island bird. *Journal of Evolutionary Biology*, ***29***, 824–836.

Brawand, D., Wagner, C. E., Li, Y. I., Malinsky, M., Keller, I., Fan, S., … Di Palma, F. (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, ***513***(7518), 375–381.

Burri, R. (2017). Linked selection, demography and the evolution of correlated genomic landscapes in birds and beyond. *Molecular Ecology*, ***26***(15), 3853–3856.

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***(1), 1–16.

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***(1), 51–61.

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*, ***99***(12), 8127–8132.

Cropper, T. E., & Hanna, E. (2014). An analysis of the climate of Macaronesia, 1865-2012. *International Journal of Climatology*, ***34***(3), 604–622.

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

Delahaie, B., Cornuault, J., Masson, C., Bertrand, J. A. M., & Eb, C. T. H. (2017). Narrow hybrid zones in spite of very low population differentiation in neutral markers in an island bird species complex. *Journal of Evolutionary Biology*, ***30***, 2132–2145.

Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., … Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, ***491***(7426), 756–760.

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., … Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, ***26***(19), 5160–5172.

Exposito-Alonso, M., Vasseur, F., Ding, W., Wang, G., Burbano, H. A., & Weigel, D. (2018). Genomic basis and evolutionary potential for extreme drought adaptation in *Arabidopsis thaliana*. *Nature Ecology & Evolution*, ***2***(February), 352–358.

Flint-Garcia, S. A., Thornsberry, J. M., & Edward IV, S. B. (2003). Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, ***54***, 357–374.

Gabrielli, M., Nabholz, B., Leroy, T., Milá, B., & Thébaud, C. (2020). Within-island diversification in a passerine bird. *Proceedings of the Royal Society B: Biological Sciences*, ***287***(1923).

Gonzalez-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***(17), 4234–4246.

Gonzalez-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***(5), 1091–1102.

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

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., … Yang, J. (2018). Global genetic differenciation of complex traits shaped by natural selection in humans. *Nature Communications*, ***9***(1865), 1–9.

Hill, W. G., & Weir, B. S. (1988). Variances and covariances of squared linkage disequilibria in finite populations. *Theoretical Population Biology*, ***33***(1), 54–78.

Hinrichs, A. L., Larkin, Ã. E. K., & Suarez, B. K. (2009). Population stratification and patterns of linkage disequilibrium. *Genetic Epidemiology*, ***33***(Supplement 1), 88–92.

Hsieh, C. L., Lin, C. L., Liu, H., Chang, Y. J., Shih, C. J., Zhong, C. Z., … Tan, B. C. M. (2011). WDHD1 modulates the post-transcriptional step of the centromeric silencing pathway. *Nucleic Acids Research*, ***39***(10), 4048–4062.

Huber, B., Whibley, A., Poul, Y. L., Navarro, N., Martin, A., Baxter, S., … Joron, M. (2015). Conservatism and novelty in the genetic architecture of adaptation in *Heliconius* butterflies. *Heredity*, ***114***(5), 515–524.

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***(21), 4599–4612.

Illera, 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*, ***5****0*, 14–22.

Illera, J. C., Spurgin, L. G., Rodriguez-Exposito, E., Nogales, M., & Rando, J. C. (2016). What are we learning about speciation and extinction from the Canary Islands? *Ardeola*, ***63***(1), 15–33.

Izuno, A., Kitayama, K., Onoda, Y., Tsujii, Y., Hatakeyama, M., Nagano, A. J., & Honjo, M. N. (2017). The population genomic signature of environmental association and gene flow in an ecologically divergent tree species *Metrosideros polymorpha (Myrtaceae)*. *Molecular Ecology*, ***26***, 1515–1532.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, ***484***(7392), 55–61.

Lamichhaney, S., Berglund, J., Almén, M. S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., … Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, ***518***(7539), 371–375.

Lawson, L. P., & Petren, K. (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. *Molecular Ecology*, (October 2015), 4978–4989.

Lenormand, T. (2002). Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, ***17***(4), 183–189.

Linck, E., & Battey, C. J. (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources*, ***19***(December 2018), 639–647.

Liu, S., He, S., Chen, L., Li, W., Di, J., & Liu, M. (2017). Estimates of linkage disequilibrium and effective population sizes in Chinese Merino (Xinjiang type) sheep by genome-wide SNPs. *Genes & Genomics*, ***39***(7), 733–745.

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. *Nature Review Insight*, ***457***(February).

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***(12), 981–994.

Lundregan, S. L., Hagen, I. J., Gohli, J., Niskanen, A. K., Kemppainen, P., Ringsby, T. H., … Jensen, H. (2018). Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology*, ***27***(17), 3498–3514.

Manel, S., Gugerli, F., Thuiller, W., Alvarez, N., Legendre, P., Holderegger, R., … Taberlet, P. (2012). Broad-scale adaptive genetic variation in alpine plants is driven by temperature and precipitation. *Molecular Ecology*, ***21***(15), 3729–3738.

Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., … Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, ***23***, 1817–1828.

Meier, K., Hansen, M. M., Bekkevold, D., Skaala, & Mensberg, K. L. D. (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta L.*): Footprints of selection at microsatellite DNA loci. *Heredity*, ***106***(3), 488–499.

Milá, B., Warren, B. H., Heeb, P., & Thébaud, C. (2010). The geographic scale of diversification on islands : Genetic and morphological divergence at a very small spatial scale in the Mascarene grey white-eye (*Aves : Zosterops borbonicus*). *BMC Evolutionary Biology*, ***10***(158), 1–13.

Moody, K. N., Hunter, S. N., Blob, R. W., Schoenfuss, H. L., Biology, E., & Orleans, N. (2015). Local adaptation despite high gene flow in the waterfall-climbing Hawaiian goby, *Sicyopterus stimpsoni*. *Molecular Ecology*, ***24***, 545–563.

Pardo-Diaz, C., Salazar, C., & Jiggins, C. D. (2015). Towards the identification of the loci of adaptive evolution. *Methods in Ecology and Evolution*, ***6***(4), 445–464.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., … Reich, D. (2012). Ancient admixture in human history. *Genetics*, ***192***(3), 1065–1093.

Pickrell, J. K., & Pritchard, J. K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genetics*, ***8***(11).

Pilot, M., Greco, C., Vonholdt, B. M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., … Wayne, R. K. (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity*, ***112***(4), 428–442.

Ray, J. M., & King, R. B. (2006). The temporal and spatial scale of microevolution : Fine-scale colour pattern variation in the Lake Erie watersnake, Nerodia sipedon insularum. *Evolutionary Ecology Research*, ***8***, 915–928.

Reich, D. E., Cargili, M., Boik, S., Ireland, J., Sabeti, P. C., Richter, D. J., … Lander, E. S. (2001). Linkage disequilibrium in the human genome. *Nature*, ***411***(6834), 199–204.

Reich, D., Thangaraj, K., Patterson, N., Price, A. L., & Singh, L. (2009). Reconstructing Indian population history. *Nature*, ***461***(7263), 489–494.

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.

Rominger, A. J., Goodman, K. R., Lim, J. Y., Armstrong, E. E., Becking, L. E., Bennett, G. M., … Gillespie, R. G. (2016). Community assembly on isolated islands : Macroecology meets evolution. *Global Ecology and Biogeography*, ***25***, 769–780.

Safran, R. J., Scordato, E. S. C., Wilkins, M. R., Hubbard, J. K., Jenkins, B. R., Albrecht, T., … Kane, N. C. (2016). Genome-wide differentiation in closely related populations: The roles of selection and geographic isolation. *Molecular Ecology*, ***25***(16), 3865–3883.

Salces-Castellano, A., Patiño, J., Alvarez, N., Andújar, C., Arribas, P., Braojos-Ruiz, J. J., … Emerson, B. C. (2020). Climate drives community-wide divergence within species over a limited spatial scale: Evidence from an oceanic island. *Ecology Letters*, ***23***, 305–315.

Sendell-Price, A. T., Ruegg, K. C., Anderson, E. C., Quilodrán, C. S., Van Doren, B. M., Underwood, V. L., … Clegg, S. M. (2020). The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes (*Zosterops lateralis*). *bioRxiv*, ***21***(1), 1–9.

Shannon, L. M., Boyko, R. H., Castelhano, M., Corey, E., Hayward, J. J., McLean, C., … Boyko, A. R. (2015). Genetic structure in village dogs reveals a Central Asian domestication origin. *Proceedings of the National Academy of Sciences of the United States of America*, ***112***(44), 13639–13644.

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*(5), 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***(3), 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***(24), 5213–5225.

Tegeder, I., Costigan, M., Griffin, R. S., Abele, A., Belfer, I., Schmidt, H., … Woolf, C. J. (2006). GTP cyclohydrolase and tetrahydrobiopterin regulate pain sensitivity and persistence. *Nature Medicine*, ***12***(11), 1269–1277.

Tigano, A., & Friesen, V. L. (2016). Genomics of local adaptation with gene flow. *Molecular Ecology*, ***25***(10), 2144–2164.

Tokumasu, Y., Iida, A., Wang, Z., Ansai, S., Kinoshita, M., & Sehara-Fujisawa, A. (2016). ADAM12-deficient zebrafish exhibit retardation in body growth at the juvenile stage without developmental defects. *Development Growth and Differentiation*, ***58***(4), 409–421.

Tusso, S., Nieuwenhuis, B. P. S., Weissensteiner, B., Immler, S., & Wolf, J. B. W. (2020). Experimental evolution of adaptive divergence under varying degrees of gene flow. *bioRxiv*, 2020.11.02.364695.

Vincent, B., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. (2013). Landscape genomics in Atlantic Salmon (Salmo salar): Searching for gene-envrionment interactions driving local adaptation. *Evolution*, ***67***(12), 3469–3487.

Voelker, G. (1999). Dispersal, vicariance, and clocks: Historical biogeography and speciation in a cosmopolitan passerine genus (*Anthus: Motacillidae*). *Evolution*, ***53***(5), 1536–1552.

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

Warren, W. C., Clayton, D. F., Ellegren, H., Arnold, A. P., Hillier, L. W., Künstner, A., … Wilson, R. K. (2010). The genome of a songbird. *Nature*, ***464***(7289), 757–762.

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, ***38***(6), 1358–1370.

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics*, ***88***(1), 76–82.

Yoshida, G. M., & Yáñez, J. M. (2021). Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics*, ***22***(1), 1–13.

8 Supplementary material

Table S1. Summary of RAD-seq analyses. **A)** Variant filtering and **B)** number of individuals per population and SNPs for each analyses. * = loci successfully mapped to chromosomes with MAF >0. BP = Berthelot's pipit populations, not including tawny pipits.

A

	Population Genetics						Selection		
Trimming/Filtering	Relatedness	PCA	*TreeMix* Berthelot's	*TreeMix* All Pipits	BP Pop F_{ST}	LD	EigenGWAS	Loci F_{ST}	MAF
Low MAF (<0.03)	X	X	X		X	X	X	X	
Low MAF (<0.01)				X					
Loci in strong LD (r^2 >0.4, 50 kb window, 10 marker step)			X		X				
Markers on Z sex chromosome	X	X	X	X	X				
Individuals with high within-population Relatedness (>0.2, Plink)		X	X	X	X		X	X	

B

	Population Genetics						Selection		
n SNPs and individuals	Relatedness	PCA	*TreeMix* Berthelot's	*TreeMix* All Pipits	BP Pop F_{ST}	LD	EigenGWAS	Loci F_{ST}	MAF
n SNPs Canary Islands	4367	4367	2850	1650	2633 - 2850	2603 (GRA) - 2963 (TF)	4470 (3531*)	4470 (3531*)	-
n individuals Canary Islands	182	176	176	176	176	182	176	176	182
n SNPs Madeira	2877	2877	2850	1650	1822 - 2850	1600 (DG) - 1752 (M)	2938 (2295*)	2938 (2295*)	-
n individuals Madeira	60	50	50	50	50	60	49	49	60
n SNPs Selvagens	2021	-	2850	1650	-	1273	-	-	-
n individuals Selvagens	20	-	20	20	-	20	-	-	20

Table S2. Evolutionary admixture and shared drift between Berthelot's pipit populations. Three-population, f_3 statistics for all population comparisons, defined as the product of allele frequency differences between population A to B and C, respectively. All populations with a Z-score < -0.5 are reported using a block jackknifing approach in 50 SNP windows. Results are ordered by Z-score within population groups (i.e., all Madeira island comparisons first).

Population comparison (A; B,C)	f_3 statistic	f_3 Standard error	Z-score
M; LP, PS	-4.2×10^{-4}	+/- 0.00033	- 1.271
M; SG, PS	-5.3×10^{-4}	+/- 0.00056	- 0.956
M; GOM, PS	-2.7×10^{-4}	+/- 0.00037	- 0.727
M; PS, TEID	-2.3×10^{-4}	+/- 0.00034	- 0.668
M; FV, PS	-1.9×10^{-4}	+/- 0.00033	- 0.584
FV; LP, PS	-3.2×10^{-4}	+/- 0.00032	- 0.994
FV; GRA, DG	-4.5×10^{-4}	+/- 0.00053	- 0.851
FV; M, GRA	-3.9×10^{-4}	+/- 0.00048	- 0.827
FV; GRA, TEID	-1.4×10^{-4}	+/- 0.00017	- 0.796
GC; LP, PS	-3.0×10^{-4}	+/- 0.00034	- 0.873
GC; GRA, DG	-2.8×10^{-4}	+/- 0.00048	- 0.578
TF; LP, PS	-3.0×10^{-4}	+/- 0.00042	- 0.733
TEID; LP, PS	-2.8×10^{-4}	+/- 0.00037	- 0.764
TEID; GRA, DG	-3.7×10^{-4}	+/- 0.00057	- 0.645
LZ; GRA, PS	-3.2×10^{-4}	+/- 0.00047	-0.677
LZ; M, GRA	-2.8×10^{-4}	+/- 0.00047	-0.585

Table S3. Minor allele frequency by population for EigenGWAS outlier SNPs in Berthelot's Pipit.
Archipelago Acronyms: Canary Islands (CAN), Madeiran archipelago (MAD) and Selvagens (SEL). Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID), Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagens: Selvagem Grande (SG).

		Minor Allele Frequency												
SNP	Archipelago under selection	CAN									MAD			SEL
		EH	GOM	LP	TEID	TF	GC	FV	LZ	GRA	M	PS	DG	SG
219s24	CAN	0.45	0.16	0.00	0.03	0.02	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00
1585s94	MAD	0.00	0.10	0.08	0.05	0.00	0.13	0.03	0.00	0.08	0.00	0.00	0.48	0.03
1585s112	MAD	0.00	0.10	0.08	0.05	0.00	0.10	0.03	0.00	0.08	0.00	0.00	0.48	0.03
790s54	MAD	0.08	0.13	0.05	0.06	0.02	0.10	0.00	0.03	0.03	0.00	0.00	0.48	0.00

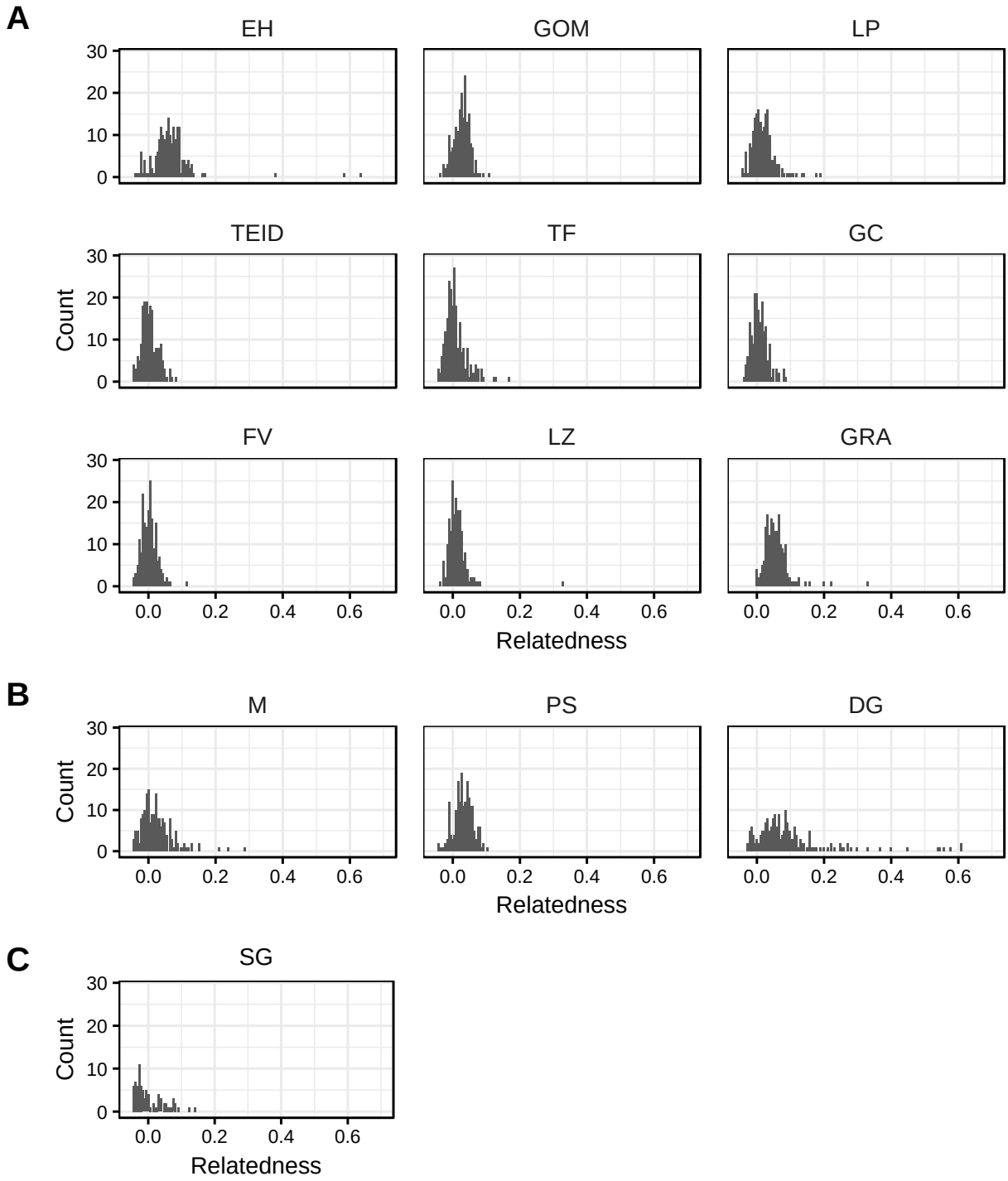
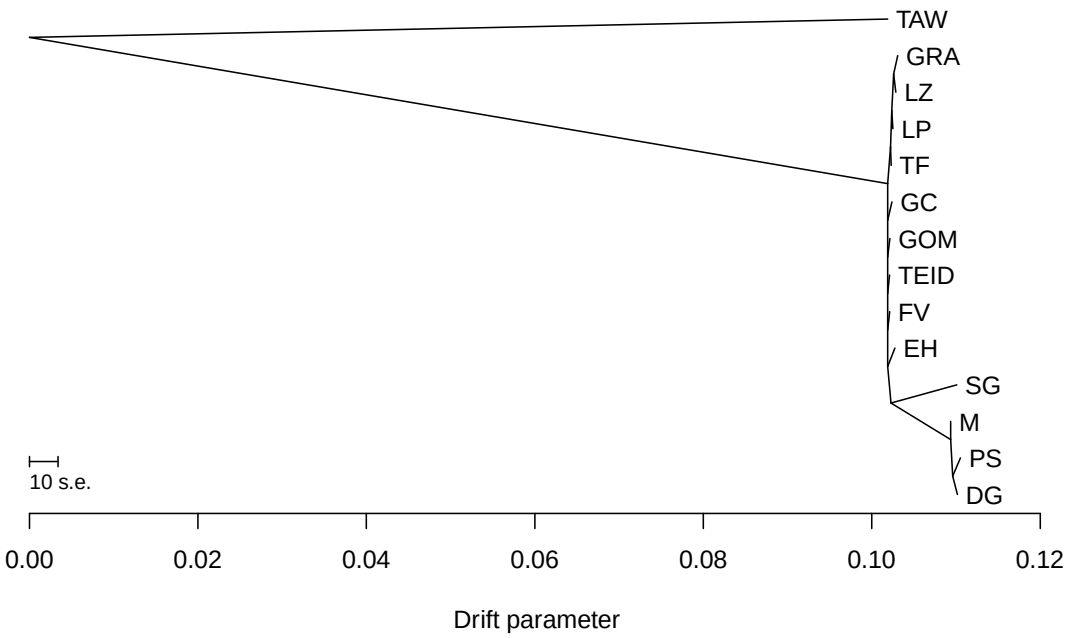


Figure S1. Relatedness of each pair of Berthelot's pipit individuals within **A)** the nine Canary Island populations **B)** the three Madeiran islands and **C)** the Selvagens. Canary Islands: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID), Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), La Graciosa (GRA). Madeiran archipelago: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagens population: Selvagem Grande (SG).

A



B

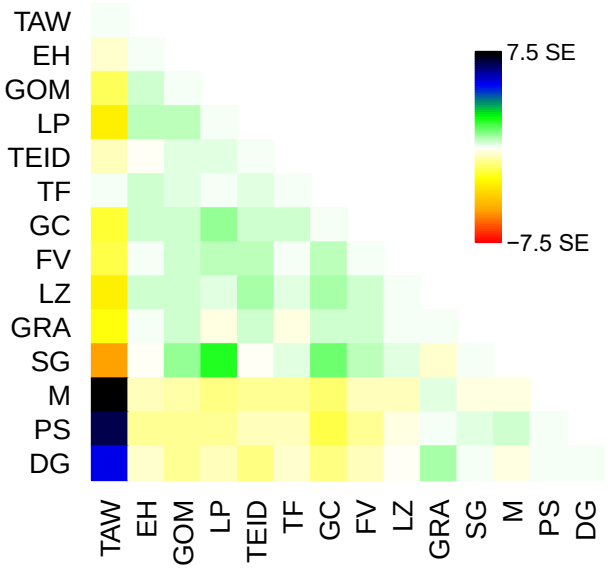


Figure S2. Evolutionary divergence and speciation from the tawny pipit across island populations of the Berthelot's pipit. **A)** Evolutionary maximum likelihood tree, without gene flow, between island populations of the Berthelot's pipit, with tawny pipit root. **B)** Residuals fitted for pairs of populations modelled with no migration. Residuals above zero represent populations that are more closely related to each other in the data than in the best-fit tree and are candidates for admixture. Negative residuals indicate that a pair of populations are less closely related, based on the data, than represented in the best fit tree.

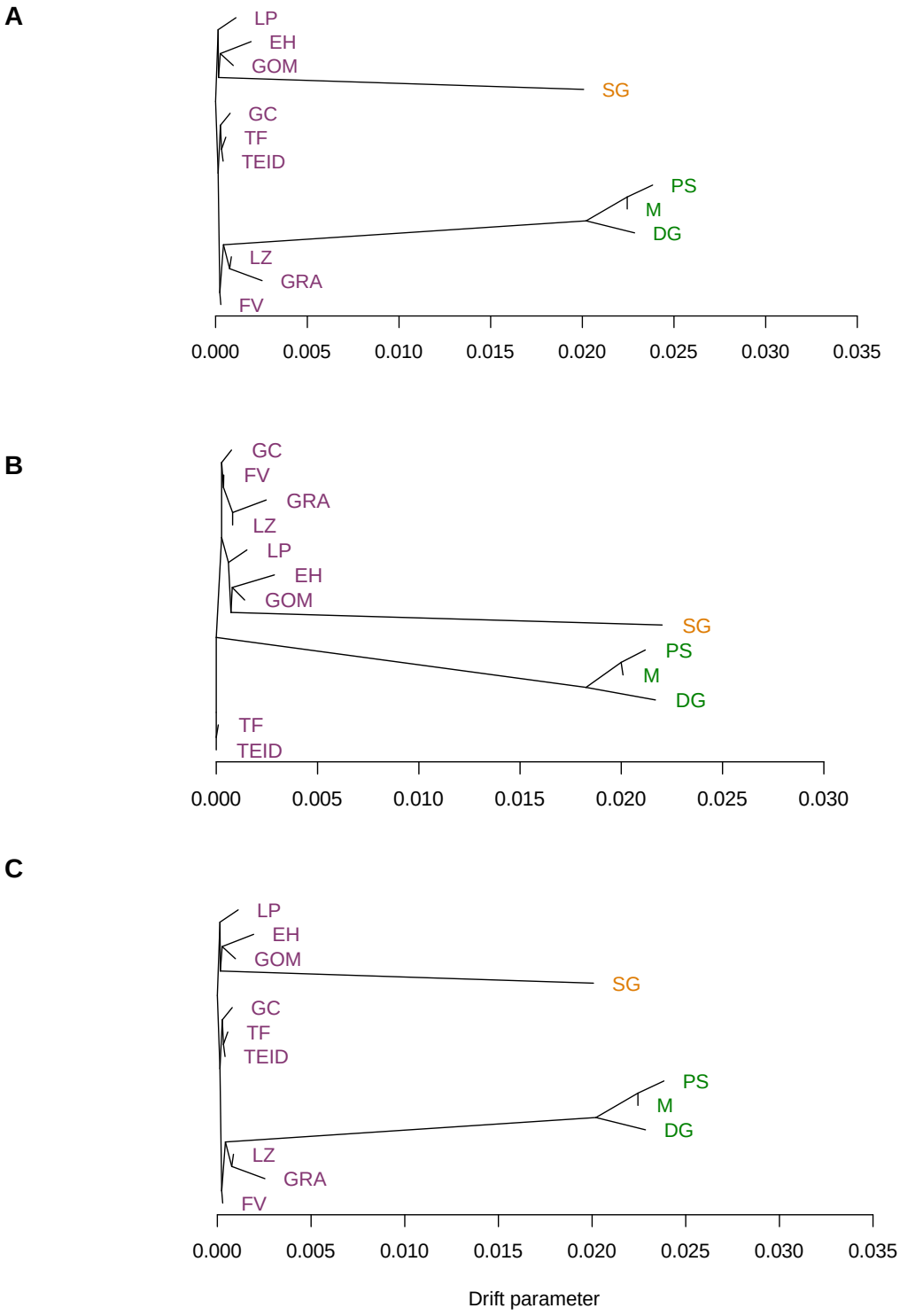


Figure S3. Evolutionary relationships between island populations of Berthelot's pipit using different SNP window sizes. Maximum likelihood bifurcating tree, without subsequent gene flow, across the pipit colonisation range as inferred by *TreeMix*. SNP averaging window sizes **A)** 20. **B)** 50. **C)** 80. Populations are coloured by archipelago; Canary Islands (purple), Madeiran archipelago populations (green) and Selvagens (orange).

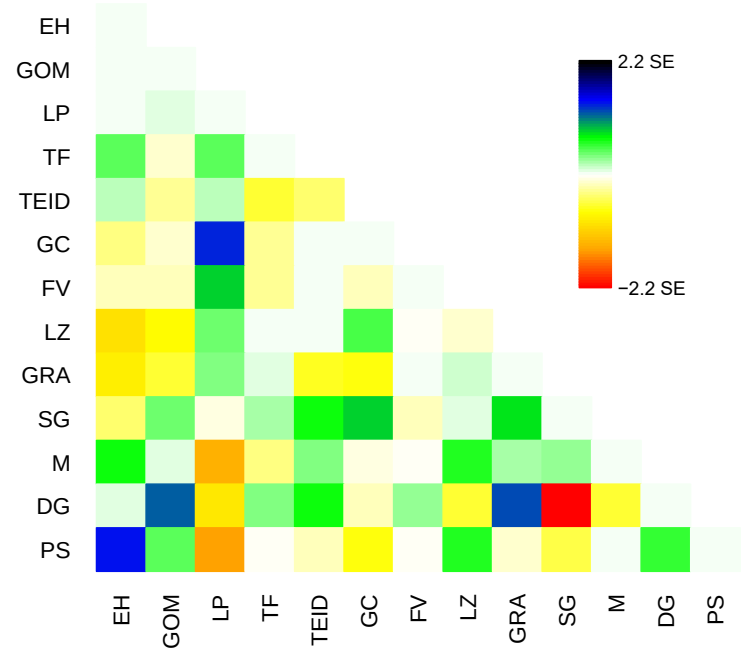


Figure S4. Evolutionary *TreeMix* model fit to Berthelot's pipit populations, with no migration edges, using windows of 50 SNPs. Population pairs with positive residuals (navy) were taken to be genetically more closely related than represented in the best fit tree and are candidates for admixture, while negative residuals (yellow-red) suggest population pairs too closely represented by the tree.

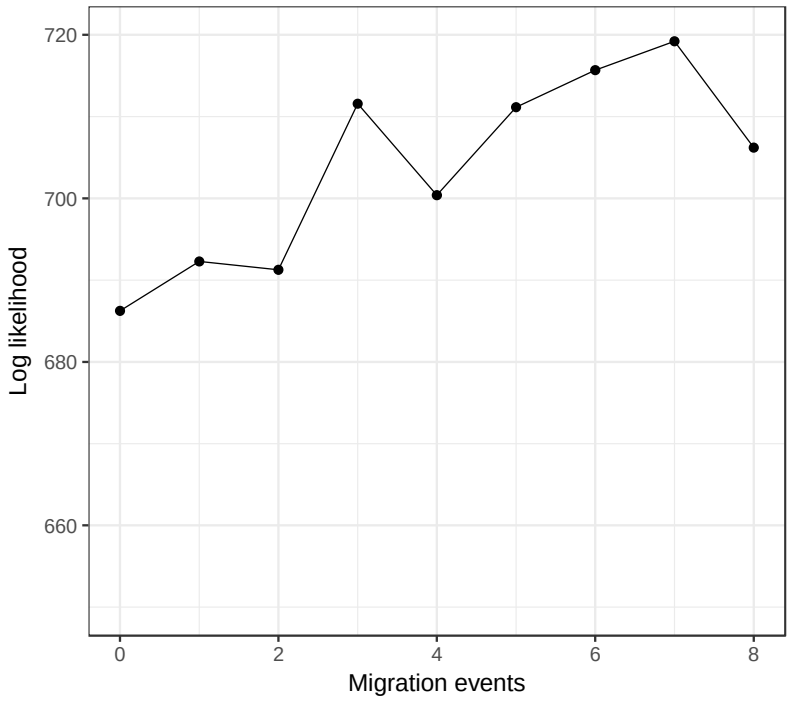
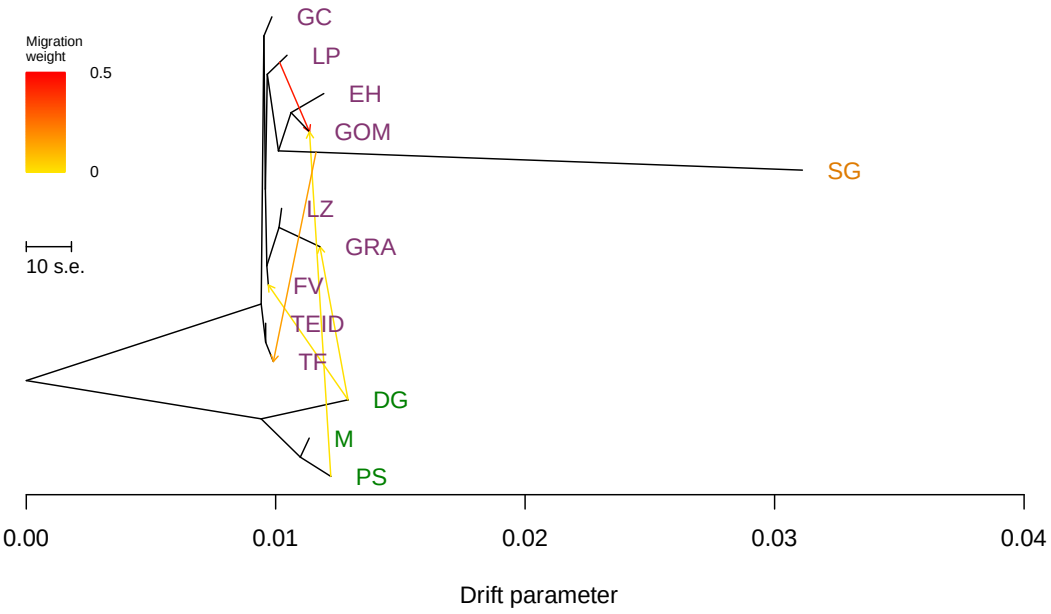


Figure S5. Log likelihood score for the *TreeMix* "Berthelot's" tree with zero to eight migration events modelled using 50 SNP windows.

A



B

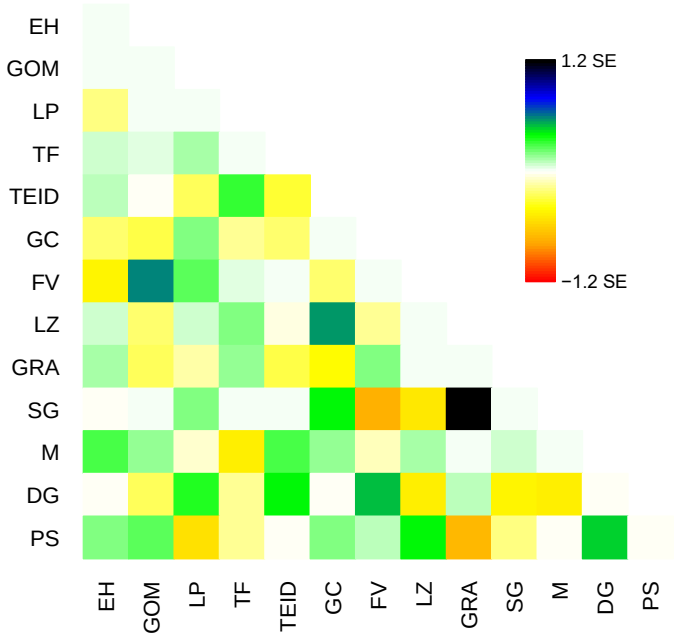


Figure S6. Evolutionary relationships between island populations of the Berthelot's pipit, with five inferred migration edges. **A)** Maximum likelihood tree across the pipit colonisation range as inferred by *TreeMix*. **B)** Pairwise residual plot.

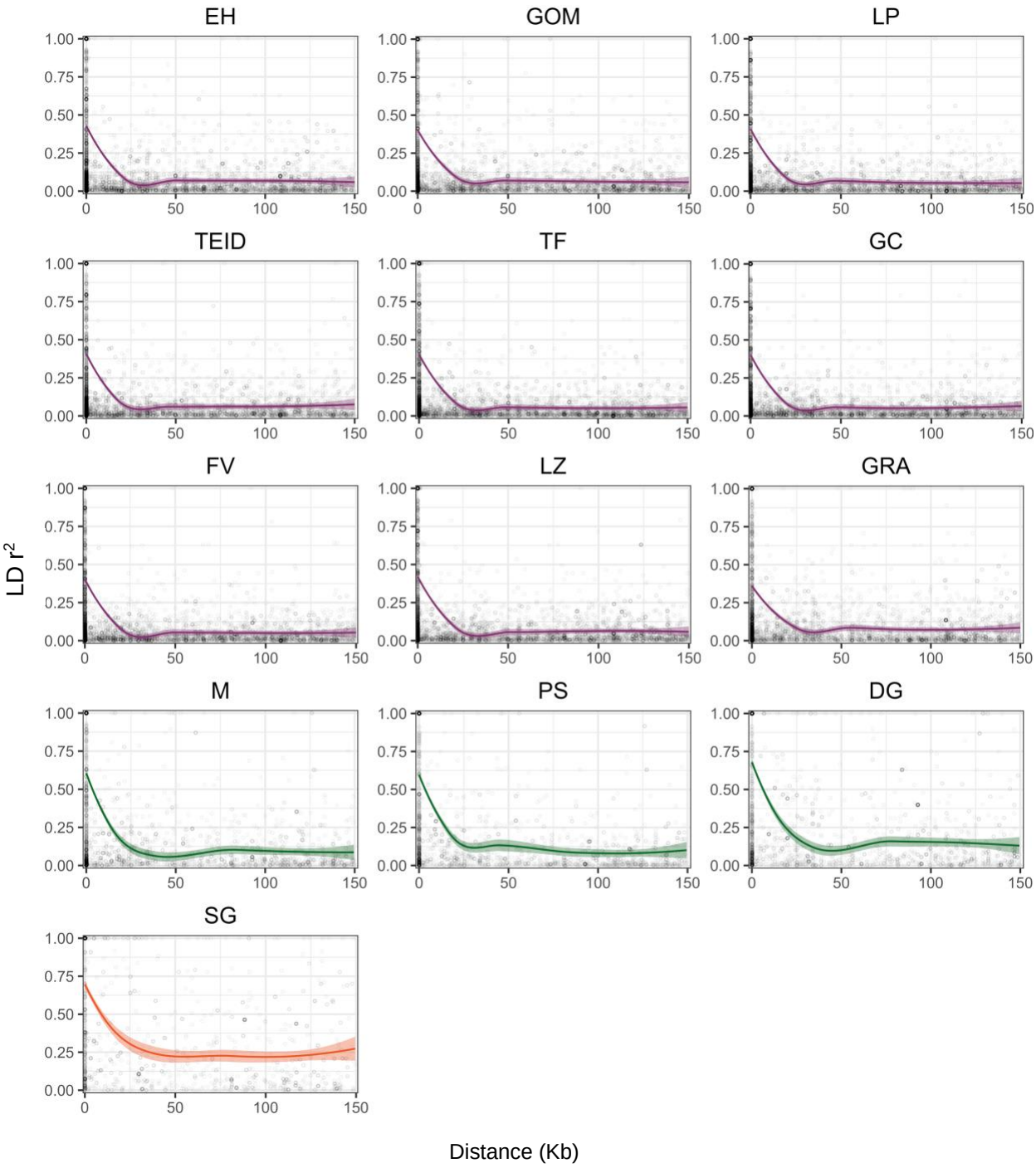
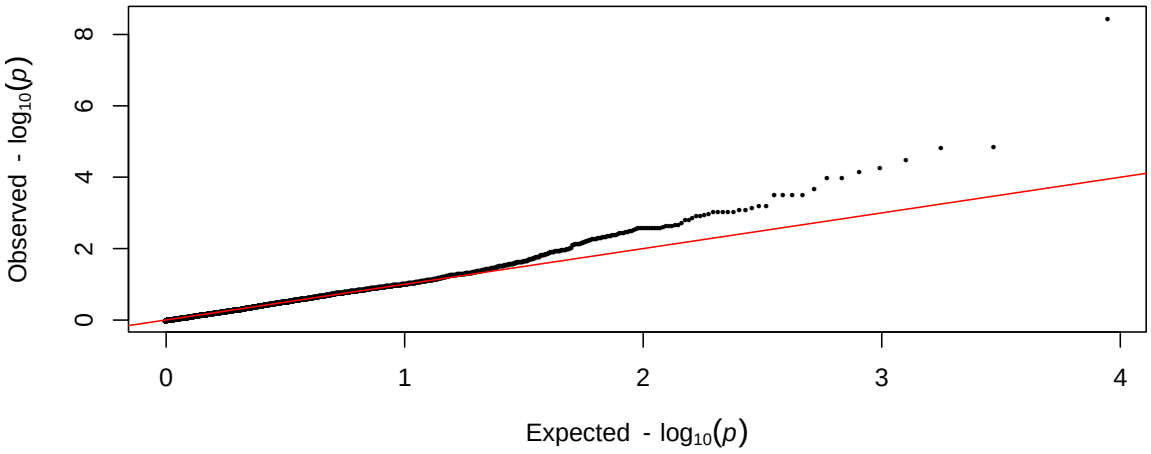


Figure S7. Population level linkage disequilibrium decay across 13 island populations of Berthelot's pipit. Pairwise linkage disequilibrium and base-pair distance for SNPs across the Canary Island populations (purple), Madeiran archipelago (green) and Selvagens (orange). The fit lines show a local regression model, with a shaded band indicating 95% confidence intervals.

A



B

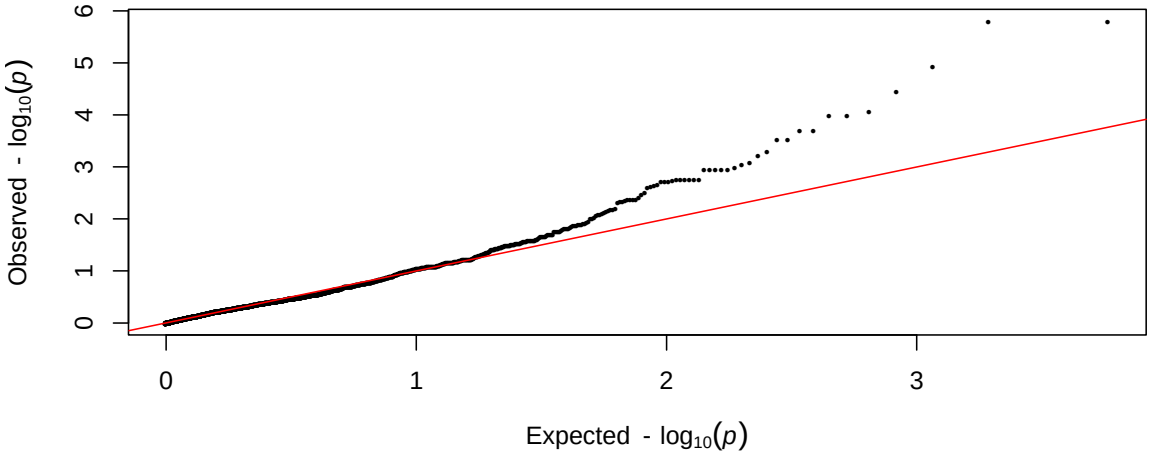
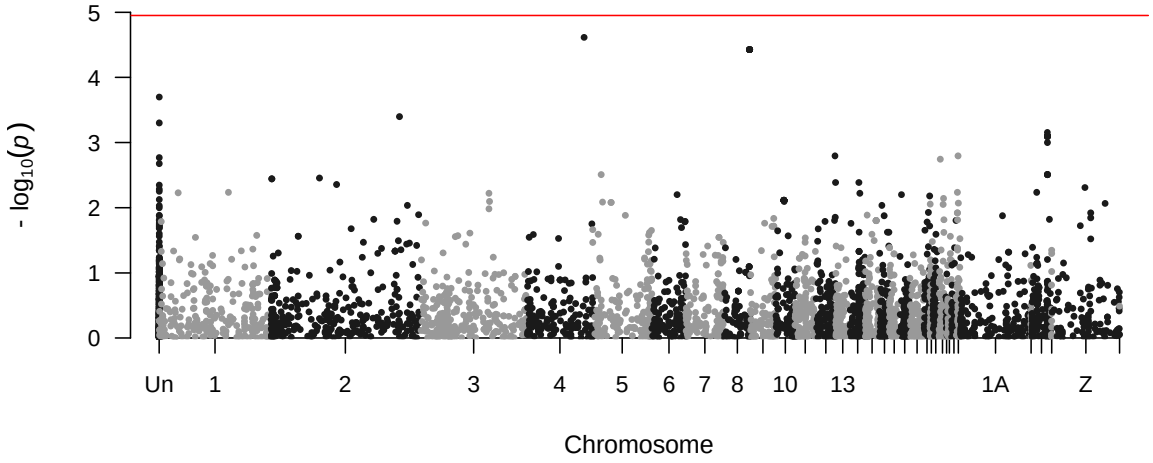
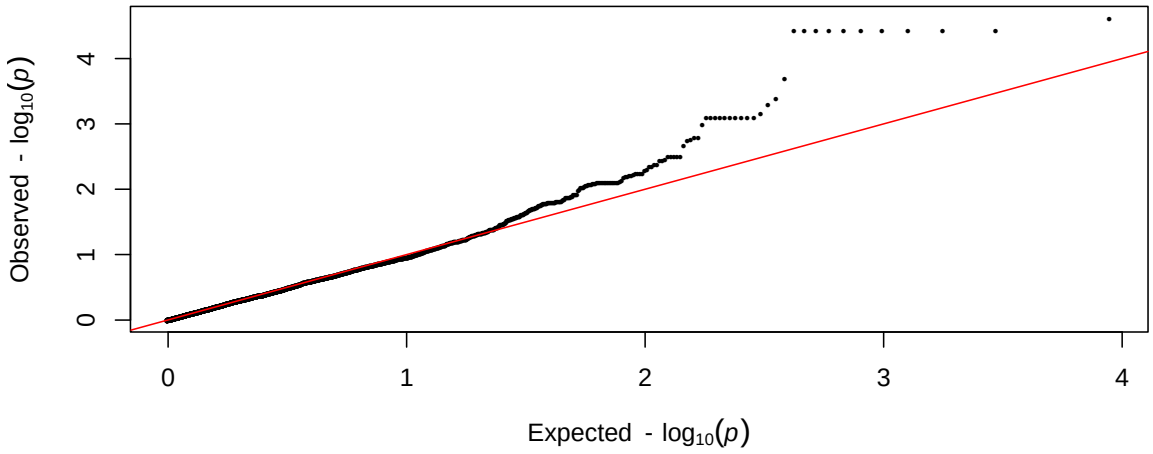


Figure S8. QQ-plots for PC1 showing deviation of SNP *P* values (with genomic inflation correction, λ_{GC}) from the null expectation (red line) for EigenGWAS analysis among archipelago populations of the Berthelot's pipit. **A)** 4470 RAD loci across the Canary Islands and **B)** 2938 RAD loci across the Madeiran archipelago.

A


5
4
3
2
1
0
- log10(p)
Un
1
2
3
4
5
6
7
8
10
13
1A
Z
Chromosome

B


Observed - log10(p)
0
1
2
3
4
Expected - log10(p)

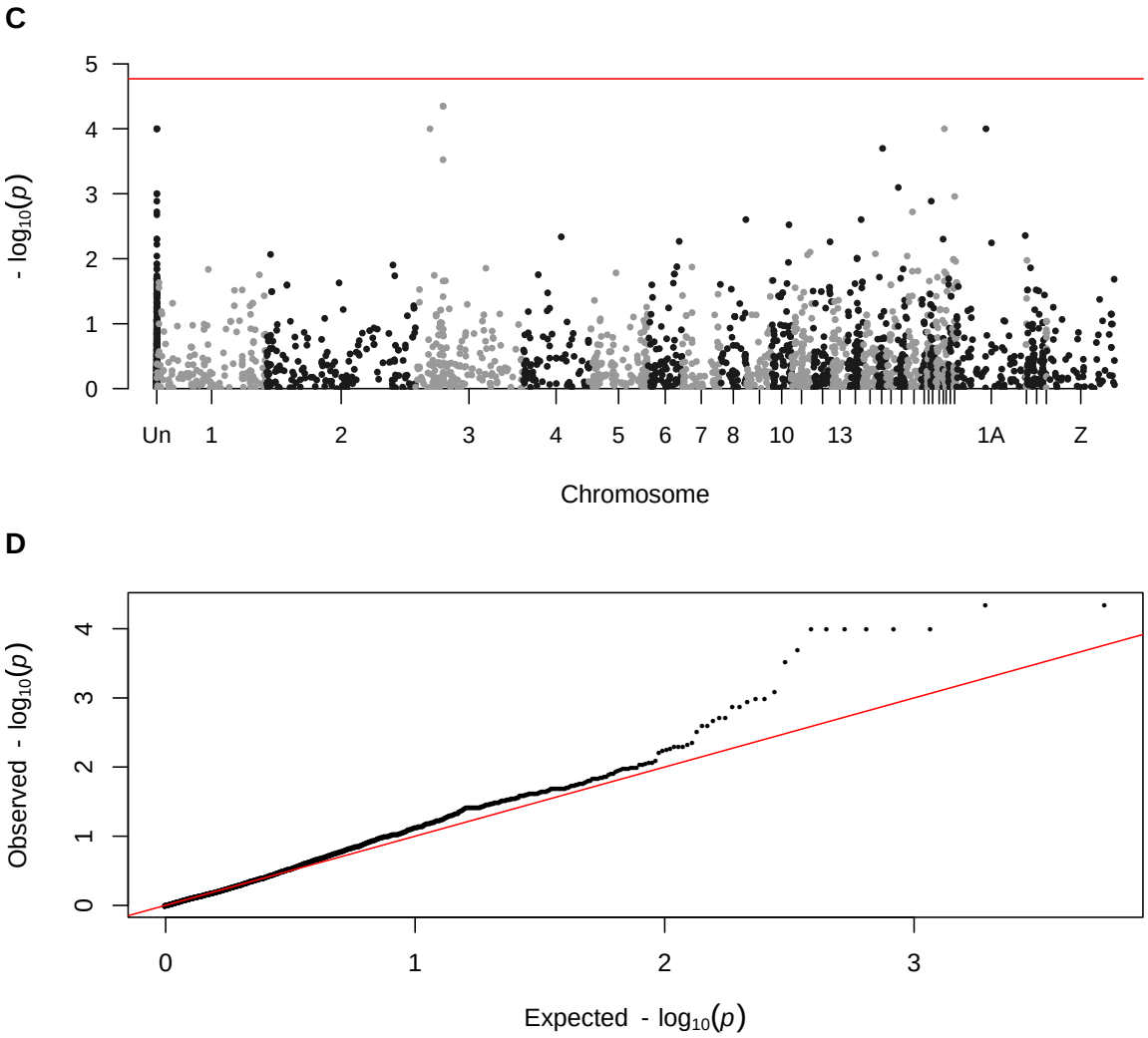


Figure S9. Genome scan EigenGWAS analysis along PC2 axis of population structure among archipelago populations of the Berthelot's pipit. See Figure 3 for PC2 axis of within-archipelago genetic structure. **A)** Manhattan plot and **B)** QQ-plot Canary Islands. **C)** Manhattan plot and **D)** QQ-plot Madeiran archipelago. QQ-plots show deviation of genomic inflation corrected *P* values from expectation for EigenGWAS analysis. Red lines represent Bonferroni significance thresholds and null expectations in the Manhattan and QQ-plots, respectively. Black-grey colouring in the Manhattan plots indicates chromosomal limits.

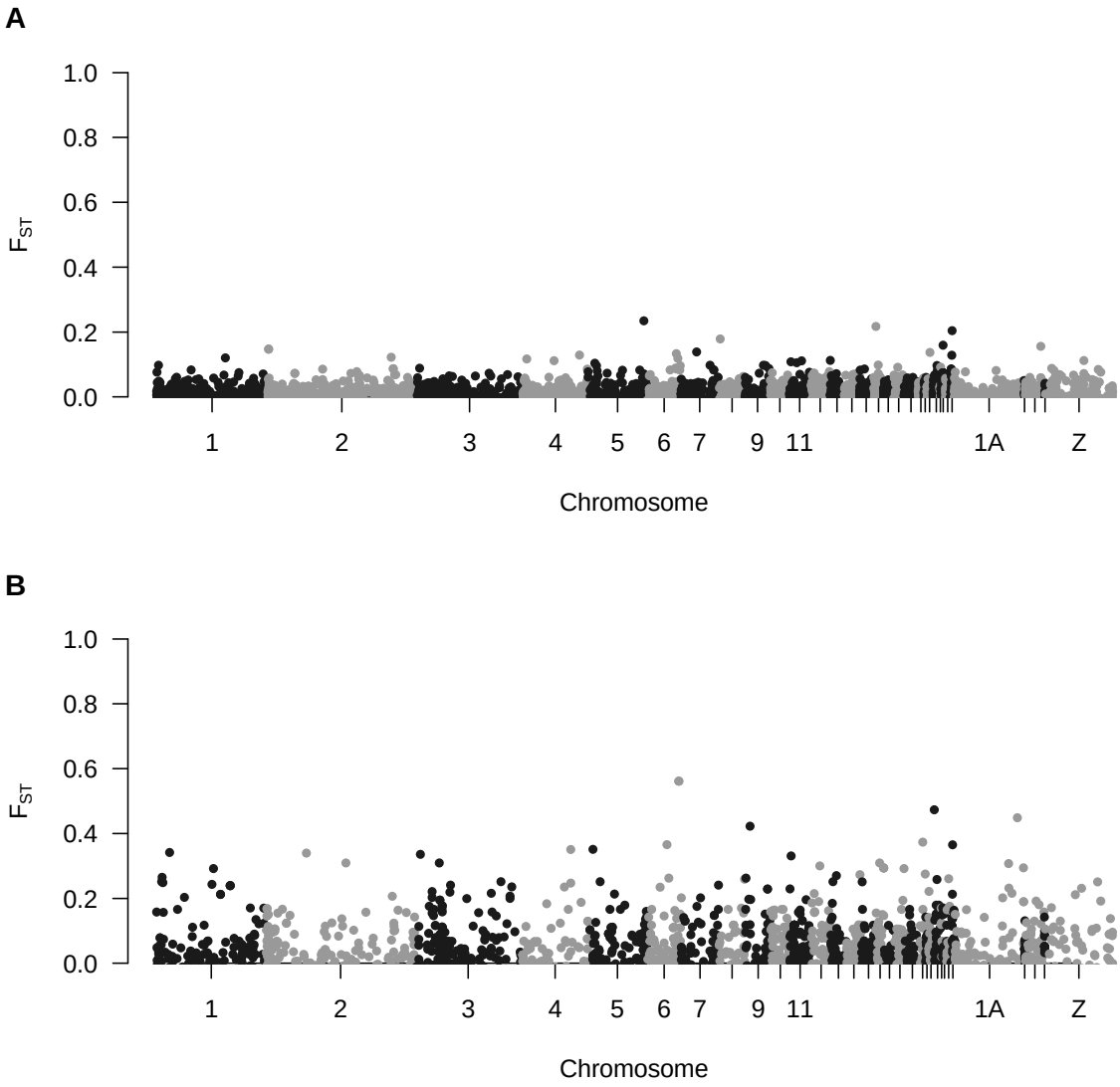


Figure S10. Genome-wide locus F_{ST} between Berthelot's pipit populations across **A)** the Canary Islands (n SNPs = 3531) and **B)** the Madeiran archipelago (n SNPs = 2295). Black-grey colouring indicates chromosomal limits.

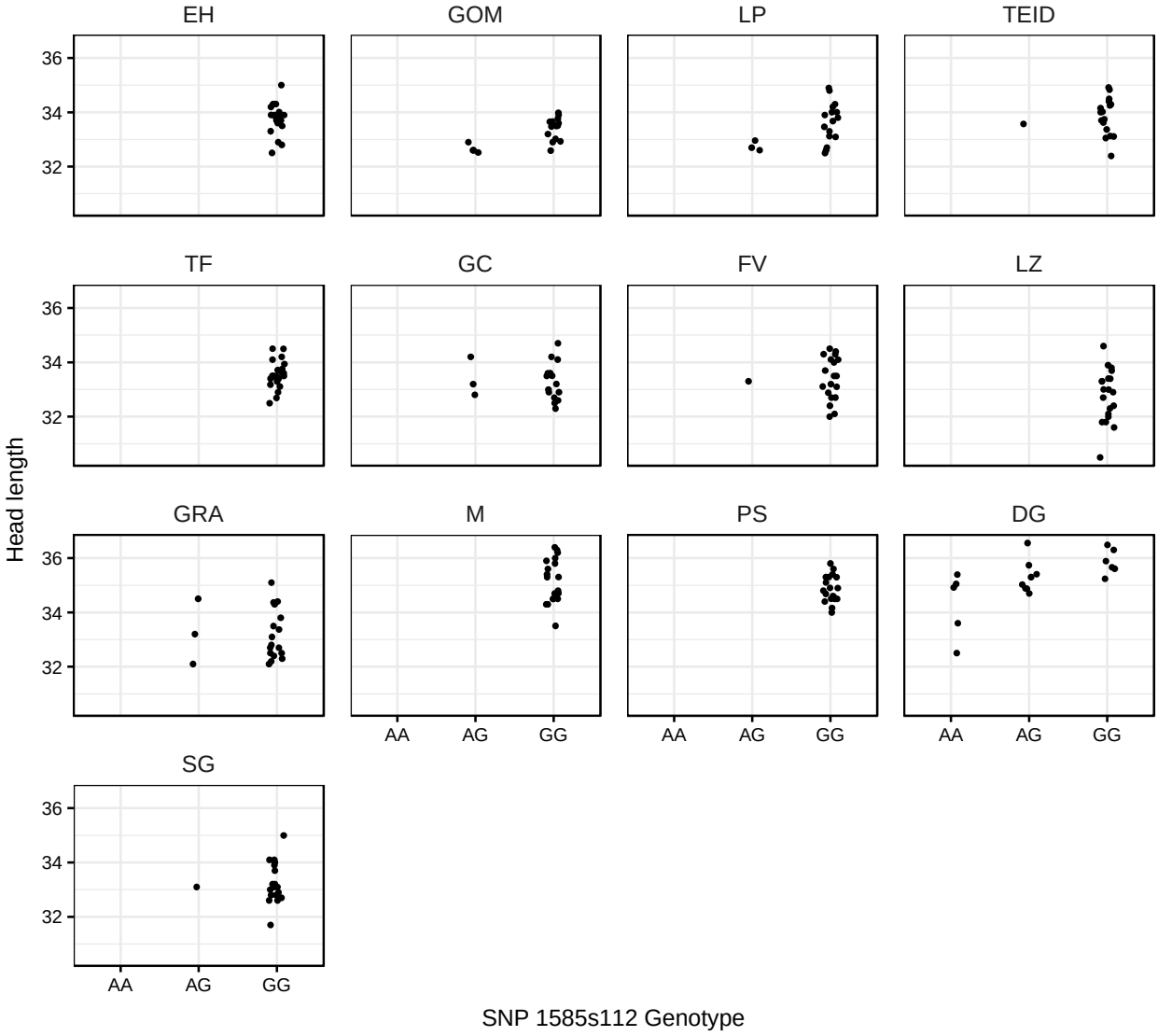


Figure S11. Population level genotype at the *ADAM12* SNP 1585s112, putatively under selection across the Madeiran archipelago, and it's association with head length across 13 island populations of the Berthelot's pipit. 20 samples for each population, excluding TF that has 22. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), El Teide (TEID), Lowland Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagens: Selvagem Grande (SG).

3

Genomic inference of colonisation timescales, associated bottlenecks and contemporary inbreeding across island bird archipelagos

Claudia A. Martin, Brent C. Emerson, Juan Carlos Illera, Alexander Suh, Krystyna Nadachowska-Brzyska, Lewis G. Spurgin and David S. Richardson.

Enticing pipits with playback calls in Lanzarote, Canary Islands

1 Abstract

Genomes retain records of demographic history and evolutionary forces that shape species and their populations. Across island systems, complex population demography may have shaped contemporary patterns of genetic diversity, including colonisation events and associated bottlenecks, gene flow and genetic drift, which may act particularly strongly in small and isolated populations. Populations that have recently colonised oceanic island archipelagos provide excellent opportunities to understand how evolutionary forces shape diversity across populations. Here, we use whole genome resequencing from six populations across three island archipelagos of Berthelot's pipit (*Anthus berthelotii*) - a passerine which has undergone island speciation relatively recently - to characterise and date past divergence through to contemporary demography. Pairwise Markovian coalescent (PSMC) analyses estimated divergence from the mainland approximately two million years ago (Mya), similar to estimates from mitochondrial cytochrome *b* sequences. Berthelot's pipit populations across archipelagos had estimated shared ancestry until approximately 50 thousand years ago (kya), when results suggest the Madeiran archipelago populations were founded, while the Selvagens colonised much more recently. We identify extensive runs of homozygosity (ROHs) extending >1 Mb across genomes in the most recently colonised populations which have experienced sequential island founder events. The size and distribution of ROH are in agreement with these estimated population bottlenecks. Subsequent mutation and recombination may have eroded long ROH in the Madeiran archipelago, and/or this may reflect moderate background levels of contemporary inbreeding. Extensive long and short ROH in the Selvagens reflects strong recent inbreeding and bottleneck effects, with as much as 38% of autosomes comprised of ROH >250 kb in length. Taken together, these findings highlight the importance of demographic history, as well as selection, in shaping contemporary patterns of genetic diversity, and speciation events.

Keywords: Whole genome resequencing, island endemic, speciation, founder events, runs of homozygosity (ROH), bottleneck, effective population size (N_e), PSMC.

2 Introduction

Patterns of genetic diversity within individual genomes and across populations are a consequence of evolutionary history, and specifically by a history of neutral and selective processes (see review by Ellegren and Galtier 2016). While selection acts locally on specific loci and linked genomic regions, demographic processes shape diversity more evenly across the genome (Maynard-Smith and Haigh 1974; Ma *et al.* 2017). Populations founded by few individuals or having experienced a drastic population decline may undergo genetic bottlenecks, which may lead to strong genetic drift and/or inbreeding (reviewed by Weaver *et al.* 2021). Geographic separation of populations, where there is limited gene flow, may further promote differentiation between populations especially where there is strong selection (Martin *et al.* 2013; Pilot *et al.* 2014). Genetic diversity within, and divergence among, populations can provide insight into past and present demography, which is, in turn, important for our understanding of what evolutionary processes shape diversity across natural environments (Luikart *et al.* 2003).

In island systems, populations may be the product of multiple founding steps, for example during a stepwise range expansion (Halkka *et al.* 1974; Sendell-Price *et al.* 2021). Genome-wide diversity of island populations is predominantly shaped by colonisation events and associated bottlenecks, gene flow and genetic drift (Carson 1971; Nei *et al.* 1975). Island populations often experience reduced genetic diversity relative to their mainland populations or ancestors (e.g., as oberved in island foxes, Robinson *et al.* 2018; and island songbirds, Leroy *et al.* 2021; Frankham 1997). The potential for loss of genetic diversity is expected to be exaggerated by sequential bottlenecks and drift as a result of long-term isolation and small population size (Gautschi *et al.* 2002). Limited gene flow between islands may result in population structure and facilitate divergence or local adaptation (Clegg *et al.* 2002). Over time genetic differences between populations may result in reproductive isolation and speciation (Nosil and Feder 2012; Warren *et al.* 2012; Comeault *et al.* 2015). Variation in demographic history and selective pressures across island populations can drive divergence and speciation, thus creating island systems that are some of the most biologically diverse habitats globally (Paulay 1994). Understanding the evolutionary processes acting across island populations, with a range of colonisation histories, is important for understanding how genetic diversity is shaped in small isolated populations.

Studying contemporary patterns of inbreeding may provide insight into recent demography and population size. Inbreeding not only reflects population level processes, but can have

negative consequences for individual fitness and survival termed ‘inbreeding depression’ (Charlesworth and Willis 2009), which may have implications for population persistence (Oostermeijer *et al.* 1995; Frankham 2005). Traditionally inbreeding has been investigated using pedigrees to estimate genetic relatedness (Ballou 1983). However, inbreeding can now more accurately be estimated using genomic approaches (Kardos *et al.* 2015). One powerful way to measure inbreeding is to identify chromosome segments which are identical by descent (IBD) (McQuillan *et al.* 2008; Stoffel *et al.* 2020). These runs of homozygosity (ROH) arise due to inheritance of IBD haplotypes from both parents (without recombination or mutation), which happens more frequently with increasing parental relatedness. However, ROH can also emerge when shorter IBD haplotypes are inherited from apparently unrelated individuals due to background relatedness in the population (Korf 2013). Consequently, ROH may arise from populations that are, or have been, small in size due to evolutionary processes including population bottlenecks, inbreeding, genetic drift, as well as non-random mating and strong selection to maintain a single haplotype (Gibson *et al.* 2006). Thus, long ROH segments are expected in populations which have experienced contemporary inbreeding, while shorter segments indicate loss of genetic diversity from a historic founder effect or genetic bottleneck (McQuillan *et al.* 2008; Kardos, Qvarnström, *et al.* 2017; Gómez-Sánchez *et al.* 2018; Islam *et al.* 2019; Stoffel *et al.* 2020).

In addition to providing insight into contemporary or recent demographic processes, whole genome sequences can be used to reconstruct demographic history and estimate fluctuations in historical effective population sizes (N_e) over longer evolutionary time periods (see reviews by; Beichman *et al.* 2018; Mather *et al.* 2020). A range of modelling approaches can be used to reveal ancient dispersal, speciation events and population contractions or expansions, while comparisons of shared population history can be made to infer divergence timescales between populations or species (Terhorst *et al.* 2017; Patton *et al.* 2019; Excofffier *et al.* 2021). For example, pairwise sequentially Markovian coalescent (PSMC) models use patterns of heterozygosity to identify historical recombination events across a single diploid genome by inferring the time to the most recent common ancestor (TMRCA) for each independent DNA segment (Li and Durbin 2011). These have been used to infer times of dispersal or colonisation events and changes to population size across a wide range of animal systems (Xue *et al.* 2015; Nadachowska-Brzyska *et al.* 2016; Patton *et al.* 2019; Hooper *et al.* 2020; de Jager *et al.* 2021; Deng *et al.* 2021; Escoda and Castresana 2021; Kirch *et al.* 2021) and some plants (Izuno *et al.* 2016; Patil *et al.* 2021). A major benefit of this modelling approach is that it does not require specification of competing demographic models like many site-frequency spectra-based approaches (e.g.,

Fastsimcoal2; Excofffier *et al.* 2021). Together with recent estimates of demographic history using, for example, ROH and inbreeding, studies are able to produce estimates of both contemporary and ancient demographic history.

The island endemic Berthelot's pipit (*Anthus berthelotii*) together with its mainland sister species, the tawny pipit (*Anthus campestris*) offer an excellent model for understanding evolutionary processes (e.g., Illera *et al.* 2007; Spurgin *et al.* 2014; Gonzalez-Quevedo *et al.* 2015; Armstrong *et al.* 2018). The ancestor of Berthelot's pipit and tawny pipit likely colonised the Canary Islands from mainland Africa (Voelker 1999; see Fig. 1), and subsequently expanded to the Madeira and Selvagens archipelagos (Spurgin *et al.* 2014; Martin *et al.* 2021). Previous work, using microsatellites (Spurgin *et al.* 2014) and restriction-site associated DNA sequencing (Armstrong *et al.* 2018; Martin *et al.* 2021) has revealed strong bottlenecks associated with the two independent colonisation events from the Canary Islands to Madeira and Selvagens, estimated to have occurred ~8.5 thousand years ago (kya). Strong genetic population structure now exists between, but not within, Berthelot's pipit archipelagos, with no evidence of subsequent gene flow (Illera *et al.,* 2007; Spurgin *et al.,* 2014; Armstrong *et al.,* 2018; Martin *et al.* 2021), thus allowing us to study independent divergence histories and incipient speciation across the species range. As yet, little detail is known about genetic divergence from the tawny pipit to Berthelot's pipit, or how population history has shaped patterns of genome-wide genetic diversity across the Berthelot's pipit range.

Here we use whole genome resequencing to investigate patterns of contemporary genome-wide diversity and structure, and combine this with demographic reconstruction modelling, to quantify divergence timescales and ancient population history across Berthelot's pipit range. In addition, we sequence a genome of the tawny pipit, to assess genomic patterns of divergence through island colonisation and speciation. Specifically, we determine: (1) how genome-wide diversity and structure vary between populations and across archipelagos of Berthelot's pipit, and from the tawny pipit; (2) ancient demographic history of Berthelot's pipit (~5 Mya until 10 kya), and consider how results from whole-genome data compare with previous estimates derived using reduced marker sets; (3) how genetic diversity varies across individual genomes, and how these genomic landscapes differ between individuals. Finally, (4) using ROH detected across the genome, we investigate how their length and frequency increases with number of founding steps, bottleneck severity and population isolation.

3 Methods

3.1 Sample collection and reference genome sequencing

Berthelot's pipit samples from six island populations across the three archipelagos of its range (samples from Illera *et al.* 2007; Spurgin *et al.* 2012) (Fig.1) were selected for genome sequencing. One sample from a tawny pipit from coastal Mauritania was also sequenced. Birds, caught using mealworm larve (*Tenebrio molitor*) baited spring traps, were sampled from different locations across each island population to reduce the probability of sampling closely related individuals. Blood (~25 µl) was taken from each bird by brachial venipuncture and stored in 800 µl absolute ethanol at room temperature. We then extracted DNA using the salt extraction protocol described by Richardson *et al.* (2001), and molecularly sexed individuals (Griffiths *et al.* 1998). The quality of DNA extractions was confirmed by visualising the genomic DNA after electrophoresis on 1% agarose gels with ethidium bromide, and tested for impurities or protein contamination within 260/280 (1.8 – 2.0 nm) and 260/230 (1.75 – 2.4 nm) absorbance ratios using NanoDrop.

A draft Berthelot's pipit reference genome from a Porto Santo sample in the Madeiran Archipelago, generated by Armstrong *et al.* (2018), was used to map genome-wide sequence reads and to call genomic variants. This bird had low level of genome-wide heterozygosity. Sequencing of this reference genome was performed using Illumina paired-end reads (2 x 125 bp) on an Illumina HiSeq 2500 sequencer, assembled using DISCOVAR *de novo* (Weisenfeld *et al.* 2014), and assembly statistics were calculated with the *abyss-fac* utility in ABySS (Simpson *et al.* 2009). Genome completeness was assessed using CEGMA (Parra *et al.* 2007) which searched for 248 highly conserved core eukaryotic genes and BUSCO (Simao *et al.* 2015), to search for 3023 vertebrate-specific single copy orthologs. For full extraction, sequence and bioinformatics details see Armstrong *et al.* (2018).

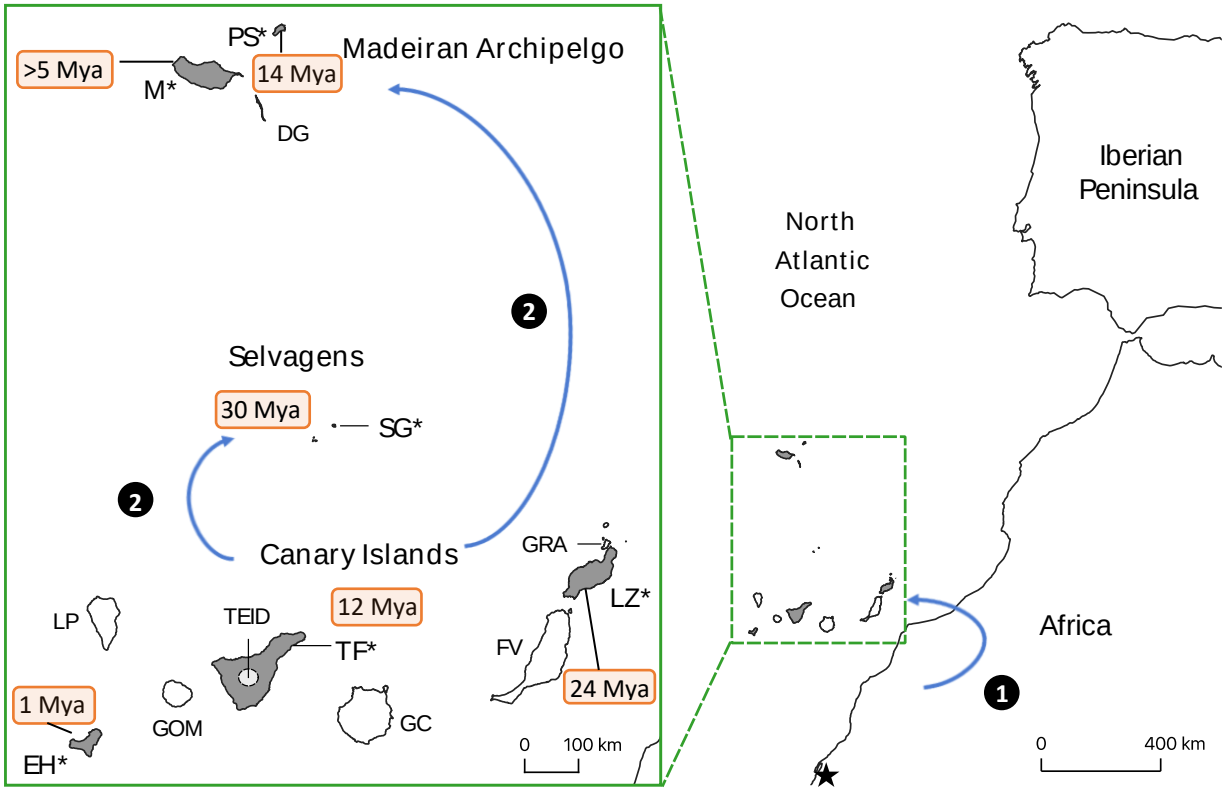


Figure 1. Berthelot's pipit range across three archipelagos and the sampling location of its sister species, the tawny pipit. Berthelot's pipit sample locations used for whole genome resequencing are denoted with an asterisk and the island shaded grey and geological age in orange boxes (data from Hoernle and Carracedo 2020). The tawny pipit was sampled on migration in Mauritania (Latitude: 17.991703°, Longitude: -16.016672°, see black star). The timing and direction of colonisation events is indicated by numbered arrows, with numbers indicating the number of between-archipelago founding steps separating the Berthelot's pipit populations from mainland Africa. Canary Island populations: El Hierro (EH), La Palma (LP), La Gomera (GOM), Teide (TEID) mountain population on Tenerife, Tenerife (TF), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ), La Graciosa (GRA). Madeiran populations: Madeira (M), Porto Santo (PS) and Deserta Grande (DG). Selvagens: Selvagem Grande (SG).

3.2 Genome resequencing, read alignment and variant calling

For Berthelot's pipit whole genome resequencing samples were selected to maximise the geographical range across archipelagos (Fig. 1). We selected two individuals per population, one male and one female, from El Hierro, Tenerife and Lanzarote (Canary Islands), Madeira and Porto Santo (one sample in addition to the reference sample) in the Madeiran archipelago and Selvagem Grande (Selvagens archipelago). All individuals chosen for sequencing were adult birds with no pox lesions and in which the presence of

haemoprotozoa parasites was not detected using a nested PCR approach (Waldenstrom *et al.,* 2014). One tawny pipit sample was also included.

Low Input, Transposase Enabled (LITE) Illumina compatible libraries were constructed at the Earlham Institute, Norwich, UK with a bespoke protocol using the Illumina Tagment DNA TDE1 enzyme and buffer kit (small 20034197/large 20034198). A total of 1ng of DNA was combined with 0.9 µl of Tagment DNA buffer and 0.1 µl Tagment enzyme TDE1 and 2 µl nuclease free water in a reaction volume of 5 µl and incubated for 10 minutes at 55˚C. Following the initial incubation, 5 µl of combined 2 µM custom barcoded P5 and P7 compatible primers, 4 µl 5x Kapa Robust 2G reaction buffer B, 0.4 µl 10mM dNTPs, 0.1 µl Kapa Robust 2G enzyme (Sigma Aldrich: KK5005) and 5.5 µl water were added and mixed, giving a total PCR volume of 20 µl. The DNA was then enriched with 14 cycles of PCR (72˚C for 3 minutes, 98°C for 3 minutes, 14 cycles of:95°C for 10 seconds, 62°C for 30 seconds, 72°C for 3 minutes, final hold at 4˚C). Post PCR, the DNA was cleaned (1.25x) using KAPA Pure Beads using the Tecan 480 robotics platform and final libraries were eluted in Ethidium Bromide. The size distribution of each library was determined using the Perkin Elmer GX Touch DNA High Sensitivity assay (DNA High Sensitivity Reagent Kit CLS760672), and a smear analysis on a 400-600bp size range was performed, this information was used to equimolar pool the libraries. Once pooled the samples were then subjected to size selection on a Blue Pippin 1.5% agarose cassette (R2 marker) from SAGE Science (BDF1510) which recovers molecules between 450-650bp. If more than one plate of samples was submitted the individual size selected plate pools were equimolar pooled creating the final sequencing pool. Where only one plate of samples had been submitted multiple lanes of the pool were size selected and combined with a 1x clean up. The quality of the final pool was determined using Agilent High Sensitivity DNA Kit from Agilent Technologies and the concentration measured with a High Sensitivity Qubit assay from ThermoFisher. Finally, a q-PCR was performed, and the pool was sequenced on the Illumina HiSeq4000 with a 150bp paired-end read metric. High throughput libraries were generated for each sample, pooled across 4 lanes of an Illumina HiSeq4000, for paired-end sequencing (2 x 150 bp). Read quality was assessed using FastQC with Phred quality score > Q30, indicating per-read base call accuracy > 99.9% (www.bioinformatics.babraham.ac.uk/projects/fastqc/).

Raw paired-end Illumina sequence reads (2 x 150 bp) were merged at the individual level (across 4 sequence lanes) and aligned to the indexed reference Berthelot's pipit genome using the "bwa mem" algorithm (suitable for sequence reads between 70 bp and 1 Mb) in Burrows-Wheeler Aligner (BWA) v. 0.7.12 (https://github.com/lh3/bwa), with default

parameters (Li 2013). Once mapped, potential duplicate PCR reads were flagged using Picard tools (http://broadinstitute.github.io/picard/) *MarkDuplicates* function in Genome Analysis Toolkit (GATK) v. 4.1 (McKenna *et al.* 2010). The Picard software was subsequently used to assign read group information and to validate binary alignment files (.bam) before variant discovery. Variant calling was then performed on each sample using GATK *HaplotypeCaller* in GVCF ("Genotype VCF") mode, removing flagged duplicated or poor quality reads based on default parameters. Joint genotyping was then performed across samples for each contig using GATK's *GenomicsDBImport* and *GenotypeGVCF* tools. To improve the accuracy of variant discovery and genotyping, variants were determined simultaneously across the 11 Berthelot's pipit samples and the tawny pipit sample for each dataset, according to GATK's best practice recommendations. Contig-level VCF files were then combined using GATK *SortVcf*, with variants mapped to contigs less than 500 bp removed. Base quality score recalibration and indel realignment, may improve variant discovery and genotype calls but require knowledge of true variant sites that are unavailable for many non-model species including Berthelot's pipit. To ensure high quality in our datasets, we applied stringent post- variant calling filtering to our VCF files. Unmapped reads and mapped reads with a root-mean-squared read mapping quality (MQ) below 25 were discarded. Variants were then filtered for read strand bias (Fisher's exact test > 60 and Strand Odds Ratio > 3) and quality by depth (QD < 2) using GATK, to account for errors in read mapping.

3.3 Variant mapping and filtering for genetic diversity analyses

As the draft reference genome for Berthelot's pipit is only assembled to the contig level (Armstrong *et al.* 2018), Berthelot's pipit contigs were mapped to chromosomes of the Zebra finch (*Taeniopygia guttata)* genome assembly bTaeGut1_v1.p (NCBI Assembly GCA_003957595.1) using SatsumaSynteny (Grabherr *et al.* 2010), which performs well on fragmented genome assemblies (Lui *et al.* 2018). Output from Satsuma Synteny was used to assign contigs to chromosomes, and determine their order, location and orientation. Finally, variants from GATK outputted VCF files were mapped against the Satsuma Synteny output and reassigned to chromosomes using custom R scripts (RStudio Team, 2016).

We generated three final VCF files to maximise variants that could be included in each analysis; "All Pipits" with variants joint called across the 11 Berthelot's pipit and one tawny pipit, "Berthelot's" dataset with variants joint called across 11 Berthelot's pipit, and "Tawny

pipit" dataset. We then filtered for genotype quality and coverage in VCFtools v. 0.1.15 (Danecek *et al.* 2011), to ensure high quality of SNPs in our datasets. We first removed unmapped sites (--not-chr 0), sites with >2 alleles (--max-alleles 2), indels (--remove-indels) and variants with Phred-scaled quality less than 30, i.e., variant accuracy >99.9% (--minQ 30), across all individuals. To minimize the impact of collapsed regions in the assembly, we also removed all sites at which mean read depth (among all individuals in the dataset) was less than 10 or more than twice the average read depth across the genome (>45 for all pipits, >44 for all Berthelot's pipit, >55 for tawny pipit) (--min-meanDP 10, max-meanDP 45/44/55). We removed sites with more than 4 failed genotype calls (--max-missing-count 4) and excluded the Z chromosome from all analyses as females have systematic biases related to coverage that could affect estimates of differentiation (--not-chr 31). This resulted in three mapped and quality trimmed VCF files for the 11 Berthelot's pipit, 1 tawny pipit sample and all 12 samples together. Individual level data for the quality filtered and mapped "Tawny pipit" and "Berthelot's pipit" variants is summarised in Table 1.

Table 1. Berthelot's pipit and tawny pipit sampling information and genome sequencing for variants mapped to the Zebra finch genome. Mean variant coverage and individual missingness are calculated after mapping and quality filtering of variants. *n* filtered loci = number of loci retained after mapping to Zebra finch genome and quality filtering.

Archipelago / Location	Pop. code	Sex (Indiv. ID)	Mean read coverage (X)	Mean variant coverage (X)	Individual missing variants (%)	*n* raw loci	*n* filtered loci
Mauritania, Mainland Africa	TAW	M (462)	27.3	27.6	4.0	11,912,976	7,081,760
Lanzarote, Canary Islands	LZ	M (87) F (93)	30.1 25.2	30.0 26.8	1.7 3.1	10,361,030 10,361,120	5,575,900 5,575,857
Tenerife, Canary Islands	TF	M (17) F (6)	25.0 23.4	25.8 25.0	2.0 3.0	10,360,653 10,359,854	5,575,901 5,575,869
El Hierro, Canary Islands	EH	M (179) F (161)	23.0 21.9	23.7 23.2	2.1 2.1	10,360,520 10,359,970	5,575,891 5,575,878
Madeira, Madeiran Arch	M	M (249) F (305)	24.3 22.0	25.0 23.3	1.6 2.0	10,361,593 10,361,072	5,575,902 5,575,905
Porto Santo, Madeiran Arch	PS	F (506)	20.2	21.6	3.5	10,360,994	5,575,865
Selvagem Grande, Selvagens Arch	SG	M (278) F (300)	20.9 23.1	21.9 24.2	2.8 2.2	10,360,007 10,360,060	5,575,862 5,575,896

3.4 Genome-wide inbreeding and nucleotide diversity

Genetic diversity within populations was measured as average observed heterozygosity (H_o), inbreeding coefficient (F_{IS}) and windowed nucleotide diversity (π) using VCFtools. To provide estimates of genome-wide nucleotide diversity, we first calculated per-site nucleotide diversity and then generated a genome-wide mean with 95% confidence intervals for each individual. We calculated per-individual inbreeding coefficients (F_{IS}) across the Berthelot's pipit genomes, based on the mapped and quality filtered marker sets in Plink 1.9 (Chang *et al.* 2015). This method of calculating inbreeding, which uses a single-point calculation, simply reflects the proportion of heterozygous loci and is not sensitive to the presence of LD (Polašek *et al.* 2010). We also used Plink to calculate individual inbreeding coefficients based on genome-wide estimates of heterozygosity; values were strongly correlated to those calculated by VCFtools (Pearson correlation; $r = 0.998$), so we only report the VCFtools calculated values.

We also generate estimates of individual inbreeding coefficients based on ROH (see below for methods to identify ROH) by calculating the proportion of the autosomal genome that is covered by ROH segments above a specified length, F_{ROH} , McQuillan *et al.* (2008):

$$F_{ROH} = \sum L_{ROH} / L_{TOTAL}$$

where L_{ROH} and L_{TOTAL} are the total length of all ROH segments and the genome, respectively. The size of the autosomal genome was considered as ~ 1,057 Mb according to the Zebra finch reference genome assembly bTaeGut1_v1.p (NCBI Assembly GCA_003957595.1), used in this study. The correlation between the F_{ROH} and F_{IS} were measured using Pearson's correlation.

3.5 Divergence across archipelagos

To visualise genome-wide structure between the tawny pipit and among the three archipelagos of Berthelot's pipit, we conducted a principal component analysis (PCA) implemented in Plink using the "All Pipits" quality trimmed dataset. The strength of genetic divergence between each population was then assessed using Wright's fixation index (F_{ST}), a measure of relative divergence (accurate divergence estimates using genomes, Willing *et al.* 2012). Pairwise F_{ST} values were calculated in 50 kb SNP windows, between each

population using VCFtools (Danecek et al., 2011). Mantel tests were used to test for associations between PSMC colonisation time estimates (see below) and mean pairwise F_{ST} values. Mantel test *p*-value estimates were generated from 100,000 randomised permutations, performed using the ade4 package in R (Dray and Dufour 2007).

3.6 Divergence timescales and N_e over time

Historical fluctuations in effective population size (N_e) were estimated from 5 Mya until approximately 10 kya using single genomes in Pairwise sequential Markovian coalescent (PSMC) models. Therefore, we followed population trends from before the point of island colonisation and speciation of Berthelot's pipit and through initial stages of divergence across the three north Atlantic archipelagos. Although MSMC's implementation of PSMC' has improved estimation of recombination rates, we chose to use PSMC as the algorithm currently performs more accurately on highly fragmented genome assemblies (Gower *et al.* 2018). PSMC analyses require a consensus genome sequence (fastq) that can be filtered for coverage and sequencing errors. Using individual level .bam files with duplicate PCR reads marked, we generated consensus sequences for one individual per population using the *mpileup* command (with -C50 to adjust the mapping quality for reads containing excessive mismatches) in SAMtools (Danecek *et al.* 2021) and the *vcf2fq* command from vcfutils.pl, with the Berthelot's pipit reference genome assembly as the reference. We filtered each consensus sequence by excluding sites at which the root-mean-square mapping quality of reads covering the site was below 25, the inferred consensus quality was below 20, and the variant read depth was either more than twice the average or less than 10X across the genome. All genomes had a mean read coverage >20X, variant coverage >19X, and very low levels of individual missingness (<5 %), enabling accurate estimation of genotype states for most sites (Han *et al.* 2014), which follows filtering recommendations as used to infer demographic history from PSMC modelling in other avian studies (Nadachowska-Brzyska *et al.* 2015; Nadachowska-Brzyska *et al.* 2016).

Recent and drastic bottlenecks, and associated inbreeding, can lower recent PSMC-based N_e estimates and erase information about ancient dynamics. Therefore, for each population, the individual with the lowest estimated inbreeding coefficients (F_{IS} and F_{ROH}, see Table 2) was used. However, it is unlikely that inbreeding causes drastically differing estimates of past demography using PSMC (Nadachowska-Brzyska *et al.* 2015; Nadachowska-Brzyska *et al.* 2016). We performed the PSMC analyses using the following fixed parameters across

each individual: maximum number of iterations (N) of 30, maximum coalescent time (t) of five, initial theta/rho ratio (r) of one and parameter pattern (p) of '4+30*2+4+6+10'. The above parameters were able to provide good resolution and showed more than 10 recombination events in each of the atomic time intervals within 20 iterations. These values were chosen in line with PSMC analyses conducted across other avian species (36 avian species; see Nadachowska-Brzyska *et al.* 2015).

To scale outputs from PSMC to real time we use estimates of generation time and neutral mutation rate in the psmc_plot.pl command. Berthelot's pipits reach sexual maturity at one year, with annual adult survival 0.55 and maximum longevity ~6 years. Using these estimates of life history information, Bird *et al.* (2020) provide estimates of generation length of 2.05 and 2.20 years for the Berthelot's pipit and tawny pipit, respectively. In light of these estimates we use a generation time of 2.20 to scale the outputs from PSMC analyses across the Berthelot's and tawny pipit range. However, previous estimates of generation time in the Berthelot's pipit using observational studies have been as much as 3.7 years (Garcia-del-rey and Cresswell 2007). Neutral mutation rate is not quantified for many avian non model species. Here we use recent estimates of 2.3×10^{-9} from the collared flycatcher (*Ficedula albicollis*), derived using a three-generation pedigree (Smeds *et al.* 2016), which is near the average (2.28×10^{-9}) reported across 38 avian species (Nadachowska-Brzyska *et al.* 2015).

3.7 Runs of homozygosity and patterns diversity across genomes

We identified the length and distribution of autozygous IBD regions across Berthelot's pipit genomes as a signature of recent and past demographic events. We used our SNP datasets with stringent genotype accuracy trimming (depth and quality, as above) to identify ROH across individual genomes. We implemented the *--homozyg* function in Plink to identify the length and location of ROH. Long ROHs (>1 Mb) are indicative of IBD and are a product of recent demography such as inbreeding and population size contraction, whereas ROHs across shorter chromosome fragments are indicative of ancient population processes (Boyko *et al.* 2010; Curik *et al.* 2014; Pilot *et al.* 2014).

A threshold was set for the minimum length (kb) of homozygosity for a segment to be considered a ROH. Because strong LD, typically extending up to 50 kb, is common throughout the genome, especially across bottlenecked Berthelot's pipit populations in the Madeiran archipelago and Selvagens (Martin *et al.* 2021), short tracts of homozygosity are

very prevalent. As our aim was to detect and compare IBD segments to infer differing population demography across the Berthelot's pipit geographic range, parameters for detecting ROH were consistent across all Berthelot's pipit populations. As recommended by Meyermans *et al.* (2020), we do not perform LD or MAF trimming prior to ROH detection. Instead, we consider two size categories of ROH implemented in Plink via the *--homozyg* function; 1) long ROHs >1 Mb (*--homozyg-kb* 1000), to exclude ROH likely to be derived from ancient population processes; 2) shorter ROH >250 kb (*--homozyg-kb* 250), likely to reflect ancient, as well as recent, population processes. As the Berthelot's pipit have a relatively short generation time (~2 years), long ROH would be expected to reflect the very recent past.

We then defined the following parameters based on our assessment of sequence quality and genome SNP densities (0.1 kb / SNP), following Meyermans *et al.* (2020) recommendations. We set a threshold for the minimum scanning window size to 50 SNPs (*--homozyg-window-snp* 50); a minimum density of one SNP per 200 kb on average (*--homozyg-density* 200; 200 kb /SNP); and a maximum gap between consecutive SNPs of 200 kb (*--homozyg-gap* 200). We account for occasional heterozygous positions within ROH resulting from sequencing errors, read mapping errors and occasional mutations. Specifically, we accepted that 2% of SNPs would be heterozygous within IBD segments (*--homozyg-window-het* 2) and allowed for up to 5 missing genotype calls within a scanning window (*--homozy-window-missing* 5).

To visualise the landscape of genetic diversity across individual genomes, we calculated nucleotide diversity across two window sizes (250 kb and 2 Mb, to assess diversity patterns at different genomic scales), each with a 20% smoothing step. We then map the locations of long ROH (>1 Mb) and short ROH (>250 kb) against genomic patterns of nucleotide diversity, to visually compare ROH distribution between individuals and populations.

4 Results

4.1 Whole Genome resequencing

Sequencing of Berthelot's pipit and tawny pipit samples resulted in 1,030,115,042 paired-end reads (80 $x10^6$ – 120 $x10^6$ per individual), with a mean insert size of 401 bp. Genome alignment and mapping resulted in mean read coverage of 23.6 X ± 2.6 s.d. per individual,

relative to the Zebra finch's 1.1 Gb genome (Warren *et al.* 2010). Reads were then mapped to the contig level assembly of the Berthelot's pipit reference genome and genotypes joint called, resulting in 19,781,461 raw "All Pipits" variants of which 13,253,579 (67.6%) were mapped to the Zebra finch chromosomes. The "Berthelot's" dataset resulted in 10,363,127 raw variants of which 6,953,309 (67.1%) were mapped to the Zebra finch chromosomes.

Subsequent quality filtering resulted in a "All Pipits" data set with 11,575,905 autosomal mapped SNPs and "Berthelot's" dataset with 5,575,905 (where indels and SNPs with > 2 genotypes were removed, the minor allele count was ≥ 1; > 99.9% genotype variant accuracy; genotype coverage range (i.e., depth per allele) = 10-44/45; and maximum of four missing genotypes across all individuals). Individuals had low levels of missing data even prior to variant quality filtering, with no individuals having > 5% missing data. Final depth of coverage for the quality filtered SNPs was high, with a mean 24.6 for the "Berthelot's" dataset (Table 1) and 22.6 X for the "All Pipits" dataset.

4.2 Loss of genetic diversity during island colonisation and bottlenecks

Genome-wide nucleotide diversity, heterozygosity and inbreeding for each individual are shown in Table 2. The largest reduction in all diversity measures was between the tawny pipit and Berthelot's pipit. In the mainland tawny pipit average heterozygosity (H_o) across polymorphic SNPs was high, 0.405. Across Berthelot's pipit populations in the Canary Islands heterozygosity was 0.127 - 0.135 with lowest diversity in the eastern island of El Hierro and highest in Lanzarote on the western edge of the archipelago (Table 2). Heterozygosity was much lower in the Madeiran archipelago (0.101 – 0.107) and the Selvagens (0.082 – 0.092). Genome-wide nucleotide diversity showed a similar pattern with reduced diversity across the Madeiran (0.0011 – 0.0012) and Selvagens (0.0008 – 0.0010) archipelagos compared to the Canaries. Inbreeding coefficients varied substantially between Berthelot's pipit populations and tawny pipit (Table 2), with near-absence of inbreeding in the mainland tawny pipit, increasing an order of magnitude across the Canary Islands, and high levels of inbreeding in both Madeiran populations (F_{IS} = 0.233 - 0.261); and exceptionally high levels in Selvagens (F_{IS} = 0.325 - 0.480).

Table 2. Genome-wide genetic diversity per individual across Berthelot's pipit populations. Mean π = mean per-site nucleotide diversity, H_O = proportion of heterozygous sites. * denotes individuals presented in Figure 7.

Archipelago / Location	Pop. Code	Sex (Indiv. ID)	Mean π	95% CI π	H_O	$F_{ROH >250 kb}$ / F_{IS}
Mauritania, Mainland Africa	TAW	M (462)	0.0047	+/- 0.0010	0.405	0.002/0.000
Canary Islands	LZ	M (87) F (93)	0.0015 0.0015	+/- 0.0003 +/- 0.0003	0.135 0.133	0.015/ 0.019 0.016/ 0.010
Canary Islands	TF	M (17)* F (6)	0.0015 0.0014	+/- 0.0004 +/- 0.0001	0.134 0.130	0.008/ 0.001 0.039/ 0.044
Canary Islands	EH	M (179) F (161)	0.0014 0.0014	+/- 0.0000 +/- 0.0002	0.127 0.132	0.039/ 0.051 0.032/ 0.047
Madeiran	M	M (249)* F (305)	0.0011 0.0012	+/- 0.0000 +/- 0.0000	0.101 0.107	0.138/ 0.248 0.130/ 0.233
Madeiran	PS	F (506)	0.0011	+/- 0.0000	0.101	0.136/ 0.261
Selvagens	SG	M (278)* F (300)*	0.0010 0.0008	+/- 0.0000 +/- 0.0000	0.092 0.082	0.248/ 0.325 0.377/ 0.480

A PCA using the "All Pipits" SNPs (Fig. 2A), showed that the strongest levels of genomic differentiation are between the tawny pipit and Berthelot's pipit, with the first principal component explaining 7.8% of variation. Using only the "Berthelot's" dataset to perform a PCA, populations separated by archipelago along the first principal component, with a gradient from Selvagens to the Canary Islands to Madeira describing just 2% of genomic variation. Pairwise F_{ST} results reflect those from the PCAs. Pairwise F_{ST} between the tawny pipit and the Berthelot's pipit populations were high, with F_{ST} >0.42 across the Canary Islands and 0.54 and 0.51 for the Selvagens and Madeira, respectively. Closely located islands, separated by within-archipelago founding events had low F_{ST} < 0.088, while those with a single between-archipelago founding event were marginally higher (F_{ST} = 0.033 – 0.119); and two independent between-archipelago founding events, which separate Madeira and the Selvagens had substantially higher differentiation (F_{ST} = 0.214 – 0.261; Table 3). Mantel tests showed there was a significant association between genome-wide F_{ST} and log-transformed divergence timeframes since colonisation across the range (r = 0.93, p = 0.044).

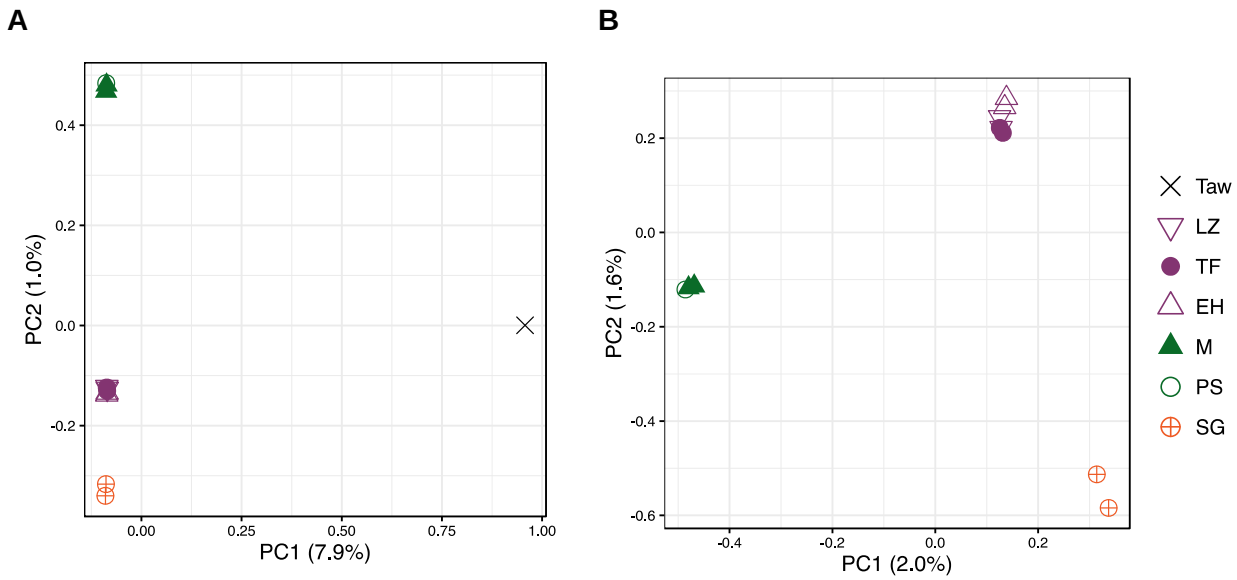


Figure 2. A) Principal component analysis (PCA) based on the "All Pipits" genomic autosomal SNPs among six Berthelot's pipit populations and one tawny pipit; and **B)** "Berthelot's" PCA between Berthelot's pipit island populations across the three colonised north Atlantic archipelgos. Populations are coloured by their archipelgo location to highlight geographic population clustering.

Table 3. Pairwise F_{ST} between island populations of Berthelot's pipit with differing levels of founding events (above diagonal). Divergence times are estimated from shared ancestry based on PSMC effective population sizes, N_e (below diagonal). Populations separated by within-archipelago founding events with potential for gene flow are coloured grey, a single founding event between archipelagos in blue and two between-archipelago founding steps in orange.

	TAW	TF	LZ	EH	SG	M	PS
TAW	-	0.424	0.431	0.438	0.538	0.511*	0.511*
TF	2.1 Mya	-	0.019	0.026	0.106	0.096	0.033
LZ	2.1 Mya	45 kya	-	0.026	0.106	0.098	0.037
EH	2.1 Mya	35 kya	<10 kya	-	0.119	0.109	0.054
SG	2.1 Mya	15-25 kya	45 kya	40 kya	-	0.214	0.261
M	2.1 Mya	50 kya	50 kya	40 kya	40 kya	-	0.088
PS	2.1 Mya	50 kya	50 kya	40 kya	40 kya	<10 kya	-

* = Pairwise F_{ST} comparison between the tawny pipit and Madeiran archipelago populations calculated across both populations, as pairwise F_{ST} cannot be calculated between just two individuals for PS.

4.3 PSMC analysis of population history

We used PSMC modelling to infer fluctuations in population size from five Mya until 10 kya in Berthelot's and tawny pipits. Looking backwards in time, our results showed that the N_e curves of Berthelot's and tawny pipits converged from about 2.1 Mya, at N_e 200,000, indicating a shared ancestry and demography, and then started to diverge to form distinct and non-overlapping population histories since (Fig. 3). The curve for the tawny pipit indicates a higher ancestral N_e than across the island range of the Berthelot's pipit, with strong population growth until 150 kya, and more recent N_e estimates at least ten-fold higher than for the Berthelot's pipit (Fig. 3).

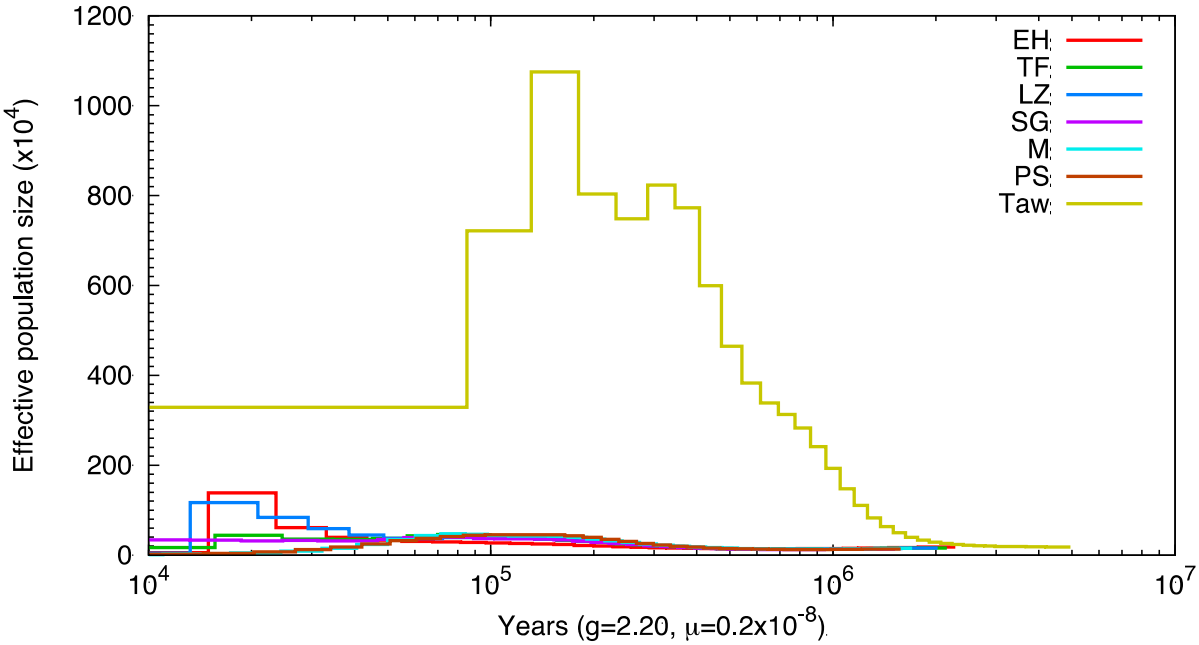


Figure 3. PSMC modelled divergence between the mainland tawny pipit and the Berthelot's pipit. Estimates of effective population size, N_e, through time for six Berthelot's pipit populations and the tawny pipit genome. Each line represents one individual, coloured by population. Individuals had no more than 5% missing data and a mean genome-wide coverage >19X.

Across the contemporary Berthelot's pipit range, the PSMC results clearly indicated that all six analysed populations shared ancestry and demography for most of the investigated time period (Fig. 4). The species N_e started to increase from ~200 000 approximately 1 Mya, until approximately 50 kya when the populations of Madeira and Porto Santo experienced a gradual decline in N_e. Meanwhile, two Canary Island populations experienced continued

population expansion until 20 kya. During these last 50 thousand years N_e have remained fairly stable across the central Canary Island of Tenerife and the Selvagens. Differences in divergence histories between the Selvagens and Madeiran archipelago are consistent when the upper estimate of generation time (3.7 years) is used, while timescales for reductions in N_e shift deeper into the past (Fig. S3). This also reveals strong decline in N_e within both Selvagens individuals approximately 11-14 thousand years ago, which is not plotted when the standard generation time of 2.2 years is used.

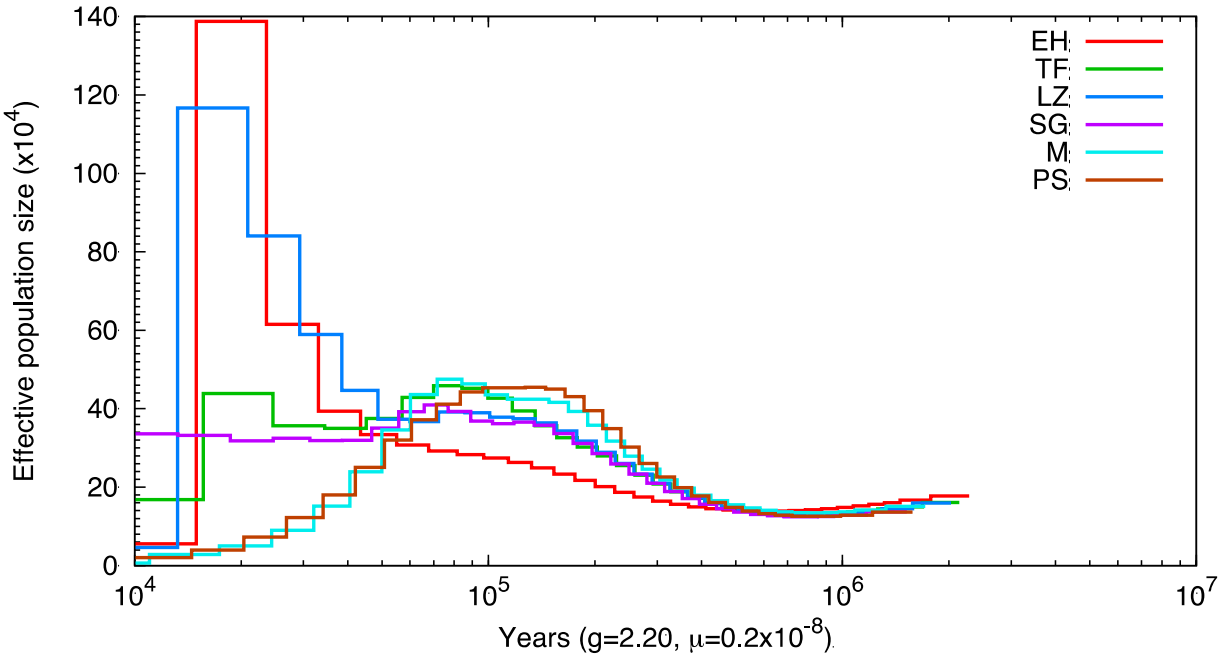


Figure 4. PSMC estimates of changes in effective population size, N_e, over time for six contemporary Berthelot's pipit populations across three archipelagos in the north Atlantic, from 2 Mya until 10 kya.

4.4 Landscapes of diversity and signatures of indentity-by-descent across genomes

The landscape of nuclotide diversity (π) varied significantly across individual genomes, with peaks and valleys of diversity within individual chromosomes (Fig. 5B). Broadly, patterns of diversity within chromosomes are reflected across individuals (i.e., shared locations of peaks and valleys between individuals), with similar patterns in the tawny pipit and across the three Berthelot's pipit archipelagos (Fig. 5B and S1). However, while patterns of diversity are similar, absolute diversity is three-fold higher in the tawny pipit compared to the average in

Berthelot's pipit and there are only a few regions of the genome where diversity is comparable. Low levels of genetic diversity, characterised by large regions with very low diversity, are observed in the more recently colonised archipelagos of Maderia and the Selvagens.

Signatures of IBD were statistically quantified by identifying ROH within two length criteria; >250 kb and >1 Mb. No long ROHs, >1 Mb, were detected in the tawny pipit genome and only a small number were detected across the Canary Island populations (Table 4, Fig. S1). These signatures were very strong in the Selvagens with long ROHs, >1 Mb, extending across 10.8 – 12.1% (130-145 Mb) of the genome, with the highest density of ROH on chromsomes 2 and 3 (Fig. 5B; Fig. S1). ROH are also prevelant across genomes of the Maderian archipelago, with similar prevalence across the two sampled popualtions covering 11.4 – 12.2% (137 – 146 Mb) of the genome (Table 4; Fig. S1). The location of ROH varies strongly even between individuals within the same populations (Fig. 5B; Fig. 7). However, some ROH locations are shared between individuals (see for example the Madeiran individuals, Fig. S1).

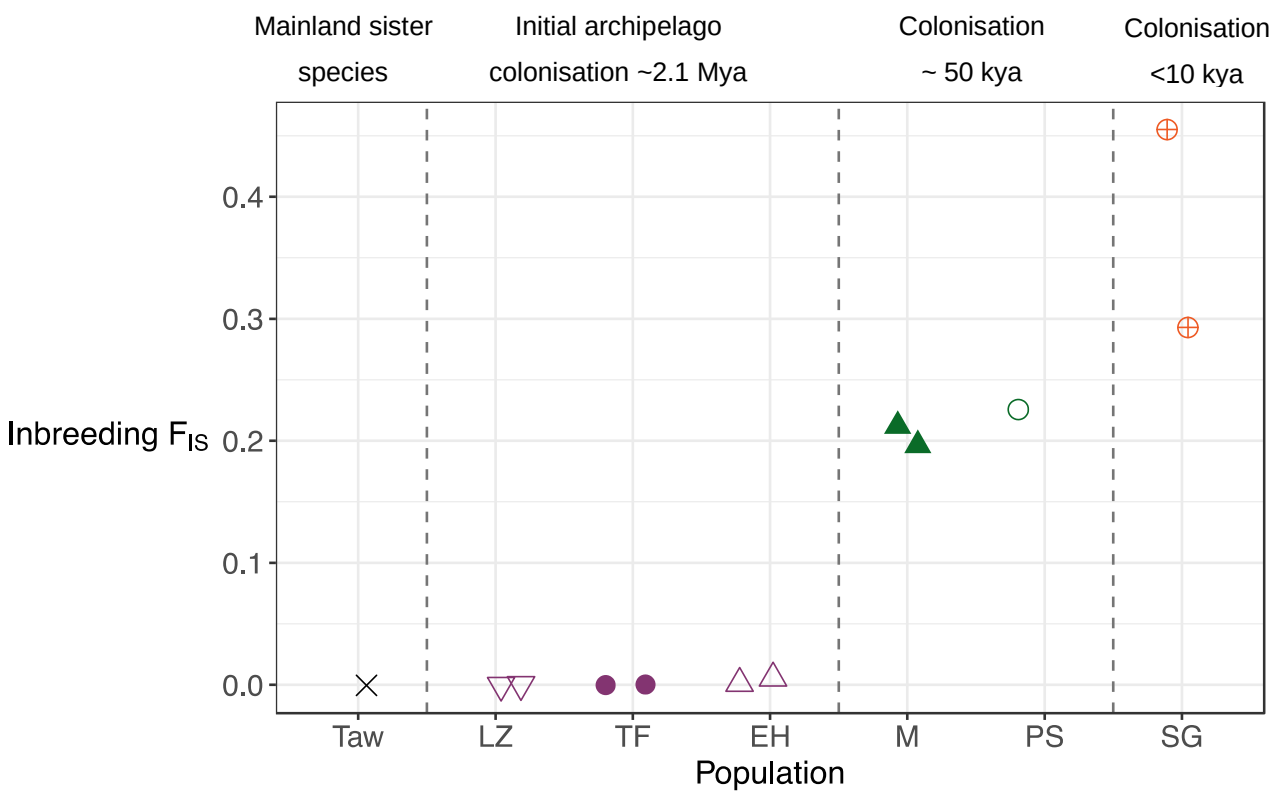
Table 4. Long ROH (>1 Mb) across populations of Berthelot's pipit. Archipelago populations are separated by grey dotted lines. * = Long ROH plotted in Fig. 5B.

Archipelago / Location	Pop. Code	Sex	# All ROH (>250 kb)*	Total length All ROH (kb)	# Short ROH (250 kb – 1 Mb)	Total length Short ROH (kb)	# long ROH (>1 Mb)	Total length long ROH (kb)
Mauritania, Mainland Africa	Taw	M*	5	1,749	5	1,749	0	0
Canary Islands	LZ	M	27	16,331	24	11,549	3	4,781
	LZ	F	33	16,537	30	12,863	3	3,673
Canary Islands	TF	M*	19	8,808	19	8,808	0	0
	TF	F	65	41,552	54	26,993	11	14,559
Canary Islands	EH	M	70	41,381	63	27,102	7	14,278
	EH	F	63	33,398	58	26,726	5	6,672
Madeiran	M	M*	286	146,339	266	118,311	20	28,028
	M	F	285	137,354	264	110,350	21	27,004
Madeiran	PS	F	296	143,844	280	122,945	16	20,899
Selvagens	SG	M*	327	262,107	254	131,667	73	130,440
	SG	F*	594	398,181	280	253,352	94	144,829

The average length of ROH was similar across the species range but much more numerous in the Madeiran and Selvagens archipelagos (Table 4). Within individuals, ROH are clustered in chromosomal regions, suggesting these originated as larger autozygous chunks, which have been eroded by mutations and recombination. As well as clustered ROH, we observe few stretches of 2 - 5.5 Mb IBD segements across both Selvagens birds, the Maderian islands and one bird from El Hierro (a small and isolated population in the Canary Islands).

We calculated a measure of inbreeding using the proportion of the genome in ROH >250 kb, F_{ROH} (Table 2). Across all populations, measures of inbreeding based on genome-wide heterozygosity (F_{IS}) were strongly correlated with those calculated based on the proportion of an individual's genome in ROH ($F_{ROH\ >250\ kb}$) (Pearson correlation; r = 0.977, Fig. S2). It is important to note that substantially lower F_{ROH} are inferred when only ROH >1 Mb are considered, as only a small proportion of autozygous segments are in contiguous loci at least 1 Mb in length (Table 4). The proportion of the genome in ROH (F_{ROH}) was very low for the tawny pipit, with only five short segments detected (F_{ROH}= 0.002), and few generally short segments were detected across the Canary Island populations (F_{ROH} = 0.008 – 0.039). Populations that experienced historic founder effects had substantially greater proportion of their genome in ROH (Fig. 5B and 6, Selvagens F_{ROH} = 0.248 – 0.377; Porto Santo F_{ROH} = 0.136, Madeira F_{ROH} = 0.130 - 0.138). Short ROH were far more prevalent than ROH >1 Mb across all populations of the Berthelot's pipit, representing approximately 1/3 of total ROH detected across genomes (Fig. 6B). While a similar number of short ROH were detected across the Madeiran and Selvagens archipelagos, many more (> 3x) long ROH were detected across the Selvagens. Finally, variance in inbreeding within populations is also apparent within Tenerife, the Selvagens and El Hierro - the total number, length and genomic location of ROH segments varied between individuals within a population (Fig. 6 and Fig. S1).

A


Mainland sister species
Initial archipelago colonisation ~2.1 Mya
Colonisation ~ 50 kya
Colonisation <10 kya
Inbreeding F_{IS}
0.4
0.3
0.2
0.1
0.0
Taw
LZ
TF
EH
M
PS
SG
Population

B

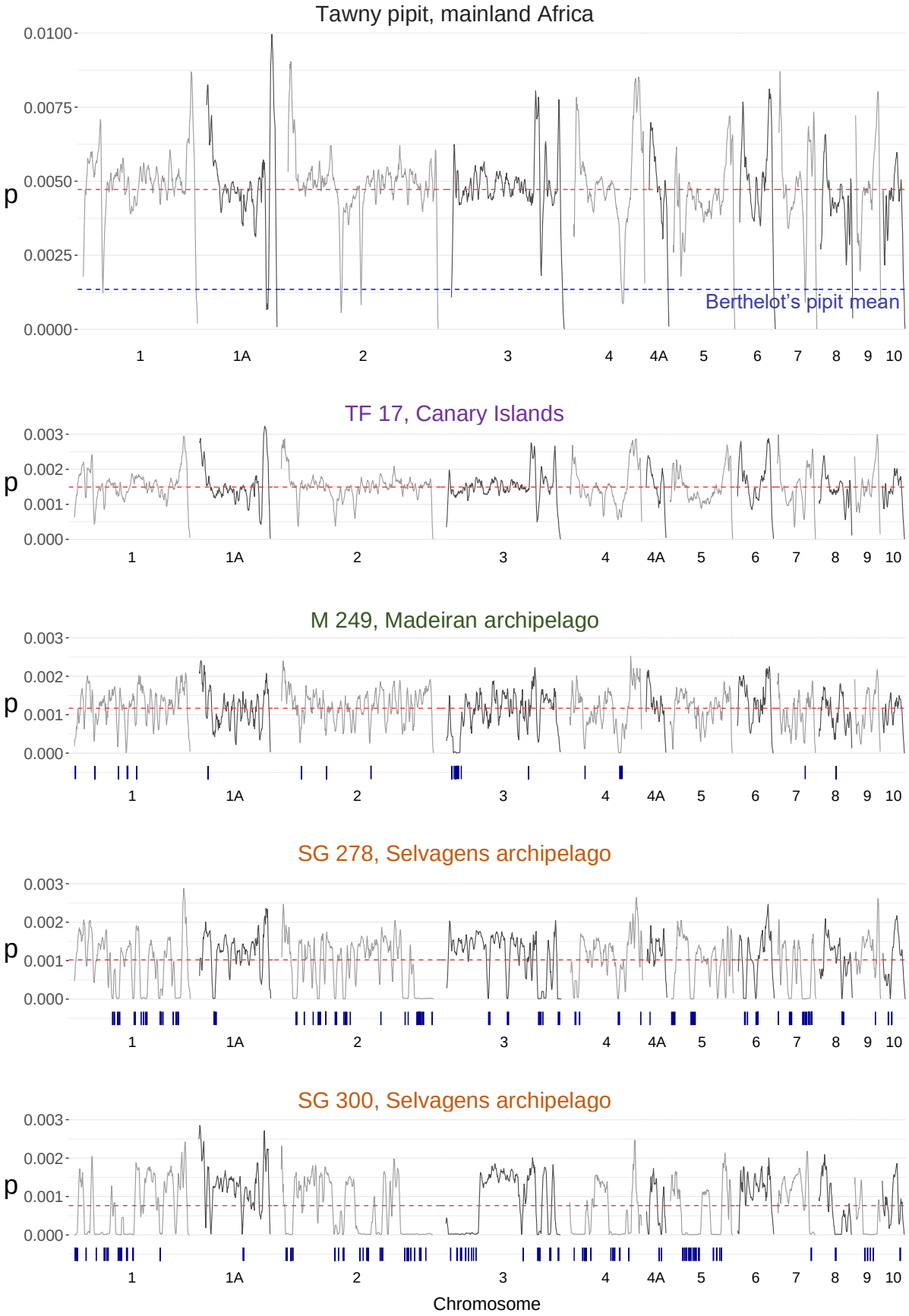

Tawny pipit, mainland Africa
p
0.0100
0.0075
0.0050
0.0025
0.0000
Berthelot's pipit mean
TF 17, Canary Islands
0.003
0.002
0.001
0.000
M 249, Madeiran archipelago
SG 278, Selvagens archipelago
SG 300, Selvagens archipelago
1
1A
2
3
4
4A
5
6
7
8
9
10
Chromosome

Figure 5. Genome-wide patterns of nucleotide diversity (π), inbreeding and ROH across island populations of Berthelot's pipit and its mainland sister species, the tawny pipit. **A)** Individual inbreeding coefficient, F_{IS}, across 6 island populations of Berthelot's pipit and the tawny pipit. $F_{IS} = 0$ suggests random mating. Archipelagos are separated by grey vertical dotted lines. **B)** Nucleotide diversity across the genome of the tawny pipit, and four Berthelot's pipit individuals with low, moderate and high levels of inbreeding using 2 Mb windows with 20% overlap (grey lines). Horizontal red dotted lines represent mean nucleotide diversity per sample calculated across all autosomes, and blue blocks are runs of homozygosity (ROHs) of at least 1 Mb in length. Macrochromosomes 1A, 4A and 1-10 are presented for visual comparison between individuals.

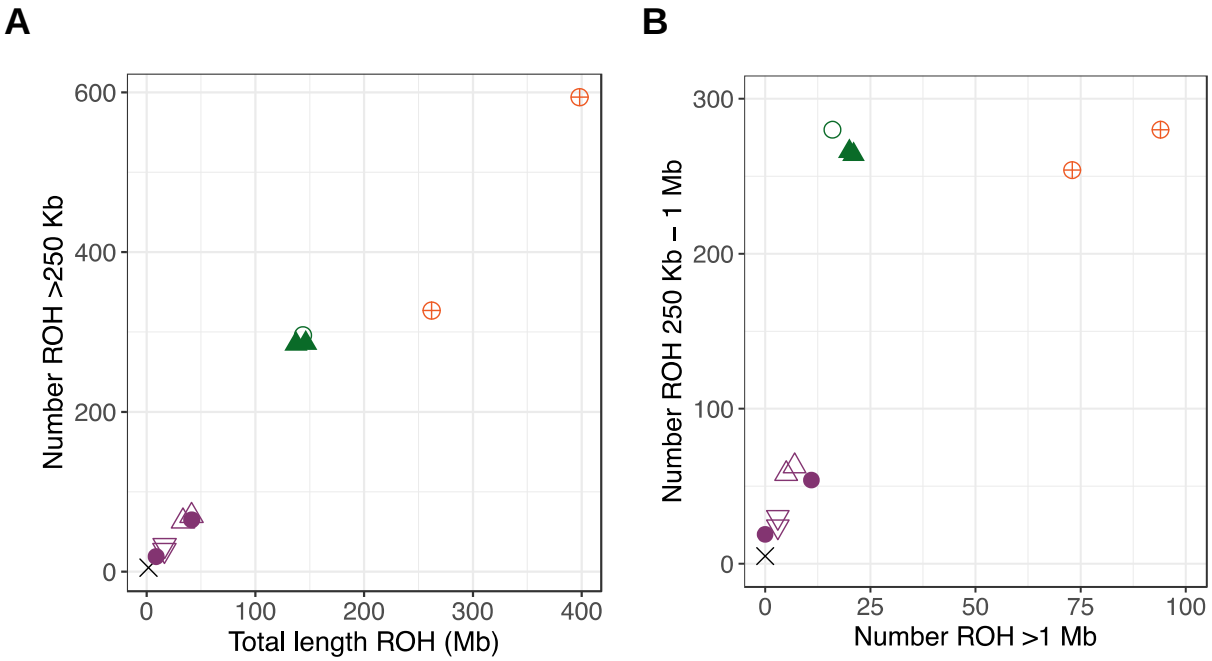


Figure 6. Size classes of genome-wide runs of homozygosity (ROH) for the tawny pipit and populations of Berthelot's pipit. **A)** The relationship between the number of ROH detected covering a chomosome segement at least 250 kb in size and the cumulative length of the genome in ROH. **B)** Number of ROH detected in individual genomes between 250 kb and 1 Mb relative to the ROH of at least 1 Mb in length. Populations are coloured by their archipelago; black cross = tawny pipit, filled circle = Tenerife, empty upwards triangle = El Hierro, downwards empty triangle = Lanzarote, empty circle = Porto Santo, filled triangle = Madeira, crossed circle = Selvgens.

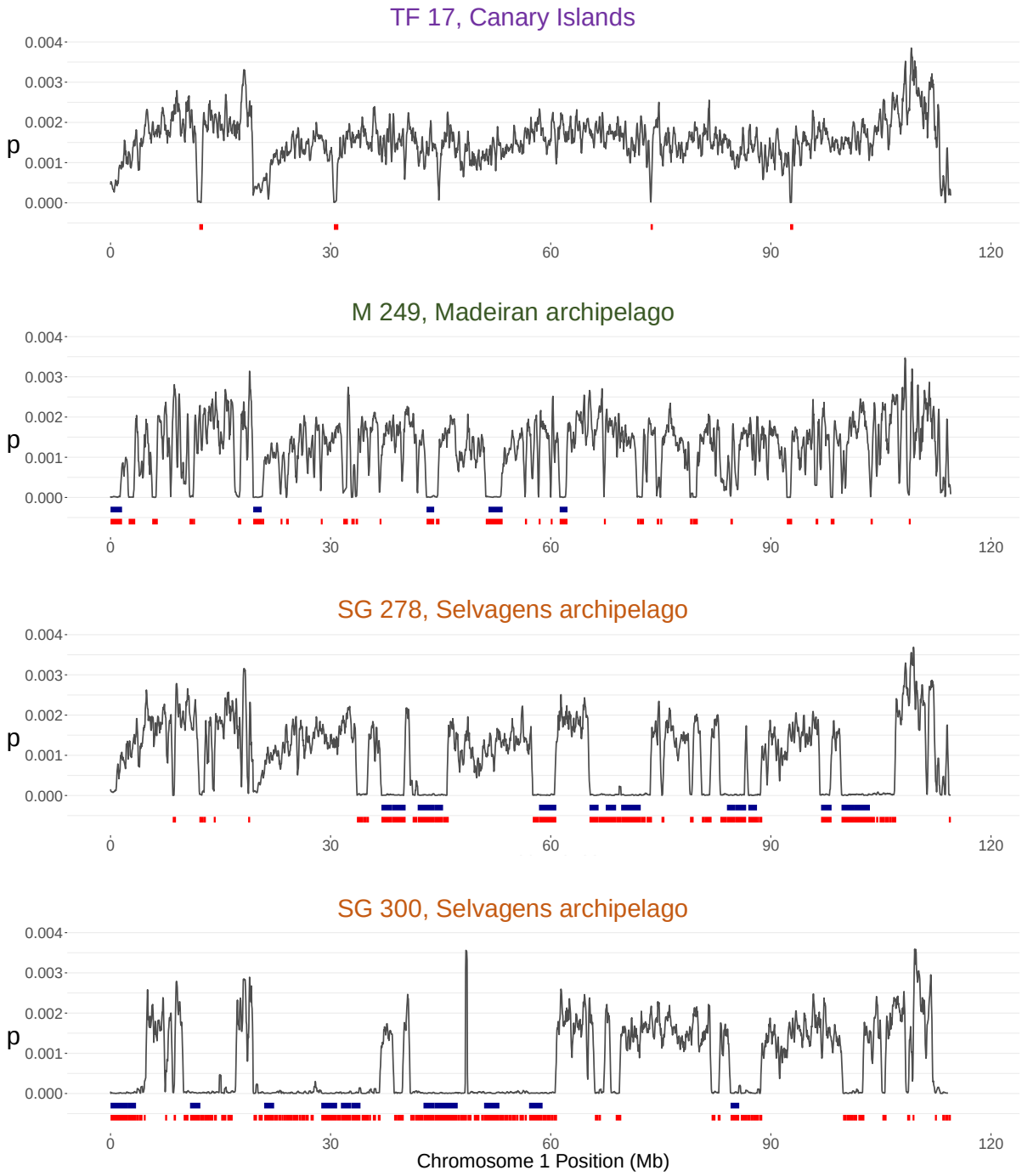


Figure 7. Nucleotide diversity across Chromosome 1 in four Berthelot's pipits. Nucleotide diversity was measured in 250 kb windows with 20% step across the entire chromosome. The position and length of runs of homozygosity (ROH) are plotted; blue blocks are ROH at least 1 Mb in length and red blocks indicate all ROH at least 250 kb in length.

5 Discussion

Using whole genome resequencing, we examined genetic diversity and demographic history through speciation and sequential island colonisation events, across three archipelagos of the Berthelot's pipit. We find loss of genetic diversity through colonisation events, with the most significant drop from the mainland to island populations, and identify genome-wide signatures of ROH as a result of ancient bottlenecks and contemporary inbreeding. Examining distribution of ROH and effective population sizes, N_e, using PSMC modelling revealed that: (i) sequential colonisation events are likely to be associated with strong founder effects resulting in ROH distributed across the genome; (ii) genomic signatures of inbreeding as a result of bottlenecks may persist over at least 20,000 generations, likely resulting from constrained population size, limited post-colonisation gene flow and high background relatedness within populations; and, iii) long ROH >1 Mb gradually degrade to form shorter more numerous ROH that correspond to relative times of colonisation bottlenecks across the archipelagos. Furthermore, we find evidence of initial colonisation to the Canary Islands dating ~2.1 Mya which closely supports previous estimates based on mitochondrial DNA (Voelker 1999), and confirm distinct secondary colonisation events to the Madeiran and Selvagens archipelagos (Spurgin *et al.* 2014). Our findings, using whole genome data, suggest an earlier colonisation and bottleneck to Madeira dated 50 kya, while Selvagens was likely colonised in the more recent past.

Our comparisons of genome-wide diversity and structure in Berthelot's and tawny pipits support previous evidence of colonisation and bottleneck history of Berthelot's pipit from reduced representation RAD-sequencing (Armstrong *et al.* 2018; Martin *et al.* 2021), as well as microsatellites and mitochondrial DNA (Illera *et al.* 2007; Spurgin *et al.* 2014). Using whole genomes, measures of genetic diversity (heterozygosity, nucleotide diversity, F_{IS} and F_{ROH}) show the most dramatic reduction between the tawny pipit and Berthelot's pipit, and are lowest in the populations that have experienced sequential archipelago level colonisations and associated population bottlenecks (Table 2). Overall, genetic diversity is at the lower end of what has been reported among other vertebrates (Yu *et al.* 2004; Dutoit *et al.* 2017; Kardos, Qvarnström, *et al.* 2017). We find relatively weak signatures of inbreeding (few short ROH and $F_{IS} < 0.06$) across the Canary Islands - the first archipelago the Berthelot's pipit colonised - compared to the Madeiran and Selvagens archipelagos (Fig. 5). That said, levels of inbreeding across the Canary Islands were comparable to avian island populations which have experienced strong population bottlenecks (see Jamieson *et al.* 2007, founder population of 33 North Island Robins (*Petroica longipes*); Lawson *et al.* 2017, mangrove finch (*Camarhynchus heliobates*) long-term bottleneck and small *Ne*; Swinnerton

et al. 2004, historic bottleneck of 20 pink pigeon (*Columba mayeri*)). Using PCA and pairwise F_{ST} measures, we were able to further describe population structure: The mainland tawny pipit diverged strongly from the Berthelot's pipit, with moderate divergence between the three Berthelot's pipit archipelagos, especially Madeira and the Selvagens (see Fig. 2 and Table 3).

Previous studies have estimated the N_e of the Berthelot's pipit across the colonisation range using approximate Bayesian computation (ABC) modelling (Spurgin *et al.* 2014) and divergence from the tawny pipit has been dated using mitochondrial cytochrome *b* evolution (Voelker 1999). Using PSMC modelling, we estimate Berthelot's pipit diverged from the mainland tawny pipit ~2.1 Mya, N_e was small, ~ 25,000 (Fig. 3). These species have since become distinct, with no genomic signatures of shared ancestry since divergence and substantially lower N_e across island populations. Total effective population size of the Berthelot's pipit steadily increased from one Mya until 150 kya across the range, likely reflecting an expansion of their habitable range as volcanic activity and climate across the Canary Islands stabilised and further islands formed (see Fig. 1). Recent population estimates suggest further ancestral splits between Berthelot's pipit populations within the last 50 thousand years, which may point to earlier divergence across the three archipelagos than previously estimated, with small population sizes across the range estimated 10 thousand years ago (N_e 33,000 in Selvagens, 17,000 across the Canary Islands and <2,000 in the Madeiran archipelago) (Fig. 4). Colonisation of the Selvagens is likely to have occurred more recently than the last 10 thousand years, since population declines (similar to that seen in the Madeiran archipelago) are only captured in the model using longer generation times (see Fig. S3). Reasons for dramatic population contraction across the range are unknown but may include pathogenic pressures, fitness effects of low diversity (i.e. inbreeding depression) or volcanic and climatic disturbances. It is important to note that PSMC did not provide information on recent N_e (present - 10 kya) and is sensitive to ancestral population structure and admixture, so it is likely that we need to interpret estimates within the last 50,000 years with caution (Li and Durbin 2011). Differing estimates of generation time and mutation rate also affect the PSMC interpretations. However, these estimates have a predictable effect on the PSMC curves: they do not change the relative shape of the curves but instead only move the curve along the axes (Nadachowska-Brzyska et al. 2015). For example, a halved mutation rate per year will move the curve to older times and also double the estimate of N_e (given a fixed generation time) and doubling the generation time will half the estimate of N_e.

We next aimed to study patterns of genetic diversity across individual genomes, to reveal signatures of demographic history. Despite differences in genome-wide levels of nucleotide diversity, peaks and troughs of diversity were generally consistent between individuals both within the same population and across the Berthelot's pipit populations and the tawny pipit (Fig. 5B and S1), which has been reported in other avian studies (Dutoit *et al.* 2017). However, this is not the case for the location of ROH, for which prevalence, but not genomic location, correlates within populations (Fig. S1). It is very likely that these regions represent true inbreeding instead of being consequences of shared chromosomal features (e.g., centromeres) as ROH in these regions are absent within the genomes of outbred pipits, for example across the Canary Islands (Fig. 5B; Fig. S1). The location of ROH varies strongly even between individuals within the same populations (Fig. 5B and 7), suggesting that these signatures are not solely a result of strong selection within particular islands, but due to recent ancestry of chromosomal segments as a result of inbreeding. We detected ROH across the genome of all individuals and generally find, (i) few short ROH across the tawny pipit and large Canary Island populations, (ii) with increased proportion of the genome in ROH (F_{ROH}) across Madeira and the Selvagens; and, (iii) longer ROH in the Selvagens relative to all other islands. This suggests an ancient bottleneck across Madeira with moderate contemporary background level inbreeding and a more recent severe bottleneck in the Selvagens, with an absence of post colonisation gene flow.

Signatures of inbreeding vary between species and with population demography. Where population contraction is very rapid it is possible there may be no signs of inbreeding such as ROH (Gelabert *et al.* 2020), while many severely inbred species have many short ROH and few covering vast chromosomal regions, such as ~17 Mb ROH in the California condor (*Gymnogyps californianus*; Robinson *et al.* 2021) and ~95 Mb ROH in a highly inbred population of Grey wolves (*Canis lupus*; Kardos, Åkesson, *et al.* 2017). In the Selvagens Berthelot's pipit population, we detect several ROH ~6 Mb despite bottlenecks dating several thousand years ago (see for example Fig. 7). While we do not detect vast autozygous regions as reported by some other studies of recent and severe inbreeding, comparisons across the Berthelot's pipit colonisation range clearly show the longest ROH in the isolated and bottlenecked population in the Selvagens.

We detect variance in inbreeding between individuals based on individual level observed heterozygosity and inbreeding estimates from ROH, which is particularly clear within the Selvagens archipelago. Such variance is common in populations where there are just a handful of family groups remaining and population-wide genetic diversity is very low

(Jamieson *et al.* 2007), as we expect in the Berthelot's pipit population in the Selvagens. Individual inbreeding is expected to fluctuate over short timescales in such populations, depending on the level of close relative mating. In a wild population it is likely that a high population average inbreeding reflects high background relatedness in the population as a result of founder effects or historic inbreeding, with individuals with exceptionally high inbreeding as a result of close parental ancestry (Brzeski *et al.* 2014).

High levels of inbreeding may result in inbreeding depression, which has been shown to be associated with phenotypic variation, survival and reproductive success in many natural populations (Richardson *et al.* 2004; Jamieson *et al.* 2007; Brzeski *et al.* 2014; Sin *et al.* 2021). Extreme and prolonged bottlenecks are thought to result in the purging of deleterious alleles (Stoffel *et al.* 2020), but this is not always the case. Arguably the most inbred wild species of bird, the Chatham Island black robin (*Petroica traversi*), was reduced to only one breeding pair from which all surviving individuals descended. Despite this extreme genetic bottleneck, severe inbreeding depression (measured as reduced juvenile survival) persists with no evidence of purging (Kennedy *et al.* 2014). In some situations, extreme genetic bottlenecks may instead result in the fixation of deleterious alleles (see van Oosterhout 2020). We cannot link inbreeding directly to fitness in Berthelot's pipit populations as we have not monitored individuals throughout their lives. However, it is likely that high levels of inbreeding in the Madeiran and Selvagens archipelagos have led to inbreeding depression (Szpiech *et al.* 2013), while genome-wide reductions in genetic diversity will impact the populations' adaptive potential. Both these effects may threaten the long-term viability of these populations, especially if they are exposed to new selection pressures, such as introduced infectious diseases (cf. Hawaiian avifauna, Jarvi *et al.* 2001). Further research is required to understand loss of or altered gene function in regions with exceptionally low diversity to uncover potential traits where variation has been lost.

When considered with the PSMC results, ROH findings suggest that dispersal to the Madeiran archipelago occurred earlier than previously estimated, approximately 50 kya, with a strong genetic bottleneck which has resulted in high levels of population level contemporary inbreeding. Meanwhile population structure between islands formed, potentially indicating geographic isolation and lack of gene flow, within the Canary Islands prior to dispersal to the Selvagens which is likely to have occurred within the last 10,000 years. It appears that signatures of founder effects persist for many generations in Berthelot's pipit, which adds to evidence of single colonisation events to each of the archipelagos and absence of genetic rescue from recent gene flow. These inferences made

using whole genomes are similar to those from previous studies using reduced representation methods but provide greater detail to the mechanisms and timings of divergence across the Berthelot's pipit colonisation range.

6 Conclusions

Genomic tools can be used to study contemporary and historic population demography, providing an opportunity to understand how genetic diversity is shaped across populations. We assessed patterns of genetic diversity between the Berthelot's pipit and tawny pipit, and across the Berthelot's pipit contemporary range, revealing that mainland to island colonisation, and sequential founder events, result in cumulative reductions in genetic diversity and inbreeding across the Berthelot's pipit island range. It is likely that post-colonisation population expansion across the Madeiran archipelago has resulted in genetic recovery which can be observed via many short ROH segments, while the Selvagens has experienced a more recent bottleneck and high background inbreeding, with ROH covering as much as 38% of the autosomal genome. Understanding the evolutionary processes behind loss of genetic diversity across small and isolated populations will be of additional importance for conservation efforts as future climatic and habitat shifts alter natural population ranges. Future studies investigating genetic diversity across and within populations should consider using whole genome sequencing as a powerful way to determine how past and present population history shape contemporary genetic diversity and its role in speciation.

7 References

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.

Ballou, J. (1983). Calculating inbreeding coefficients from pedigrees. *Genetics and Conservation*:509–516.

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

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.

Boyko, A.R., Quignon, P., Li, L., Schoenebeck, J.J., Degenhardt, J.D., Lohmueller, K.E., Zhao, K., *et al.* (2010). A simple genetic architecture underlies morphological variation in dogs. *PLoS Biology* **8**:49–50.

Brzeski, K.E., Rabon, D.R., Chamberlain, M.J., Waits, L.P. and Taylor, S.S. (2014). Inbreeding and inbreeding depression in endangered red wolves (*Canis rufus*). *Molecular Ecology* **23**:4241–4255.

Carson, H. (1971). Genetics, Speciation and the Founder Principle. *Stadler Symposia* **38**:279–279.

Chang, Chow, Tellier, Vattikuti, Purcell and Lee (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**:1–16.

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

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

Comeault, A.A., Flaxman, S.M., Riesch, R., Curran, E., Soria-Carrasco, V., Gompert, Z., Farkas, T.E., *et al.* (2015). Selection on a genetic polymorphism counteracts ecological speciation in a stick insect. *Current Biology* **25**:1975–1981.

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., *et al.* (2021). Twelve years of SAMtools and BCFtools. *GigaScience* **10**:1–4.

Dray, S. and Dufour, A.B. (2007). The ade4 package: Implementing the duality diagram for ecologists. *Journal of Statistical Software* **22**:1–20.

Dutoit, L., Burri, R., Nater, A., Mugal, C.F. and Ellegren, H. (2017). Genomic distribution and estimation of nucleotide diversity in natural populations : perspectives from the collared flycatcher (*Ficedula albicollis*) genome. *Molecular Ecology Resources* **17**:586–597.

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

Excoffier, L., Marchi, N., Marques, D.A., Matthey-Doret, R., Gouy, A. and Sousa, V.C. (2021). Fastsimcoal2 : Demographic Inference Under Complex Evolutionary Scenarios. *Bioinformatics*:1–4.

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. and Cresswell, W. (2007). The breeding biology of the endemic Berthelot's Pipit *Anthus berthelotii* in a harsh oceanic island environment (Tenerife, Canary Islands). *Ostrich - Journal of African Ornithology* **78**.

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

Gelabert, P., Sandoval-Velasco, M., Serres, A., de Manuel, M., Renom, P., Margaryan, A., Stiller, J., *et al.* (2020). Evolutionary history, genomic adaptation to toxic diet, and extinction of the Carolina Parakeet. *Current Biology* **30**:108–114.e5.

Gibson, J., Morton, N.E. and Collins, A. (2006). Extended tracts of homozygosity in outbred human populations. *Human Molecular Genetics* **15**:789–795.

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 gray wolf population. *Molecular Ecology*:3599–3612.

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

Gower, G., Tuke, J., Rohrlach, A.B., Soubrier, J., Llamas, B., Bean, N. and Cooper, A. (2018). Population size history from short genomic scaffolds: How short is too short? *bioRxiv*:1–11.

Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., di Palma, F. and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**:1145–1151.

Halkka, O., Raatikainen, M. and Halkka, L. (1974). The founder principle, founder selection, and evolutionary divergence and convergence in natural populations of Philaenus. *Hereditas* **78**:73–84.

Han, E., Sinsheimer, J.S. and Novembre, J. (2014). Characterizing bias in population genetic inferences from low-coverage sequencing data. *Molecular Biology and Evolution* **31**:723–735.

Hoernle, K. and Carracedo, J.-C. (2020). Canary Islands, Geology. *Encyclopedia of Islands*:133–143.

Hooper, R., Excoffier, L., Forney, K.A., Gilbert, M.T.P., Michael, D., Morin, P.A., Wolf, J.B.W., *et al.* (2020). Runs of homozygosity in killer whale genomes provide a global record of demographic histories. *bioRxiv*:1–22.

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

Izuno, A., Hatakeyama, M., Nishiyama, T., Tamaki, I., Shimizu-Inatsugi, R., Sasaki, R., Shimizu, K.K., *et al.* (2016). Genome sequencing of *Metrosideros polymorpha* (*Myrtaceae*), a dominant species in various habitats in the Hawaiian Islands with remarkable phenotypic variations. *Journal of Plant Research* **129**:727–736.

de Jager, D., Glanzmann, B., Möller, M., Hoal, E., van Helden, P., Harper, C. and Bloomer, P. (2021). High diversity, inbreeding and a dynamic Pleistocene demographic history revealed by African buffalo genomes. *Scientific Reports* **11**:1–15.

Jamieson, I.G., Tracy, L.N., Fletcher, D. and Armstrong, D.P. (2007). Moderate inbreeding depression in a reintroduced population of North Island robins. *Animal Conservation* **10**:95–102.

Jarvi, S.I., Atkinson, C.T. and Fleischer, R.C. (2001). Immunogenetics and resistance to avian malaria in Hawaiian Honeycreepers (*Drepanidinae*). *Studies in Avian Biology No* **22**:254–263.

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., *et al.* (2017). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution* **2**:124–131.

Kardos, M., Luikart, G. and Allendorf, F.W. (2015). Measuring individual inbreeding in the age of genomics: Marker-based measures are better than pedigrees. *Heredity* **115**:63–72.

Kardos, M., Qvarnström, A. and Ellegren, H. (2017). Inferring individual inbreeding and demographic history from segments of identity by descent in *Ficedula* flycatcher genome sequences. *Genetics* **205**:1319–1334.

Kennedy, E.S., Grueber, C.E., Duncan, R.P. and Jamieson, I.G. (2014). Severe inbreeding depression and no evidence of purging in an extremely inbred wild species-the chatham island black robin. *Evolution* **68**:987–995.

Kirch, M., Romundset, A., Gilbert, M.T.P., Jones, F.C. and Foote, A.D. (2021). Ancient and modern stickleback genomes reveal the demographic constraints on adaptation. *Current Biology* **31**:2027–2036.e8.

Korf, B. (2013). Runs of homozygosity and parental relatedness. *Genetics in Medicine* **15**:752–753.

Lawson, L.P., Fessl, B., Hernán Vargas, F., Farrington, H.L., Francesca Cunninghame, H., Mueller, J.C., Nemeth, E., *et al.* (2017). Slow motion extinction: inbreeding, introgression, and loss in the critically endangered mangrove finch (*Camarhynchus heliobates*). *Conservation Genetics* **18**:159–170.

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* **0**:1–8.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **0**:1–3.

Li, H. and Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*:8–12.

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

Ma, T., Wang, K., Hu, Q., Xi, Z., Wan, D., Wang, Q., Feng, J., *et al.* (2017). Ancient polymorphisms and divergence hitchhiking contribute to genomic islands of divergence within a poplar species complex. *Proceedings of the National Academy of Sciences of the United States of America* **115**:E236–E243.

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

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., *et al.* (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**:1817–1828.

Mather, N., Traves, S.M. and Ho, S.Y.W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution* **10**:579–589.

Maynard-Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**:23–35.

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. *Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning* **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. *American Journal of Human Genetics* **83**:359–372.

Meyermans, R., Gorssen, W., Buys, N. and Janssens, S. (2020). How to study runs of homozygosity using plink? a guide for analyzing medium density snp data in livestock and pet species. *BMC Genomics* **21**:1–14.

Nadachowska-Brzyska, K., Burri, R., Smeds, L. and Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology* **25**:1058–1072.

Nadachowska-Brzyska, K., Li, C., Smeds, L., Zhang, G. and Ellegren, H. (2015). Temporal dynamics of avian populations during pleistocene revealed by whole-genome sequences. *Current Biology* **25**:1375–1380.

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

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

van Oosterhout, C. (2020). Mutation load is the spectre of species conservation. *Nature Ecology and Evolution* **4**:1004–1006.

Oostermeijer, J.G.B., Van Eijck, M.W., Van Leeuwen, N.C. and Den Nijs, J.C.M. (1995). Analysis of the relationship between allozyme heterozygosity and fitness in the rare *Gentiana pneumonanthe L. Journal of Evolutionary Biology* **8**:739–759.

Patil, A.B., Shinde, S.S., Raghavendra, S., Satish, B.N., Kushalappa, C.G. and Vijay, N. (2021). The genome sequence of *Mesua ferrea* and comparative demographic histories of forest trees. *Gene* **769**:145214.

Patton, A.H., Margres, M.J., Stahlke, A.R., Hendricks, S., Lewallen, K., Hamede, R.K., Ruiz-Aravena, M., *et al.* (2019). Contemporary demographic reconstruction methods are robust to genome assembly quality: A case study in Tasmanian Devils. *Molecular Biology and Evolution* **36**:2906–2921.

Paulay, G. (1994). Biodiversity on oceanic islands: Its origin and extinction. *Integrative and Comparative Biology* **34**:134–144.

Pilot, M., Greco, C., Vonholdt, B.M., Jędrzejewska, B., Randi, E., Jędrzejewski, W., Sidorovich, V.E., *et al.* (2014). Genome-wide signatures of population bottlenecks and diversifying selection in European wolves. *Heredity* **112**:428–442.

Polašek, O., Hayward, C., Bellenguez, C., Vitart, V., Kolčić, I., McQuillan, R., Saftić, V., *et al.* (2010). Comparative assessment of methods for estimating individual genome-wide homozygosity-by-descent from human genomic data. *BMC Genomics* **11**.

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

Robinson, J.A., Bowie, R.C.K., Dudchenko, O., Aiden, E.L., Hendrickson, S.L., Steiner, C.C., Ryder, O.A., *et al.* (2021). Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology* **31**:2939–2946.e5.

Robinson, J.A., Brown, C., Kim, B.Y., Lohmueller, K.E. and Wayne, R.K. (2018). Purging of strongly deleterious mutations explains long-term persistence and absence of inbreeding depression in island foxes. *Current Biology* **28**:3487–3494.e4.

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

Sin, S.Y.W., Hoover, B., Nevitt, G. and Edwards, S. V. (2021). Demographic history, not mating system, explains signatures of inbreeding and inbreeding depression in a large outbred population. *The American Naturalist* **197**.

Smeds, L., Qvarnström, A. and Ellegren, H. (2016). Direct estimate of the rate of germline mutation in a bird. *Genome Research* **26**:1211–1218.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. and 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. and 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.

Stoffel, M.A., Johnston, S.E., Pilkington, J.G. and Pemberton, J.M. (2020). Genetic architecture and lifetime dynamics of inbreeding depression in a wild mammal. *bioRxiv* **21**:1–9.

Swinnerton, K.J., Groombridge, J.J., Jones, C.G., Burn, R.W. and Mungroo, Y. (2004). Inbreeding depression and founder diversity among captive and free-living populations of the endangered pink pigeon Columba mayeri. *Animal Conservation* **7**:353–364.

Terhorst, J., Kamm, J.A. and Song, Y.S. (2017). Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics* **49**:303–309.

Voelker, G. (1999). Dispersal, Vicariance, and Clocks: Historical Biogeography and Speciation in a Cosmopolitan Passerine Genus (*Anthus: Motacillidae*). *Evolution* **53**:1536.

Warren, B.H., Bermingham, E., Bourgeois, Y., Estep, L.K., Prys-Jones, R.P., Strasberg, D. and Thébaud, C. (2012). Hybridization and barriers to gene flow in an island bird radiation. *Evolution* **66**:1490–1505.

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.

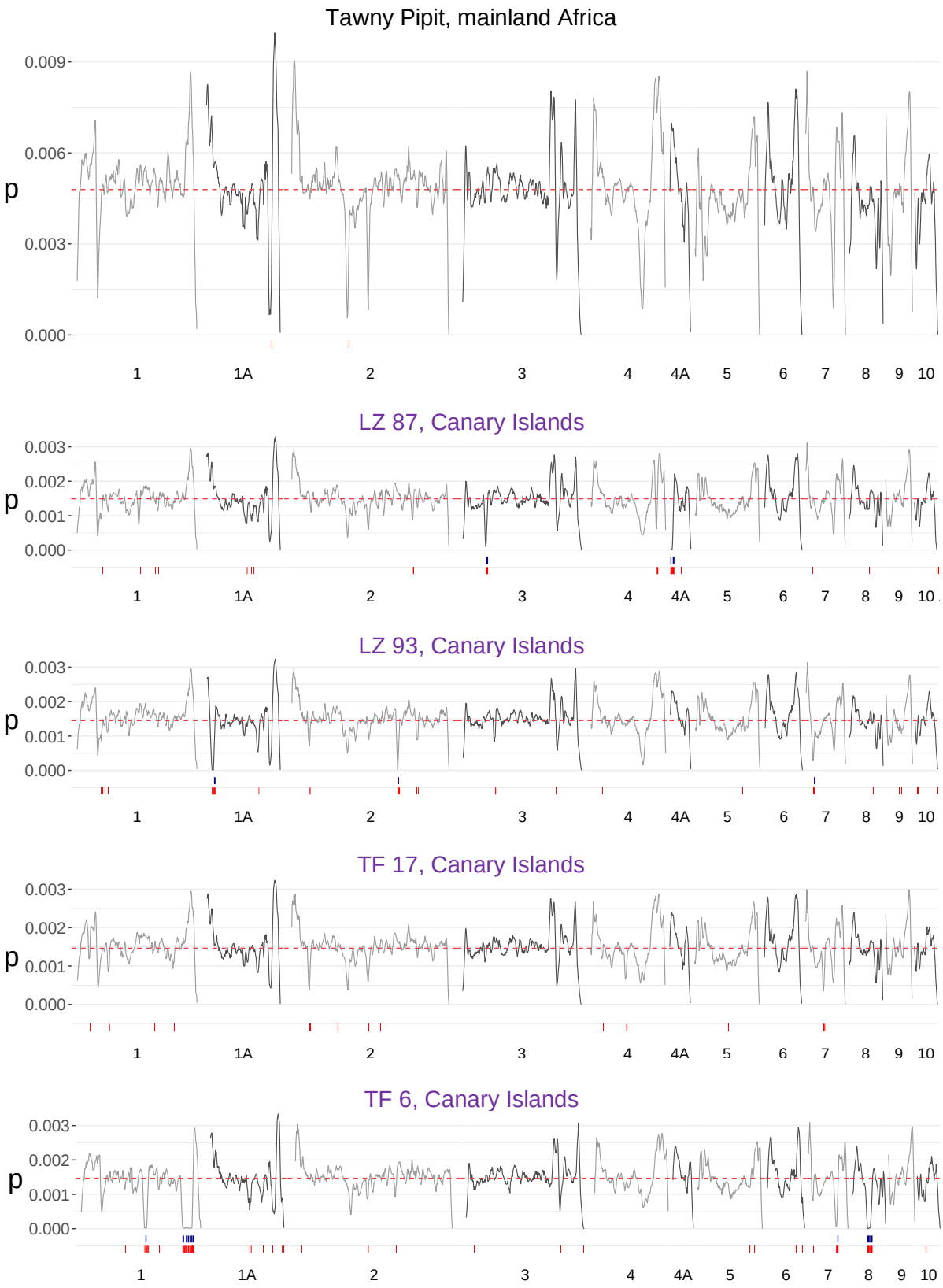
Weaver, S.C., Forrester, N.L., Liu, J. and Vasilakis, N. (2021). Population bottlenecks and founder effects: implications for mosquito-borne arboviral emergence. *Nature Reviews Microbiology* **19**:184–195.

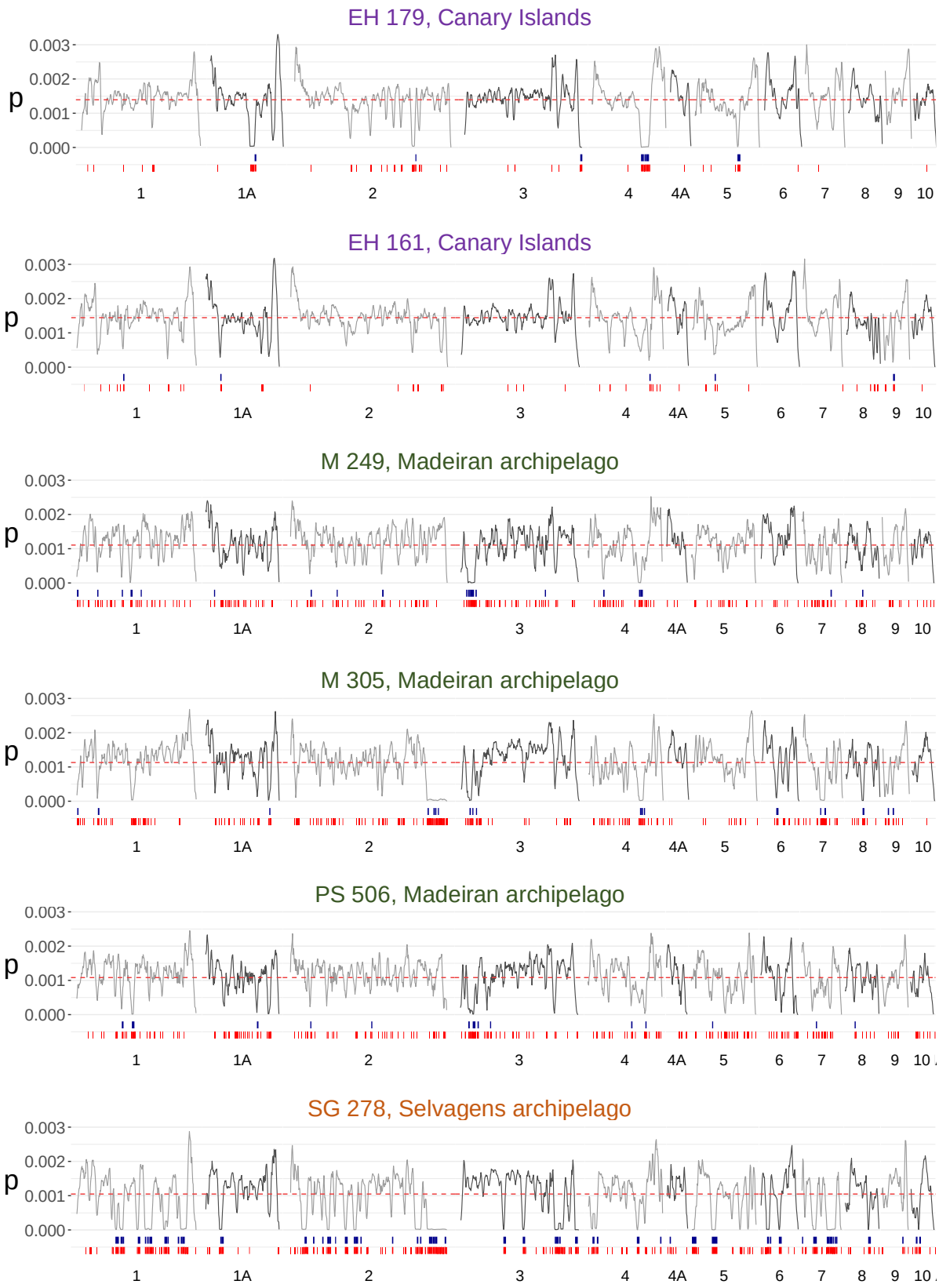
Willing, E.M., Dreyer, C. and van Oosterhout, C. (2012). Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. *PLoS ONE* **7**:1–7.

Xue, Y., Prado-martinez, J., Sudmant, P.H., Narasimhan, V., Ayub, Q., Szpak, M., Frandsen, P., *et al.* (2015). Mountain gorilla genomes reveal the impact of long-term population decline and inbreeding. *Science* **348**:242–245.

Yu, N., Jensen-Seaman, M.I., Chemnick, L., Ryder, O. and Li, W.H. (2004). Nucleotide diversity in gorillas. *Genetics* **166**:1375–1383.

8 Supplamentary material:




EH 179, Canary Islands
EH 161, Canary Islands
M 249, Madeiran archipelago
M 305, Madeiran archipelago
PS 506, Madeiran archipelago
SG 278, Selvagens archipelago
p
0.003
0.002
0.001
0.000
1
1A
2
3
4
4A
5
6
7
8
9
10

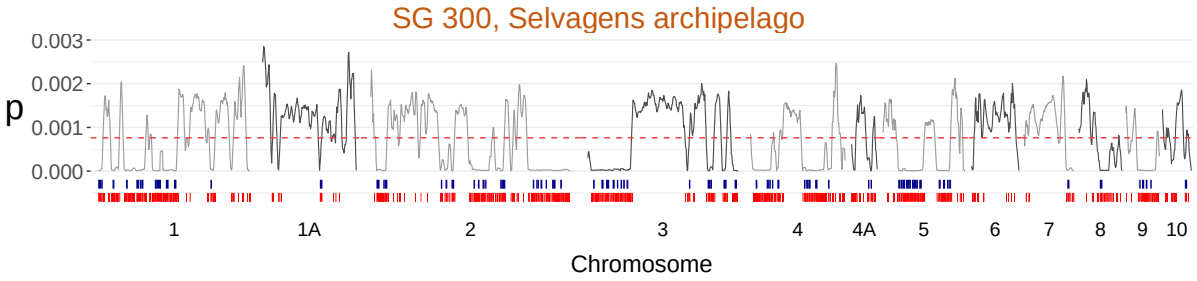


Figure S1. Patterns of genetic diversity across resequenced genomes of the tawny pipit and 11 Berthelot's pipit across the species colonisation range. Nucleotide diversity is plotted using 2 Mb windows with 20% overlap (grey lines), and the length and location of Runs Of Homozygosity (ROH) are plotted; blue blocks are ROH at least 1 Mb in length and red blocks indicate ROH at least 250 kb in length. Horizontal red dotted lines represent mean nucleotide diversity per sample calculated across the genome. Macrochromosomes 1A, 4A and 1-10 are presented for visual comparison between individuals.

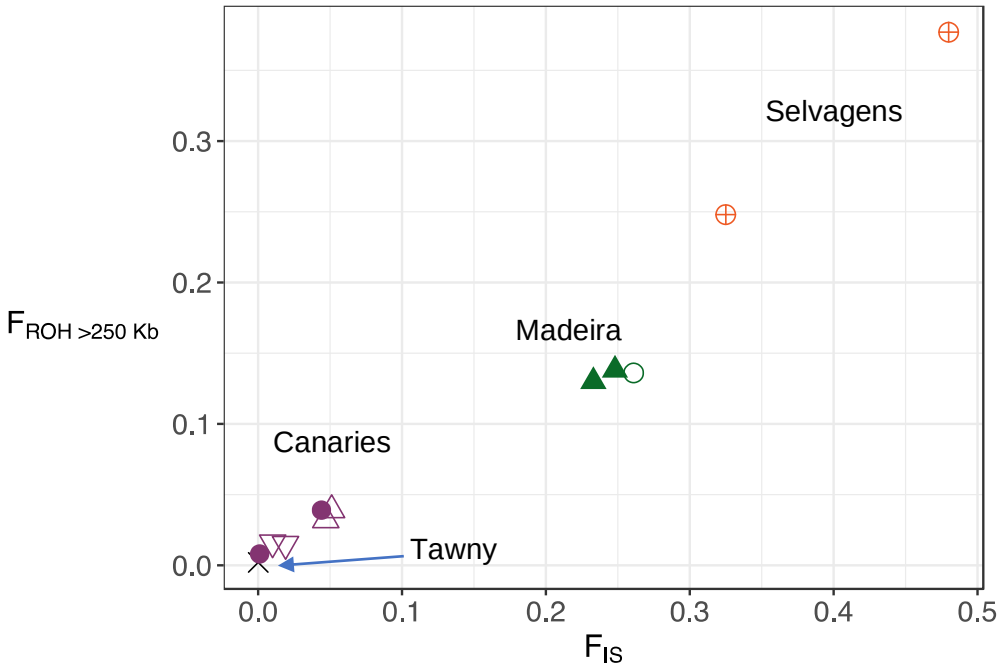


Figure S2. Inbreeding within tawny pipit and across Berthelot's pipit populations estimated using F_{IS} and $F_{ROH >250 kb}$. Populations are coloured by archipelago, with tawny pipit in black, Canaries populations in purple, Madeiran populations in green and Selvagens in orange. Shapes match those used in main text. Population symbols: tawny pipit = cross, El Hierro = upwards triangles, Tenerife = filled circles, Lanzarote = downward triangle, Madeira = filled triangles, Porto Santo = empty circle, Selvagens = hashed circles.

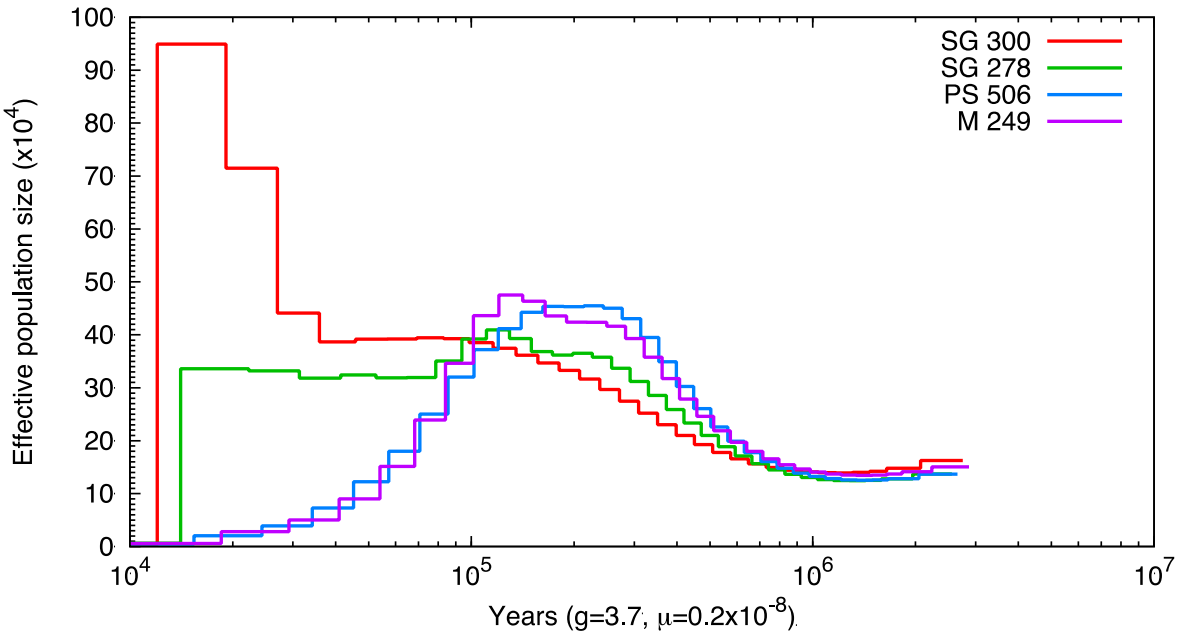


Figure S3. PSMC estimates of changes in effective population size, *Ne*, over time when the upper estimate of 3.7 years for generation time is used, for the contemporary Madeira and Selvagens archipelagos. Populations are modelled from 3 Mya until 10 kya.

4

Genomic landscapes of divergence in an island archipelago bird: identifying the role of drift and selection

Claudia A. Martin, Brent C. Emerson, Juan Carlos Illera, Alex Suh,
Lewis G. Spurgin and David S. Richardson

Berthelot's pipit. Photograph by Jean Charles De Coriolis

1 Abstract

When populations colonise new environments they may be exposed to novel selective pressures but also suffer from extensive genetic drift due to founder effects and limited subsequent gene flow. Genomic approaches provide an opportunity to empirically study how these factors drive divergence, enabling us to disentangle neutral effects from differentiation at specific loci due to selection. We investigate patterns of genetic diversity and divergence in Berthelot's pipit (*Anthus berthelotii*), a passerine bird endemic to the islands of three north Atlantic archipelagos. Strong environmental gradients, including pathogen exposure, among the different isolated populations across the species range, make it an excellent system in which to explore traits important in adaptation and/or incipient speciation. We analyse genomic differentiation, to identify genomic islands of divergence, among Berthelot's pipit populations, and between Berthelot's pipit and its mainland sister species, the tawny pipit (*Anthus campestris*). We quantify how genomic divergence accumulates across the speciation continuum and identify highly differentiated loci between population comparisons spanning different divergence timeframes (2.1 Mya – ca. 8 kya). Characteristic signatures of selection between Berthelot's and tawny pipits, and among Berthelot's pipit archipelagos, are identified in loci associated with craniofacial/bone and eye development, metabolism and immune response. We find limited evidence for repeated divergence within the same loci across the colonisation range but do identify different loci putatively associated with the same biological traits in different populations, likely due to parallel adaptation. Our findings provide evidence that loci involved in morphology, metabolism and immune defence may be involved in incipient speciation across these island bird populations.

Keywords: Genomic islands of divergence, genomic landscape, speciation, adaptation, craniofacial evolution, immune defence.

2 Introduction

Genetic differentiation among populations accumulates over time due to a combination of differing adaptive and stochastic processes (see reviews, Feder *et al.* 2013; Seehausen *et al.* 2014). The speed at which divergence occurs and the resulting genomic landscape characteristics depend on the strength of selective and neutral forces - including drift, mutation and gene flow - occurring within and among populations (Nosil *et al.* 2009). Across

the genome, loci under selection in individual or multiple populations are expected to diverge first, with neutrally evolving genomic regions typically differentiating more slowly (Walsh *et al.* 2019). Demographic history, such as previous bottlenecks or inbreeding, may also lead to radical changes in the genome, but the effects of these events are expected to occur more evenly throughout the genome (Kimura 1991; Nei 2005). Genomic approaches provide an opportunity to study the genetic landscape of divergence among populations, enabling the relative importance of differing evolutionary forces driving divergence to be determined (Stajich and Hahn 2005; Ravinet *et al.* 2017). However, to do this studies first need to understand demographic history and gene flow between populations, and characterise selection pressures and how these vary across populations (Ravinet *et al.* 2017; Nosil and Feder 2012).

Upon colonisation of new environments, populations may be exposed to different selective pressures, which may result in rapid ecological and phenotypic divergence between populations (e.g., Walsh *et al.* 2019). When a locally beneficial allele at a locus arises, positive selection may cause it to rapidly increase in frequency in one population resulting in a local selective sweep, while in the other population that allele may be lost through drift or purifying selection, or remain at low frequencies (Ravinet *et al.* 2017). During the initial stages of divergence, loci under strong selection are expected to be the first regions of the genome to become differentiated (Nosil *et al.* 2009). Furthermore, genetic differentiation is expected to be localised, with peaks of divergence around selected loci, often referred to as ‘genomic islands of divergence’ (Nadeau *et al.* 2012; Burri *et al.* 2015). Such areas can be responsible for the accumulation of genetic and phenotypic differences between populations, which may play a fundamental role in speciation (Via and West 2008; Ruegg *et al.* 2014). As well as genomic islands of divergence, highly conserved genomic regions, where differentiation is far below background levels have been identified in a range of species (Ravinet *et al.* 2017; Hofer *et al.* 2012; Van Doren *et al.* 2017; Sendell-Price *et al.* 2020). These ‘genomic valleys of divergence’ may occur because the same allele is favoured in both populations (parallel selection) (Nielsen 2005; Roesti *et al.* 2012) or linked neutral loci are favoured through background selection (Cvijović *et al.* 2018), leaving a distinct signature of reduced genetic diversity in both populations (Roesti *et al.* 2014). Importantly, reduced recombination and background selection can also result in the formation of genomic islands through linkage to selected loci (Noor and Bennett 2009). Studying highly conserved, as well as divergent, genomic regions may provide important evolutionary insight into ecological adaptation and speciation.

Linkage disequilibrium may facilitate divergence hitchhiking of neutral and weakly selected loci, resulting in broad genomic islands surrounding selected loci (Maynard-Smith and Haigh 1974; Nosil *et al.* 2009; Nosil and Feder 2012). Broad peaks of divergence often form as a result of recent selection, where there has been an absence of recombination around the selected loci. Over time recombination erases the effect of divergence hitchhiking, by reducing linkage between loci, which may result in sharp peaks surrounding selected loci (Nosil *et al.* 2009). However, recombination does not act evenly across the genome and population processes including limited gene flow and small effective population sizes (*Ne*) reduce the rate of recombination, which in turn maintains large regions of divergence or conservation over long timescales (Feder and Nosil 2010). Discerning and dating the combination of evolutionary processes that have shaped genetic diversity between populations is complex, and there is potential for the same pattern to evolve as a result of differing evolutionary scenarios. Therefore, studies need to combine observations of the divergence landscape with knowledge of past and present population processes to determine the relative roles of drift and selection in the divergence of a particular genomic region.

With the colonisation of a new environment from a source population, the new population may be strongly influenced by founder effects (Barton and Charlesworth 1984; Harrison 1991). Founder effects cause a loss of genetic diversity and result in different subsets of genetic variation existing in the source and newly founded populations (Barton and Charlesworth 1984; Berry 1986). The characteristics of divergence over time are then mediated by the level and timing of gene flow between diverging populations (Ravinet *et al.* 2017), although new species can emerge without complete geographical isolation (Kirkpatrick and Ravigné 2002; Li *et al.* 2010; Martin *et al.* 2013; Bay and Ruegg 2017) and across microgeographic (Milner *et al.* 1999; Langin *et al.* 2015) as well as broad spatial scales. Genetic drift may be particularly strong in small genetically isolated populations which may result in the rapid loss or fixation of genetic variants (Alleaume-Benharira *et al.* 2006). The impact of such processes may be exaggerated by sequential founder events and the cumulative effects of drift through these events (Clegg *et al.* 2002; Tomozawa *et al.* 2014).

Patterns of genomic divergence are typically measured using $F_{ST,}$ a measure of relative difference in allele frequencies between populations. Contrasting different measures of genetic diversity can be helpful for inferring differing modes of divergence between populations (Delmore *et al.* 2018; Irwin *et al.* 2018; Osmond and Coop 2020), and a range of

diversity statistics can be used (see review; Wolf and Ellegren 2017). Commonly used statistics include Tajima's D, to test for rare variants as a signal of directional or background selection or large-scale demographic effects; and nucleotide diversity (π), an estimate of genetic diversity, which is derived from the number of pairwise sequence differences among members of a population. Knowledge of the recombination landscape and patterns of linkage disequilibrium can be particularly helpful to estimate the timing of divergence signatures putatively under selection, but these data are rarely available for non-model systems. Combining measures of divergence with genetic diversity across populations with well-characterised history have the best ability to distinguish between divergence due to positive divergent or convergent selection, balancing selection, population bottlenecks or genetic architecture (Schneider *et al.* 2021).

Island populations of Berthelot's pipit (*Anthus berthelotii*) and its mainland sister species, the tawny pipit (*Anthus campestris*), provide an excellent opportunity to explore genomic patterns of divergence and speciation across divergence timescales and known colonisation events, with an absence of post-colonisation gene flow (Illera *et al.* 2007; Spurgin *et al.* 2014; Armstrong *et al.* 2018; Chapter 2). The ancestor of these two species colonised the Canary Islands from mainland Africa approximately 2.1 Mya and has since dispersed independently to both the Madeiran and Selvagens archipelagos approximately 50 kya and 8-10 kya, respectively (Chapter 3). Previous research suggests colonisation is associated with reduced genetic diversity through founder effects and an absence of post-colonisation gene flow across the Berthelot's pipit range (Spurgin *et al.* 2014; Gonzalez-Quevedo *et al.* 2015). This is confirmed by signatures of inbreeding (ROH >1 Mb) across the recently colonised archipelagos (Chapter 3). Across the species' range, strong genetic structuring exists between, but not within, Berthelot's pipit populations at the archipelago level (Armstrong *et al.* 2018; Chapter 2).

Importantly, across the Berthelot's pipit system there are considerable selection gradients, including gradients in climate, habitat and pathogen prevalence (reviewed by Illera *et al.* 2016). For example, different Berthelot's pipit populations have considerable, and temporally consistent, variation in the prevalence of avian pox and avian malaria (Illera *et al.* 2008; Spurgin *et al.* 2012), which has enabled previous studies of host-pathogen evolution (Gonzalez-Quevedo *et al.* 2014; Gonzalez-Quevedo *et al.* 2015; Sheppard *et al.* 2022). Such pathogens can exert strong selective pressures on avian populations (see for example, Liao *et al.* 2017). There are also significant morphological differences across the system, with reduced body and bill size in Berthelot's pipit compared to the tawny pipit, and

archipelago level variation in bill morphology and body size in Berthelot's pipit (Spurgin *et al.* 2014; Armstrong *et al.* 2018).

Reduced representation restriction-site associated DNA sequencing (RAD-seq) markers have previously been used to investigate divergence across the Berthelot's pipit range, both at broad (see Armstrong et al. 2018) and fine (Chapter 2) geographic scales. Stronger signatures of selection were identified between archipelagos, compared to between island populations within archipelagos (Armstrong *et al.* 2018; Chapter 2). That study identified genes associated with immunity, metabolism and bill length as being divergent. However, it is not clear to what extent patterns of diversity in these regions are shaped by drift and selection, and it is likely many divergent loci in the genome have gone undetected as a result of unsequenced genomic regions (Armstrong *et al.* 2018).

Here, we use whole genome resequencing to assess genomic landscapes of divergence through sequential archipelago colonisation by the Berthelot's pipit and its ancestor, the tawny pipit, to uncover loci of importance for divergence and adaption across timeframes. Our specific aims were: (1) to determine how divergence accumulates across the genome between Berthelot's pipits and its mainland relative the tawny pipit, as well as between more recently divergent archipelago level populations of Berthelot's pipit; (2) to identify genomic islands and valleys of divergence for each dyad of populations; and (3) to understand how drift and selection have interacted to shape variation across these genomic regions. Finally, to generate hypotheses about potential adaptive phenotypes, we identified candidate genes in regions under selection.

3 Methods

3.1 Sample collection and reference genome sequencing

Blood samples were collected from six Berthelot's pipit island populations (Lanzarote, Tenerife, El Hierro, Selvagem Grande, Porto Santo and Madeira) across the three archipelagos of its range, and one tawny pipit (*Anthus campestris*) sampled from coastal Mauritania (Illera *et al.* 2007; Spurgin *et al.* 2014) (Fig.1). This allowed the capture of a range of divergence levels and colonisation timeframes within Berthelot's pipit and its sister species (Table 2). Full sampling details, DNA preparation and the draft Berthelot's pipit reference protocols are provided in Chapter 3.

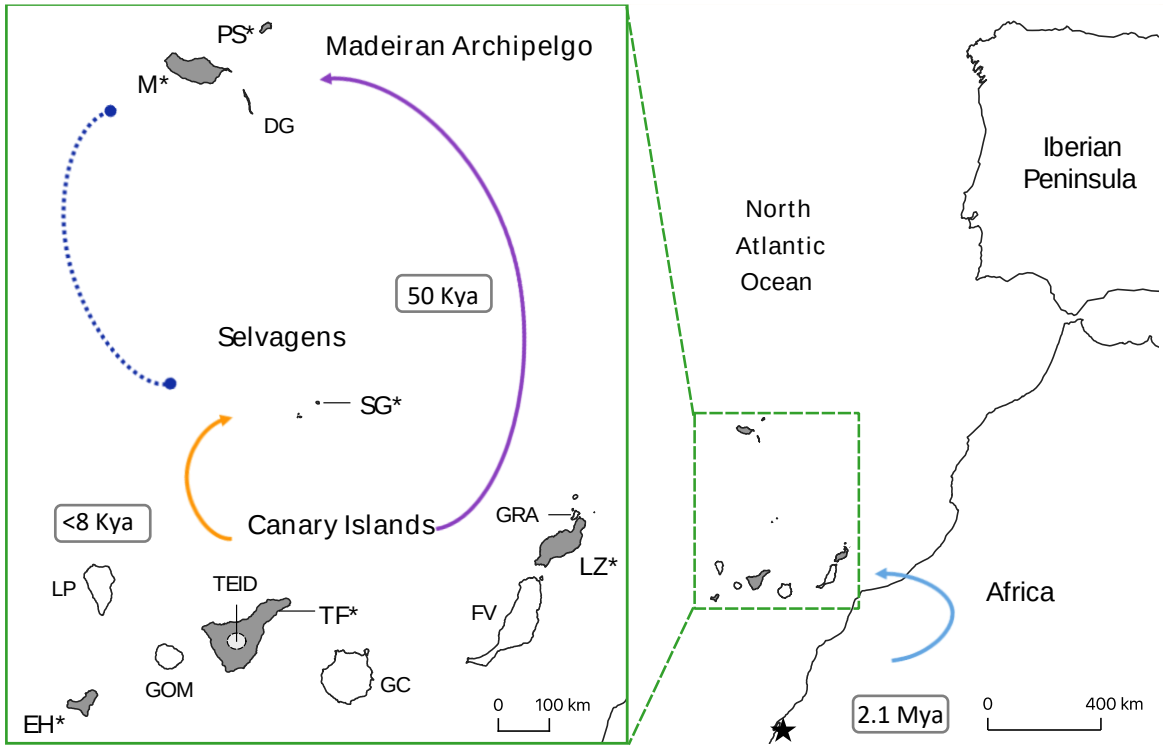


Figure 1. Berthelot's pipit range across three Macaronesian archipelagos and the sampling location for the tawny pipit. Sample locations used for whole genome resequencing are denoted with a * and the island shaded grey. The tawny pipit was sampled on migration in Mauritania (Latitude: 17.991703°, Longitude: -16.016672°, see black star). Population comparisons are indicated by coloured arrows, with the timing and direction of colonisation events highlighted, while the blue dotted line indicates the comparison made between populations after two independent colonisation events from the Canary Islands to the Selvagens and to the Madeiran archipelago. Canary Island populations: El Hierro (EH*), La Palma (LP), La Gomera (GOM), Teide (TEID) mountain population on Tenerife, Tenerife (TF*), Gran Canaria (GC), Fuerteventura (FV), Lanzarote (LZ*), La Graciosa (GRA). Madeiran populations: Madeira (M*), Porto Santo (PS*) and Deserta Grande (DG). Selvagens: Selvagem Grande (SG*).

3.2 Genome resequencing, read alignment and variant calling

We selected Berthelot's pipit samples to maximise the geographical range across archipelagos (Fig. 2A) with two individuals (one male one female) per population from El Hierro, Tenerife and Lanzarote (Canary Islands), Madeira and Porto Santo (one sample in addition to the reference sample) in the Madeiran archipelago and Selvagem Grande (Selvagens archipelago). All individuals chosen were adult birds with no pox lesions and no identified *hemoprotozoa* parasites using a nested PCR approach (Waldenstrom *et al.* 2014).

One tawny pipit sample was also included. Low Input, Transposase Enabled (LITE) Illumina compatible libraries were constructed and high throughput libraries generated for each sample at the Earlham Institute as reported in Chapter 3.

We use two VCF files generated in Chapter 3: the “All Pipits” dataset had variants joint called across the 11 Berthelot’s pipit samples and the one tawny pipit to assess divergence between these species. The second “Berthelot’s” dataset was joint called across Berthelot’s pipit samples exclusively, meaning variants unique to the tawny pipit were not included.

3.3 Variant mapping and filtering

Variants were mapped to chromosomes of the Zebra finch (*Taeniopygia guttata*) genome assembly bTaeGut1_v1.p (NCBI Assembly GCA_003957595.1) using SatsumaSynteny (Grabherr *et al.* 2010); full details in Chapter 3. When filtering the datasets we retain the Z chromosome, but otherwise filtering was conducted as in Chapter 3. As in Chapter 3 we removed all sites at which mean read depth (among all individuals; 12 for “All Pipits”, 11 for “Berthelot’s”) was less than 10 or more than twice the average read depth across the genome (--min-meanDP 10, max-meanDP 48/44). This resulted in two mapped and quality trimmed VCF files, “All Pipits” and “Berthelot’s”. Individual level data for read depth, individual missingness and number of variants from both datasets is summarised in Table 1.

Table 1. Genomic filtering for divergence analyses across populations of Berthelot's pipit and the tawny pipit. Sampling information and divergence comparisons for the "All Pipits" (top line for each individual) and "Berthelot's" (second line) dataset. Mean variant coverage and percentage missing variants per individual were calculated using the quality filtered SNPs for both datasets. Population codes: Canary Islands - LZ = Lanzarote, TF = Tenerife lowland, EH = El Hierro. Madeiran archipelago – M = Madeira island, PS = Porto Santo. Selvagens archipelago – Selvagem Grande = SG.

Archipelago / Location	Pop. code	Sex (Ind. ID)	Mean read coverage (X)	Mean variant coverage (X)	Individual missing variants (%)	*n* raw loci	*n* filtered loci
Mainland Africa							
Mauritania	TAW	M (462)	27.3	27.3	4.0	18,920,000	11,788,031
Canary Islands							
Lanzarote	LZ	M (87)	30.1	30.1 30.1	1.7 2.3	18,923,096 10,361,030	11,788,198 5,934,918
		F (93)	25.2	26.4 26.2	3.1 4.7	18,922,012 10,360,120	11,788,132 5,934,872
Tenerife	TF	M (17)	25.0	26.0 25.9	2.0 2.9	18,922,546 10,360,653	11,788,190 5,934,918
		F (6)	23.4	24.7 24.5	3.0 4.5	18,921,656 10,359,854	11,788,148 5,934,882
El Hierro	EH	M (179)	23.0	23.9 23.7	2.1 3.0	18,922,354 10,360,520	11,788,162 5,934,910
		F (161)	21.9	22.7 22.6	2.1 3.0	18,921,866 10,359,970	11,788,147 5,934,889
Madeiran							
Madeira	M	M (249)	24.3	25.2 25.0	1.6 2.3	18,923,699 10,361,593	11,788,210 5,934,921
		F (305)	22.0	22.8 22.8	2.0 3.0	18,923,105 10,361,072	11,788,189 5,934,921
Porto Santo	PS	F (506)	20.2	21.3 21.1	3.5 5.0	18,922,905 10,360,994	11,788,150 5,934,872
Selvagens							
Selvagem Grande	SG	M (278)	20.9	22.2 22.0	2.8 4.1	18,921,661 10,360,007	11,788,133 5,934,879
		F (300)	23.1	23.7 23.6	2.2 3.1	18,921,824 10,360,060	11,788,170 5,934,910

3.4 Population comparisons

We compared pairwise divergence between the tawny pipit and Berthelot's pipit and then among three dyads of Berthelot's pipit archipelago populations (Table 2). The population

dyads chosen for comparison varied in their temporal, spatial and morphological divergence, and in terms of selective pressures (Table S1) as outlined below:

Tawny pipit vs. Canary Islands: Berthelot's pipit initially colonised the Canary Islands approximately 2.1 Mya (Chapter 3; Voelker 1999) with founder effects resulting in a genome-wide reduction in genetic diversity compared to the tawny pipit. There is no evidence of subsequent gene flow. The tawny pipit is a palearctic migrant, that winters across Northern Africa while Berthelot's pipit is an island resident. Habitat types and disease prevalence varies substantially across the Berthelot's pipit Canary Islands range and across the tawny range. Both species are exposed to avian malaria, avian pox and other pathogens (tawny pipit; Calero-Riestra and Garcia 2016, Berthelot's pipit; Illera *et al.* 2008).

Canary Islands vs. Madeiran archipelago: Colonisation of the Madeiran archipelago from the Canary Islands is estimated at approximately 50 kya (Table 2, Figure 1; Chapter 3), and resulted in a strong population bottleneck. Berthelot's pipits from the Madeiran archipelago are classified as a separate subspecies, *A. berthelotii madeirensis*, based on longer bill lengths (Martın and Lorenzo 2001; Oliveira and Menezes 2004) and larger body size (Spurgin *et al.* 2014). Considerable variation in disease prevalence occurs within and among these islands archipelagos (Illera *et al.* 2008).

Canary Islands vs. Selvagens archipelago: The Selvagens was populated by Berthelot's pipit through independent colonisation from the Canary Islands approximately 8-10 kya (Table 2, Figure 1; Chapter 3). Small island size, geographic isolation and strong founder effects have resulted in low genetic diversity in the Selvagens' population and strong signatures of inbreeding (Spurgin *et al.* 2014; Chapter 3). While there is considerable variation in disease prevalence across the Canary Islands, no disease has been detected on the Selvagens (Illera *et al.* 2008).

Madeiran archipelago vs. Selvagens archipelago: Berthelot's pipit populations across these archipelagos are separated by approximately 50 thousand years and two independent bottleneck events, with no evidence of post-colonisation gene flow between the archipelagos (Illera *et al.* 2007; Spurgin *et al.* 2014; Chapter 2). Berthelot's pipit within the Selvagens are not infected with avian pox or malaria, while there are strong differences in disease prevalence between islands within the Madeiran archipelago (Illera *et al.* 2008).

Table 2. Sampling information of divergence comparisons across populations of the Berthelot's pipit and mainland tawny pipit. Berthelot's pipit population codes: Canary Islands - LZ = Lanzarote, TF = Tenerife lowland, EH = El Hierro. Madeiran archipelago – M= Madeira island, PS = Porto Santo. Selvagens archipelago – Selvagem Grande = SG. *n* retained loci = autosomal and Z mapped loci retained for each population comparison.

Divergence comparison	Population locations	Estimated divergence timeframe	Bottleneck severity	*n* retained loci
Tawny pipit, sister species / Canary Islands	Mauritania, Mainland Africa / LZ, TF & EH	2.1 Mya	Weak. Genome-wide reduction in diversity.	10,829,660
Canary Islands / Madeiran archipelago	LZ, TF & EH / M & PS	50 kya	Strong. Founder effect, contemporary population recovery.	5,590,607
Canary Islands / Selvagens archipelago	LZ, TF & EH / SG	8-10 kya	Strong. Founder effect, contemporary inbreeding.	5,266,205
Madeiran archipelago / Selvagens archipelago	M & PS / SG	50 kya	Very strong. Two independent bottlenecks.	3,733,990

3.5 Differentiation landscapes and genomic islands

We calculated pairwise F_{ST} between the tawny pipit and the Canary Island Berthelot's pipit population (one comparison) using the "All Pipits" dataset, and between the three Berthelot's pipit archipelago populations using the "Berthelot's" dataset in VCFtools (Table 1). Variation across the genome was visualised using Manhattan plots.

To identify highly diverged regions occurring between population comparisons, pairwise F_{ST} values were Z-transformed. Since population history shapes genetic variation between populations, baseline levels of divergence vary significantly between each of the population comparisons (see Results). Within each comparison, we classified windows as divergent if their mean F_{ST} was more than five standard deviations above the genome-wide mean ($_{Z}F_{ST}$ > 5) and in the top 1% of SNP windows, which is a conservative approach to identifying outliers (Lamichhaney *et al.* 2015; Han *et al.* 2017; Walsh *et al.* 2019; Choi *et al.* 2020). For each population comparison, F_{ST} was calculated in 50 kb non-overlapping windows. Genomic linkage typically extends 25-35 kb in the three archipelago populations of Berthelot's pipit (see Chapter 2). We used a 50 kb window for genomic island detection because it provided sufficiently fine resolution across the genome while containing 184-532 sites per window between each population comparison. Further, windows containing <30 sites were removed prior to conducting analysis (~1% of windows). We considered the Z

chromosome separately due to known differences in evolutionary pressures across sex chromosomes.

By investigating genomic variation across the tawny pipit and Berthelot's pipit speciation event, it was possible to identify highly conserved genomic regions which may have a role in parallel adaptation between the species. To do this we applied the same $_{Z}F_{ST}$ approach, instead identifying windows with mean F_{ST} less than 5 standard deviations below the genome-wide mean.

3.6 Detailed characterisation of variation in divergence peaks

To identify regions putatively under selection in elevated regions of differentiation, we compared values of Tajima's D and nucleotide diversity (π) inside and outside of outlier windows (i.e., assessing whether regions of elevated differentiation had corresponding dips in Tajima's D and π). We calculate 50 kb-windowed Tajima's D and π using VCFtools, for each population dyad to allow direct comparison. Tajima's D detects deviations from neutral evolution, with values around zero signifying neutral evolution, while positive values indicate balancing selection or sudden population contraction, negative values indicate a recent selection sweep or population expansion following a recent bottleneck (Tajima 1989). Nucleotide diversity (π) is defined as the number of nucleotide differences per site between sequences within a population. Estimating π across the genome may reveal population-level diversity within genomic regions, which can be used together with Tajima's D to make inferences about potential evolutionary forces acting within regions of interest.

To identify genes located within divergent windows, we viewed 50 kb regions of interest using the Zebra finch genome (v. bTaeGut1_v1.p) in NCBI Genome Data Viewer v. 4.8. (www.ncbi.nlm.nih.gov/genome/gdv/browser). Patterns of divergence across peak regions were assessed across the three archipelago dyadic comparisons. Where several windows exceeded this threshold within a genomic island, we assessed the distribution of F_{ST} within peaks, and where appropriate highlight the most likely candidate genes under selection.

4 Results

4.1 Whole Genome re-sequencing

Illumina sequencing resulted in 1,139,170,057 paired-end reads (80 x10^6–120 x10^6 per individual), with a mean insert size of 401 bp. Genome alignment and mapping resulted in mean read coverage of 23.6 X ± 2.6 s.d. per individual, relative to the Zebra finch's 1.1 Gb genome (Ellegren *et al.* 2012). Reads were mapped to a contig level assembly Berthelot's pipit reference genome and genotypes joint called, resulting in an "All Pipits" dataset with 18,925,759 raw variants, and a "Berthelot's" dataset with 10,363,127 raw variants, of which 13,797,199 (72.9%) and 6,953,309 (67.1%) could be mapped to the Zebra Finch chromosomes, respectively.

Quality filtering resulted in 11,788,225 mapped SNPs in the "All Pipits" dataset and 5,934,934 in the "Berthelot's" dataset. Individuals had low levels of missing data even prior to variant quality filtering, with no individuals >5% missing data for the quality filtered variants. Sites with more than four missing variant calls were also removed from the "Berthelot's" dataset. Final depth of coverage for the quality filtered SNPs was high, with a mean 24.3 X (Table 1).

4.2 Distributions of F_{ST} between comparisons

Autosomal mean F_{ST} was high between the tawny pipit and Berthelot's pipit (Fig. 2; Chapter 3) but low between the Berthelot's pipit archipelago populations separated by one colonisation event (Canary Islands vs. Madeira, Canary Islands vs. Selvagens). The distribution of F_{ST} was positively skewed between the Canary Islands and both subsequently colonised archipelagos (Madeira and Selvagens) and approaches a normal distribution between the tawny pipit and Canary Islands Berthelot's pipit. Between single colonisation events, standard deviations of windowed F_{ST} values were low, ranging 0.041–0.070. The spread of divergence scores was much greater between the Selvagens and Madeiran archipelago (standard deviation = 0.139), which are separated by two independent colonisation events (Fig. 2).

Like the autosomes, the Z chromosome showed increasing divergence over longer timeframes between single colonisation events, with highest divergence between the tawny

pipit and Berthelot's pipit and lowest between the most recently separated Berthelot's pipit archipelago population (i.e., the Canary Islands and Selvagens; Fig. S1A). Between the tawny and Berthelot's pipit, mean F_{ST} divergence across the Z chromosome exceed that observed across the autosomes, while the converse was found across Berthelot's pipit archipelago populations (Table 3).

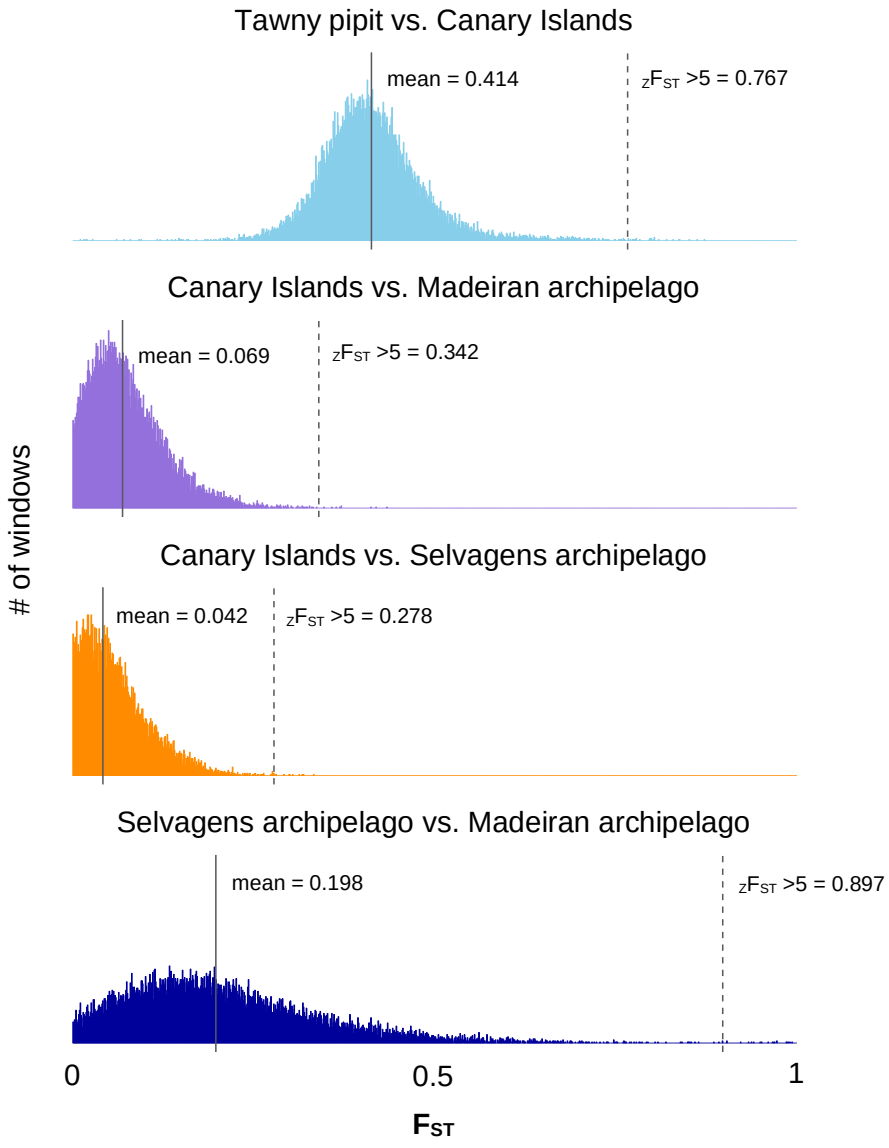


Figure 2. Distribution of pairwise genomic differentiation, F_{ST}, across Berthelot's and tawny pipit comparisons calculated in 50 kb autosomal windows. Positions of means (solid line) and $_zF_{ST}$ > 5 threshold (dotted black line) are highlighted. Maximum number of windows (x axis) = 200.

4.3 Overall correlations between genomic landscapes of divergence

The divergence at genomic loci in tawny-Berthelot's pipit comparisons was significantly but weakly correlated with divergence at these loci between Berthelot's archipelago comparisons (Fig. 3, Pearson's correlation: r = 0.02–0.15, P <0.01). As expected, the divergence of loci between dyadic comparisons was more strongly correlated when those comparisons involved a common population (because of the characteristics of loci in that common population) and not because of independently reoccurring divergence of the same loci. The highest correlation explains 34.8% of variation in dyadic F_{ST} scores (Pearson's correlation: r = 0.59, P <2.2 x 10^{-6}, Canary Islands vs. Madeira compared to Selvagens vs. Madeira).

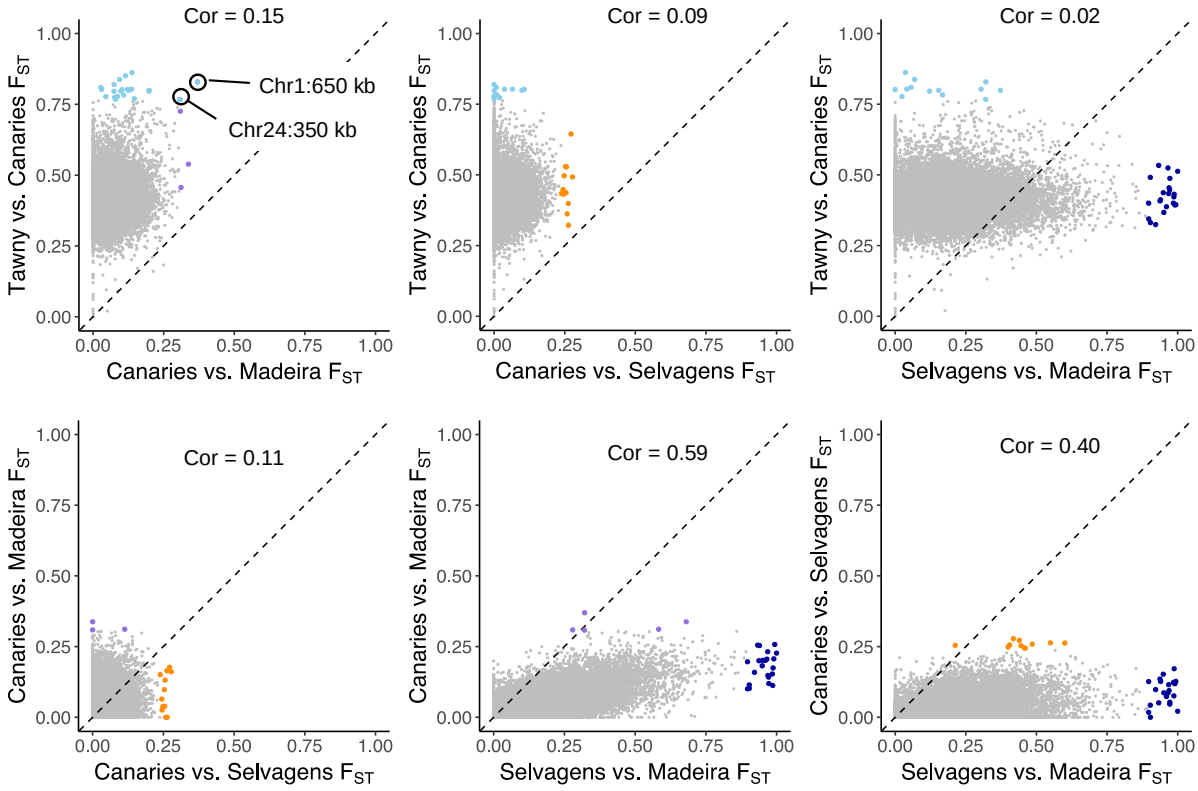


Figure 3. Correlated patterns of F_{ST} between all Berthelot's pipit population comparisons, and between Berthelot's pipit and tawny pipit. F_{ST} values were calculated in 50 kb non-overlapping windows, outliers with 5 > $_{Z}F_{ST}$ are highlighted. Chromosomal positions of the only two overlapping outlier windows ($_{Z}F_{ST}$ > 5) between multiple population comparisons are indicated. Population comparison highlights: light blue = Tawny vs. Canary Islands, Canary Islands vs. Madeira = purple, Canary Islands vs. Selvagens = orange and Selvagens vs. Madeira = navy.

4.4 Identification of peaks and valleys and correlations between population comparisons

The genomic landscape of divergence between the tawny pipit and the Canary Islands Berthelot's pipit population was broadly homogeneous across the genome, with 27 windows more than five standard deviations from the genomic autosomal mean ($F_{ST} > 0.41$) (Fig. 4, Table 3). The distribution of these areas of elevated divergence was non-random with 12 well defined clusters of high F_{ST}, seven of which exceeded the criteria to be considered 'divergence peaks' or 'genomic islands of divergence' (Fig. 4, Table 4). Peak size was in the range of 0.05–2 Mb (Table 4) and such divergence peaks covered 0.14% of the genome. Across the Z chromosome, two broad peaks (1-2 Mb) of strong divergence are also observed between the tawny and Canary Island Berthelot's pipit. The pattern of divergence within the different genomic islands fell into two categories: i) most commonly, regions of sharp divergence usually include one or two peaks where a "top peak" with the highest associated F_{ST} could be identified or, ii) a broad peak of similar F_{ST} divergence across the island (see Fig. 5). Only one such broad peak of divergence was identified approximately 2 Mb in length on chromosome 1A (Fig. 5A), with the strongest F_{ST} divergence scores across the genome.

Autosomal divergence between Berthelot's pipit archipelagos showed different patterns. Archipelago comparisons with just one founding step separating them (Canary Islands vs. Madeira and Canary Islands vs. Selvagens) were characterised by low genome-wide divergence with few strongly differentiated 'islands of divergence' (Fig. 2, Fig. 4). In contrast, the divergence landscape was highly heterogenous when the two independently bottlenecked archipelagos, Selvagens vs. Madeira were compared. We identified 9-22 genomic islands of divergence in each Berthelot's pipit archipelago comparison (Fig. 4, Table 3), which represented 0.05-0.1% of 50 kb windows. In contrast to the tawny vs. Berthelot's pipit comparison, Z chromosome divergence was consistently (but marginally) lower than autosomal divergence between Berthelot's pipit archipelagos comparisons (Table 3), and no strongly divergent genomic windows were identified.

The position of divergent windows varied between population comparisons, with only two shared windows identified between multiple population comparisons (Fig. 3). These two windows, mapped to regions of chromosome 1 and 24, and were shared between the tawny pipit vs. Canary Island Berthelot's pipit population comparison and between Canary Island

Berthelot's pipit vs. Madeiran archipelago comparison. No windows of strong divergence were shared between the different Berthelot's pipit comparisons.

Between the tawny and Berthelot's pipit, the majority of the genome exhibited strong divergence, with the highest window having F_{ST} = 0.87, and only two windows where F_{ST} was zero (Fig. 2 and 4). However, we identified seven strongly conserved genomic regions, more than five standard deviations below the genome-wide mean (so called genomic valleys of divergence), which mapped to five distinct chromosome regions (Fig. 4). Between Berthelot's pipit archipelagos comparisons, many 50 kb regions of the genome exhibited complete sequence conservation (F_{ST} = 0, Fig. 4).

Table 3. Genomic differentiation, F_{ST}, between the Berthelot's pipit and tawny pipit, and among Berthelot's pipit archipelagos.

Comparison	Autosomal mean F_{ST}	Z Chr mean F_{ST}	Autosomal $_{Z}F_{ST}$ > 5 threshold	Top 1% SNPs threshold	*n* windows $_{Z}F_{ST}$ > 5
Tawny pipit, sister species / Canary Islands	0.414	0.427	0.767	0.637	27
Canary Islands / Madeiran archipelago	0.069	0.063	0.342	0.229	9
Canary Islands / Selvagens archipelago	0.042	0.034	0.278	0.186	10
Selvagens archipelago / Madeiran archipelago	0.198	0.171	0.897	0.609	22

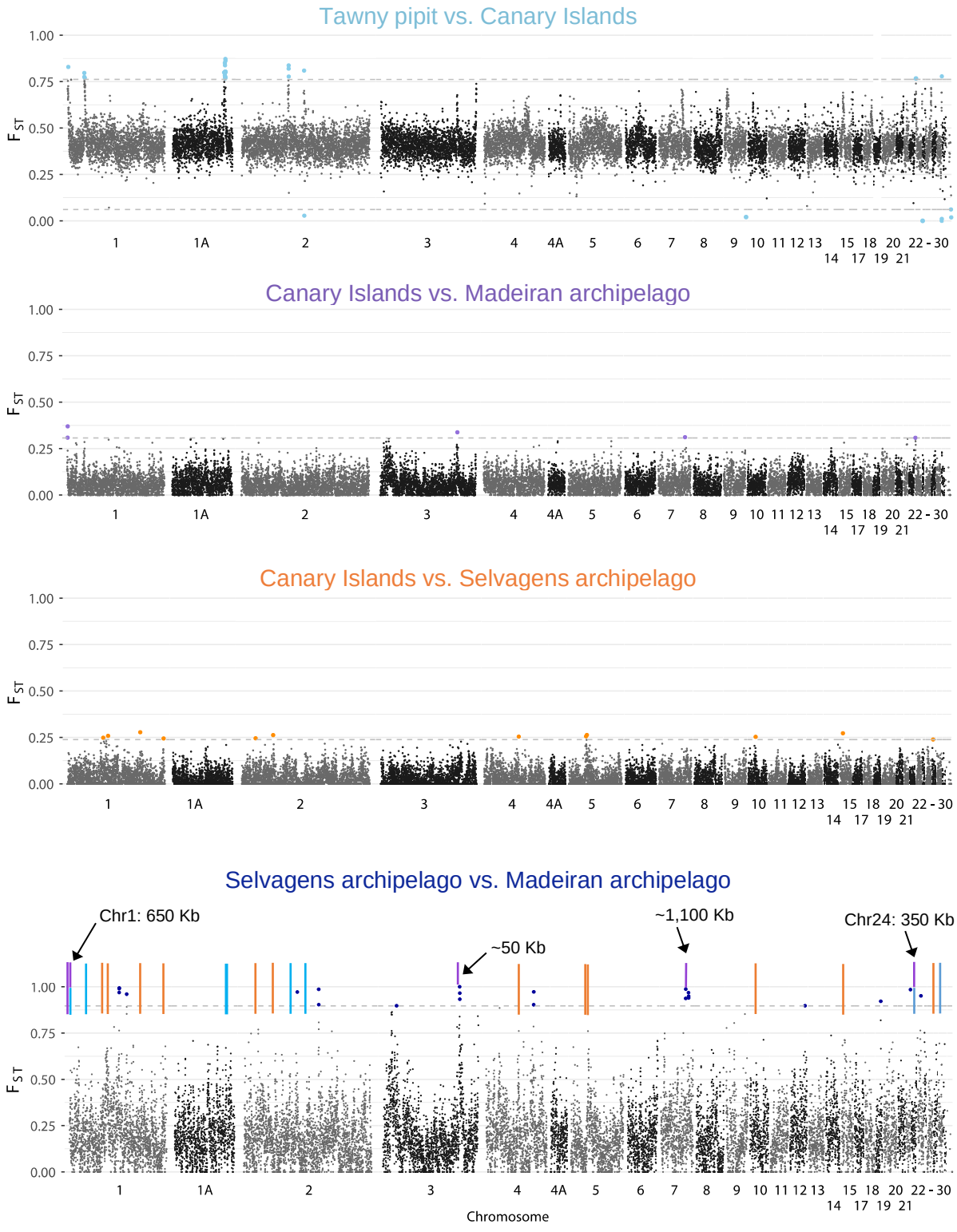


Figure 4. Pairwise genomic differentiation, F_{ST}, across the genome for Berthelot's and tawny pipit comparisons calculated in non-overlapping 50 kb windows. Genomic islands of divergence or valleys of similarity, where F_{ST} is 5 standard deviations greater than or less than the mean window value, are highlighted (indicated by the dashed horizontal line). Chromosomes (derived by comparison with the zebra finch genome) are shown in alternating light and dark shading. Vertical coloured bars indicate the location of genomic islands within each population comparison. Labelled arrows indicate the location of shared

genomic islands and the genomic distance between closely located peak windows across population comparisons. Population comparison highlights: light blue = Tawny vs. Canary Islands, Canary Islands vs. Madeiran islands = purple, Canary Islands vs. Selvagens = orange and Selvagens vs. Madeira = navy.

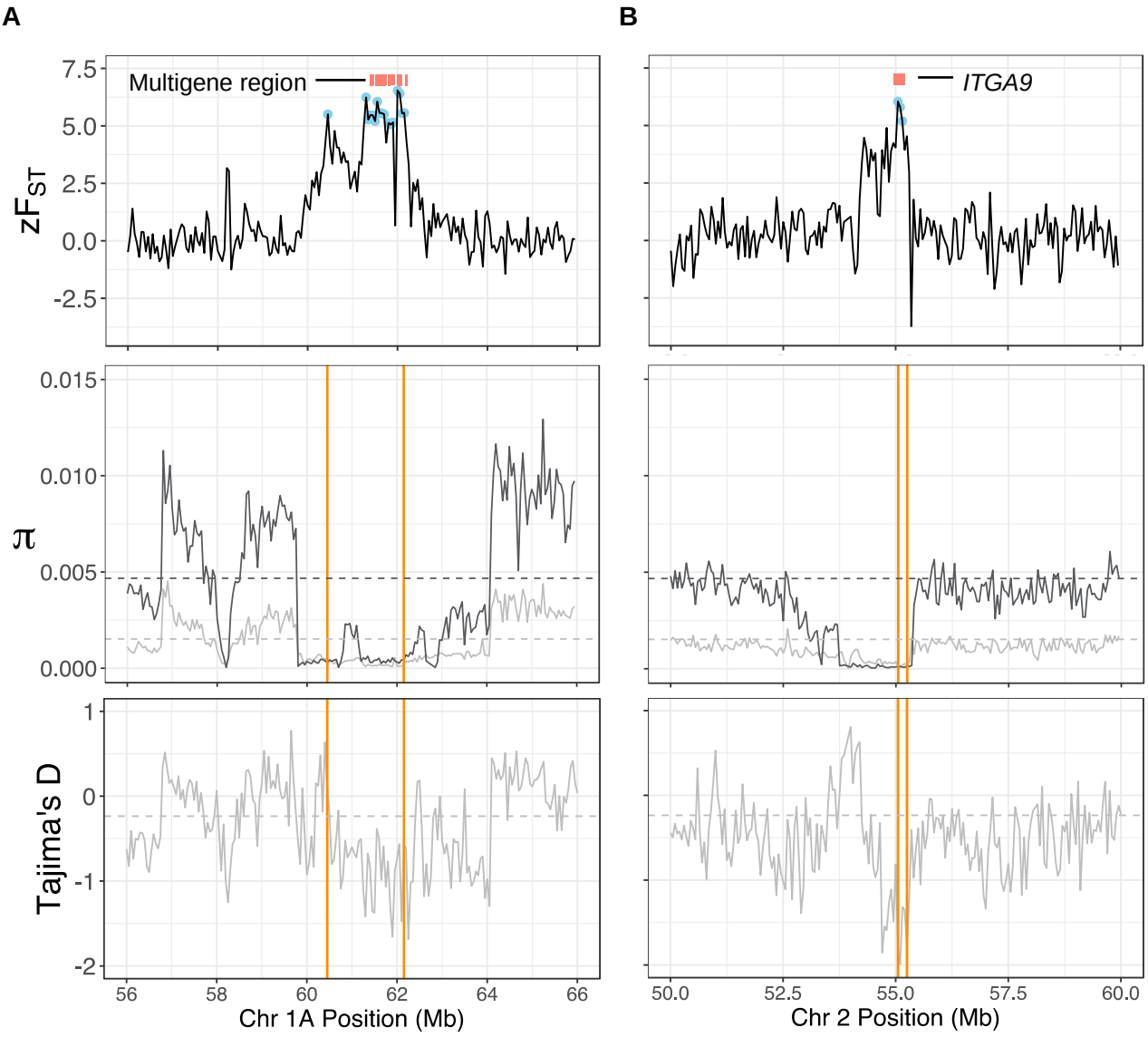


Figure 5. Patterns of divergence ($_zF_{ST}$), genetic diversity (π) and a measure of the loss of rare alleles (Tajima's D) surrounding two genomic islands of divergence between the tawny pipit and Canary Islands Berthelot's pipit comparison. These regions are presented as examples to demonstrate **A)** a broad region, or 'plateau' of elevated divergence and **B)** a narrow peak of divergence. Values were calculated in 50 kb windows, with $_zF_{ST} > 5$ highlighted in the first panel in light blue, with corresponding genomic locations of the peak start and end indicated by vertical orange lines in panel 2 and 3. Candidate gene locations within peaks are indicated; details in Table 4. Tajima's D and π are reported for the Canary Islands (light grey) and the corresponding π indicated for the tawny pipit (dark grey), with autosomal averages indicated by horizontal dotted lines.

4.5 Genes in peaks and valleys and patterns of diversity

The identified regions of elevated divergence between the tawny and Berthelot's pipit, consistently exhibit corresponding dips in Tajima's D and π in the Canary Islands Berthelot's pipit population, indicative of selective sweeps. Patterns of reduced Tajima's D and π, and elevated F_{ST} were consistent across the broad peak on chromosome 1A, suggesting strong linkage disequilibrium in this region, while the other strongly divergent peaks were narrow in width. Across all strongly divergent regions, five of seven peaks harboured named candidate genes (Table 4).

A high number of divergence windows were located within the extended genomic island mapped to chromosome 1A (17 of 27 windows identified between the tawny and Berthelot's pipits). This region contained 17 annotated genes, of which at least seven are associated with immune response, and three with craniofacial development (see Table 4). The single most strongly divergent peak in the region included two genes: *CMAS* associated with the innate immune response (O'day *et al.* 2018; Urbanek *et al.* 2020) and *ABCC9*, associated with cartilage and bone development (Czeschik *et al.* 2013) (Table 4).

To further investigate divergence across this broad region on chromosome 1A, we investigated patterns of divergence between the tawny pipit and all Berthelot's pipit populations, and among Berthelot's pipit populations. The results confirm i) divergence of this region occurs between the tawny pipit and all Berthelot's pipit archipelagos and ii) this region exhibits low divergence between Berthelot's pipit archipelagos (Fig. 6). The other narrow regions (< 150 kb) of strong divergence between the tawny pipit and Berthelot's pipit mapped to genomic regions including genes putatively associated with immune response and wound healing (Cui *et al.* 2017; Sigurðarson 2020), development of the retina (Xu *et al.* 2020) and carbohydrate metabolism (Han *et al.* 1999) (Table 4).

Across the Z chromosome, two broad peaks (1-2 Mb) of strong divergence occur between the tawny pipit and Canary Islands Berthelot's pipits. These have corresponding dips in nucleotide diversity in both populations (Fig. S1B), which suggests they may be under divergent selection. Identified genes within the top 50 kb divergent window in both peaks, both have strong association with hearing: *ADGRV1* is associated with hearing loss and retina development (Yan *et al.* 2018); and *PPIP5K2* regulates hearing through growth and

maintenance of sensory cells in the inner ear (Yousaf *et al.* 2018). Nucleotide diversity across the Z chromosome shows a consistent peak across all populations of high diversity at ~45 Mb, likely an artefact of incorrect mapping of W chromosome to the Z.

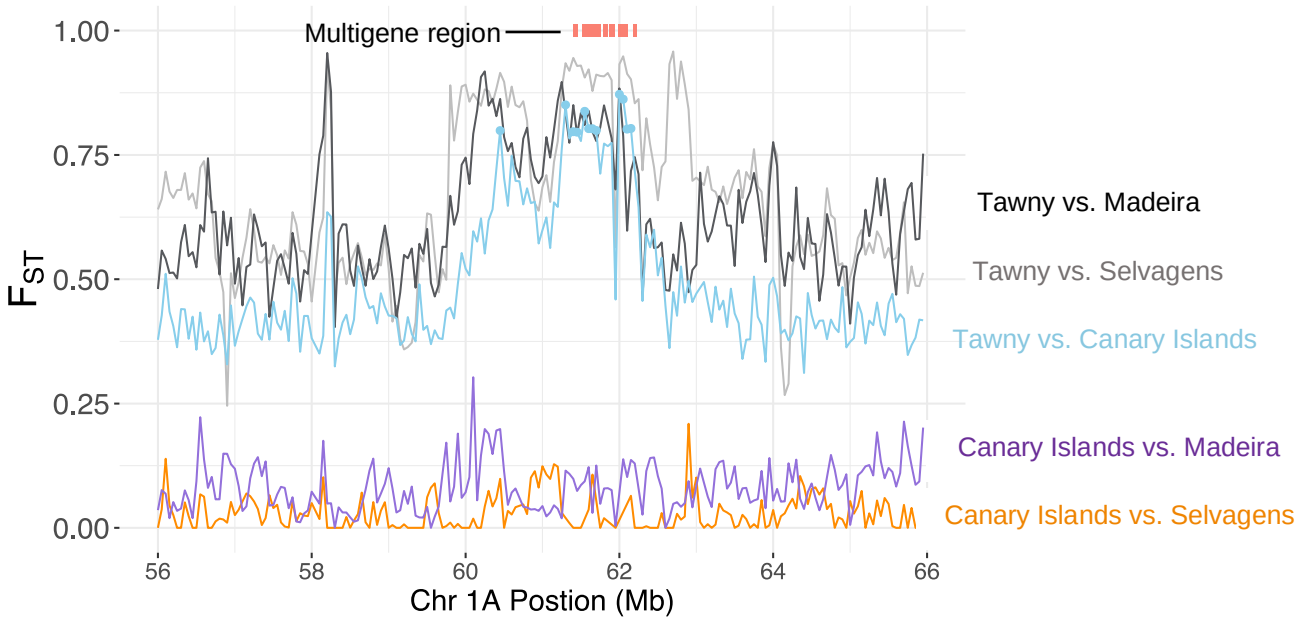


Figure 6. Consistent patterns of divergence between Berthelot's pipit and tawny pipit, and across Berthelot's pipit populations for chromosome 1A ~2 Mb genomic island of divergence. Comparisons are made between the tawny pipit and all three Berthelot's pipit archipelago populations, as well as across Berthelot's pipit archipelago populations. The light blue line represents divergence in relation to the initial island colonisation, with locations of genomic islands highlighted by circles. The corresponding location of candidate genes within strongly divergent 50 kb windows are indicated in pink rectangles.

Table 4. Genes within genomic islands of divergence identified between the tawny pipit and the first colonised Berthelot's pipit archipelago (the Canary Islands) ~ 2.1 Mya. Genes under putative positive selection in the Canary Islands are identified, due to corresponding dips in nucleotide diversity and Tajima's D when compared to the tawny pipit. F_{ST} calculated in 50 kb non-overlapping windows across autosomes, with $_zF_{ST} > 5$. All genes within a window are noted. All retained windows have > 30 SNPs. Genomic location given for the start of the 50 kb window.

Highest $_zF_{ST}/F_{ST}$	Genomic location (Chr: kb)	Candidate gene(s)	Gene name	Putative function	No. windows
Chr 1A Broad Peak: 60450-62150 kb					
6.54/ 0.87	1A: 62000	*CMAS*	Cytidine Monophosphate N-Acetylneuraminic Acid Synthetase	Innate immune response. Synthesis of cell surface proteins, pathogen binding and invasion (O'day *et al.* 2018; Urbanek *et al.* 2020).	1
		ABCC9	ATP binding cassette subfamily C member 9	Membrane transport affecting cartilage and bone development incl. craniofacial (Harakalova *et al.* 2012).	2
6.40/ 0.86	1A: 62050	*KCNJ8*	Potassium Inwardly Rectifying Channel Subfamily J Member 8	Wound healing, T cell regeneration and cardiac homeostasis during innate immune response (Zhang and Bei 2015; Zhang *et al.* 2020).	2
6.24/ 0.85	1A: 61300	Un. protein	Unknown.	Unknown.	1
6.06/ 0.84	1A: 61550	*SCUBE1*	Signal peptide, CUB and EGF-like domain-containing protein 1	Modulates bone morphogenic protein (BMP) family activity, craniofacial development (Tu *et al.* 2008; Xavier *et al.* 2009).	6
5.57/ 0.80	1A: 62150	*GYS2*	Glycogen synthase 2	Glycogen storage in liver and adipose tissue (Mandard *et al.* 2007).	1
5.52/ 0.80	1A: 61700	*TTLL12*	Tubulin Tyrosine Ligase-Like 12	Cellular antiviral signalling (Ju *et al.* 2017).	1
		TSPO	18-kDa Translocator Protein	Apoptosis and host defence, mitochondrial function (Veenman *et al.* 2007).	1
		MCAT	malonyl-CoA-acyl carrier protein transacylase	Function in vertebrates largely unknown.	1
5.52/ 0.80	1A: 60450	Un. proteins	Unknown.	Unknown.	1
5.49/ 0.80	1A: 61400	*MPPED1*	Metallophosphoesterase-domain-containing protein 1	Diverse signalling pathways; innate immune system, aging and life span (Gupta *et al.* 2020).	2
5.17/ 0.77	1A: 61900	*STYK1*	Serine/ threonine/ tyrosine kinase 1	Autophagy and cell invasion (Zhou *et al.* 2020).	1

		GABARAPL1 / GEC1	GABA Type A Receptor Associated Protein Like 1	Autophagy (Chakrama *et al.* 2010).	1
5.15/ 0.77	1A: 61800	*PACSIN2*	Protein Kinase C and Casein Kinase Substrate in Neurons 2	Vesicle mediated transport. Regulates ADAM13, cephalic (head) neural crest activity (Cousin *et al.* 2000).	1
5.07/ 0.77	1A: 61850	*YBX3*	Y-Box Binding Protein 3	Cold shock and degrade viral function (Qin *et al.* 2020).	1
		SYCE3	Synaptonemal complex central element protein 3	Mediates homologous chromosome synapsis during meiosis (Lu *et al.* 2014).	1
Chr 2 Peak: 55050-55200 kb					
6.06/ 0.84	2: 55050-55200	*ITGA9*	Integrin alpha 9	Cell adhesion and invasion. Immune response in chicken, potential link to wound healing (Gupta and Vlahakis 2010; Arnar Kári and Sandholt 2020).	2
Chr 1 Peak: 650 kb					
5.95/ 0.83	1: 650*	*IMPG2*	Interphotoreceptor Matrix Proteoglycan 2	Eye development; interphotoreceptor of the retina (Xu *et al.* 2020).	1
		SENP7	Sentrin-specific protease 7	Innate immune response; cytosolic DNA-triggered antiviral gene expression (Cui *et al.* 2017).	1
Chr 2 Peak: 7330 kb					
5.67/ 0.81	2: 7330	*PDE1C*	Calcium-dependant Phosphodiesterase	Glucose-induced insulin secretion, smooth muscle incl. cardiac (Han *et al.* 1999; Yan and Zhu 2007).	1
Chr 1 Peak: 19550-20100 kb					
5.48/ 0.80	1: 19550 1: 19500 1: 20100	-	-	-	
Chr 28 Peak: 5750 kb					
5.22/ 0.78	28: 5750	Un. protein	Unknown.	Unknown.	1
Chr 24 Peak: 350 kb					
5.06/ 0.77	24: 350	*ST14*	Suppression of tumorigenicity 14	Epidermal skin development (Alef *et al.* 2009).	1
		NFRKB	Nuclear factor related to kappa-B-binding protein	Upregulated in CD4+ T cells and B cells (Audard *et al.* 2012).	1
		PRDM10	PR domain zinc finger protein 10	Embryonic tissue differentiation, incl. craniofacial (Park and Kim 2010).	1

Strongly divergent regions between archipelago populations of Berthelot's pipit had varied patterns of π and Tajima's D across different populations, indicating a range of evolutionary processes occurring within these regions (Table 5). Across the Canary Islands vs. Selvagens comparison, several highly divergent regions associated with long runs of homozygosity (ROH) >1 Mb in the Selvagens (and negative Tajima's D), extending either side of genomic islands of divergence identified in the associated region. Across the strongly divergent regions for the Canary Islands vs. Madeiran archipelago comparison, Tajima's D was consistently elevated within the Madeiran archipelago.

Genes in islands of divergence across the Berthelot's pipit range are associated with similar biological functions (Table 4 and 5). Regions under putative selection within the Selvagens include genes associated with craniofacial shape, apoptosis and inflammation, development of the retina and teeth, metabolism, muscle and growth (Table 5). Within the Madeiran archipelago population regions potentially under selection include genes associated with facial, skin and bone development, immunity (innate immune response and regulation of B and T cells) and eye development.

Table 5. Genomic islands of divergence through archipelago colonisation in Berthelot's pipit. F_{ST} calculated in 50 kb non-overlapping windows across autosomes, with areas $_{Z}F_{ST} > 5$ identified. All genes within a window are noted. All retained windows have > 30 SNPs. Genomic location at the start of 50 kb window are given. Population abbreviations: Canary Islands (CI), Selvagens (SG) and MA (Madeiran archipelago). Statistical abbreviations: nucleotide diversity (π), Tajima's D (TD) and runs of homozygosity (ROH).

$_{Z}F_{ST}$ / F_{ST}	Genomic location (Chr: kb)	Candidate gene(s)	Putative function	Putative evolutionary driver
Canary Islands vs. Selvagens: ~ 10 kya				
5.97/ 0.28	1: 85500	Un. protein	Unknown.	-
5.84/ 0.27	15: 4100	*CIT*	Craniofacial development (Shaheen *et al.* 2016).	*Selective sweep in SG:* Low π in all pops; 0 TD CI & -0.75 TD SG.
		PRKAB1	Stress response through healing and inflammation (Sahoo *et al.* 2021); lipid and carbohydrate metabolism, growth (Kim *et al.* 2008).	
5.61/ 0.26	5: 20950	*AP15*	Stress and apoptosis (Jiang *et al.* 2019).	*Selective sweep in SG:* 5 Mb ROH in SG & average π in all other pops; 0 TD in CI & -0.75 TD SG.
5.60/ 0.26	2: 36750	-	-	-
5.50/ 0.26	1: 47600	-	-	-
5.43/ 0.26	5: 19950	-	-	-
5.40/ 0.25	4: 40550	-	-	-
5.37/ 0.25	10: 8750	*WDR72*	Teeth (El-Sayed *et al.* 2009), retina (Liu *et al.* 2019).	*Selective sweep in SG:* 4 Mb ROH in SG & average π in all other pops; 0 TD in CI & -0.75 TD SG.
5.25/ 0.25	1: 41950	-	-	-
5.17/ 0.25	2: 16150	*WAC*	Regulates autophagy pathway mediated by GEC1 (Joachim *et al.* 2015); loss-of-function causes facial, behavioural and muscle abnormalities (De Santo *et al.* 2015).	*Selection (?):* No change π in both; -0.5 TD in CI & -0.75 TD in SG.
5.15/ 0.24	1: 113050	*SMIT1/ SLC5A3*	Apoptosis and inflammation (Lu *et al.* 2021), linked to osmotic stress (Gao *et al.* 2021).	*Selective sweep in SG(?*): Genomic architecture, near end of Chr: Low π in all pops; 0 TD in CI & -0.75 TD in SG.
		ATP5PO/ ATP5O	Skeletal muscle, glucose uptake and fat retention (Rönn *et al.* 2009).	
5.01/ 0.24	27: 1400	*SOCS7 / NAP4*	Interacts with growth factor receptor and insulin signalling pathway (Elliott and Johnston 2004).	*Selective sweep in SG:* Low π in all pops; +0.75 TD in CI & -0.75 TD in SG.

		ARHGAP 23 / Rho GAP	Cellular morphology (Gingras *et al.* 2020).	
Canary Islands vs. Madeira : ~ 50 kya				
6.30/ 0.37	1: 650	*IMPG2* *SENP7*	Eye development; interphotoreceptor of the retina (Xu *et al.* 2020). Innate immune response; cytosolic DNA-triggered antiviral gene expression (Cui *et al.* 2017).	*Genomic architecture, balancing or divergent selection or recent inbreeding in MA:* Low π in all pops; -1 TD CI & +1 TD in MA.
5.65/ 0.34	3: 9140	*NRXN1*	Cognitive development and facial features (Zahir *et al.* 2008).	*Balancing selection in MA:* No change π in all pops; 0 TD in CI, +2 TD in MA.
5.10/ 0.31	7: 31300	Un. protein	Unknown.	-
5.05/ 0.31	1: 550	*ABI3BP*	Cardiovascular (Delfín *et al.* 2019) and bone development (Zhang *et al.* 2014).	*Balancing selection in MA, potential sweep in CI:* No change π in all pops, near start of Chr; -1 TD in CI, +0.75 TD in MA.
5.04/ 0.31	24: 350	*ST14* *NFBKB* *PRDM10*	Epidermal skin development (Alef *et al.* 2009). Upregulated in CD4+ T cells and B cells (Audard *et al.* 2012). Embryonic tissue and bone differentiation, incl. craniofacial (Park and Kim 2010).	*Balancing selection or recent inbreeding in MA*: Low π in all pops, near start of Chr; -0.25 TD in CI, +1.5 TD in MA.

4.6 Genes conserved through speciation

Between the tawny pipit and Berthelot's pipit, we identify seven strongly conserved genomic regions including a total of 11 candidate genes, of which seven were annotated. Regions included genes putatively associated with pathogen infection, inflammation and platelet regeneration, growth factor pathways, and muscle and limb development (Table 6). For example, *NFX1* is a transcriptional repressor of major histocompatibility complex (MHC) class II genes (Strominger *et al.* 1994; Gewin *et al.* 2004) with biological function in immune and inflamatory response. We also identify a genomic valley with corresponding peaks of nucleotide diversity in the Berthelot's pipit (and negative Tajima's D) and low diversity in the tawny pipit, putatively assocaited with recent popualtion growth across the Canary Islands.

This loci was within the genomic region containing *RANBP3* which is associated with growth factor pathways including Bone Morphogenic Protein (BMP) signalling (Dai *et al.* 2009).

Table 6. Genomic valleys of similarity through speciation between Berthelot's pipit and the tawny pipit. F_{ST} calculated in 50 kb non-overlapping windows across autosomes, windows with $_zF_{ST} < 5$ identified. All genes within a window are noted. All retained windows have > 30 SNPs. Genomic location at the start of 50 kb window is given. CI = Canary Island Berthelot's pipit populations.

Lowest $_zF_{ST}/F_{ST}$	Genomic location (Chr: kb)	Candidate gene(s)	Putative function	Putative evolutionary driver in CI
-5.88/ 0.000	25 : 0	Un. proteins	Unknown.	Neutral.
-5.88/ 0.000	28: 5650	*RANBP3*	Nuclear transport - regulation of growth factor pathways and BMP family (Dai *et al.* 2009). Viral infection (Cho *et al.* 2020).	*Selective sweep or recent population expansion.* High π in BP, low π in tawny.
	28: 5700	*NDUFA11*	Electron transport chain in mitochondria (Jang and Javadov 2018).	
-5.61/ 0.019	30: 5850	Un. proteins	Unknown.	*Balancing selection or sudden population contraction.* High π in BP, low π in tawny.
-5.00/ 0.061	30: 5550	Un. protein	Unknown.	
-5.59/ 0.020	9: 25500	*CP*	Iron transport and inflammation (Harris 2018).	*Parallel selection.* Low π in all.
		HPS3	Pigmentation and blood platelets (Liu *et al.* 2021).	
		GYG1	Muscle and limb defects (Yaou *et al.* 2017).	
-5.48/ 0.028	2: 73500	*SLC12A7*	Solute cotransporter, interacts with insulin-like growth factor (Brown *et al.* 2019).	*Near neutral, flanked by loci under parallel selection.* Low π in all.
		NFX1	Transcriptional repressor of MHC class II genes, immune and inflammatory response (Strominger *et al.* 1994; Gewin *et al.* 2004).	

5 Discussion

Here we take advantage of an island bird system where we have a detailed understanding of colonisation history, population demography and gene flow (see Chapter 3), to gain insight into how and where genomic divergence accumulates across a speciation continuum. Our results provide a characterisation of genomic differentiation across different timeframes and potential selection gradients in the absence of post-colonisation gene flow. The distributional shift in genetic divergence among populations through time was in concordance with models of divergence hitchhiking with a lack of gene flow, where few selected loci and linked regions differentiate during early stages of divergence, followed by drift causing other loci to also become differentiated, resulting in genome-wide divergence. We also identified a few highly conserved genomic regions between species, which are potentially a result of balancing selection acting on loci associated with Bone Morphogenic Protein (BMP) signalling and major histocompatibility complex (MHC) class II regulation - key regulators of growth and the immune response. We find only weakly correlated patterns of genetic divergence between populations through founder events (i.e., the same loci diverging though each event), detecting mostly unique divergence patterns across archipelagos. We do however detect genomic islands of divergence commonly associated with ecologically important traits of body/head size and immune defence across the range, as well as eye development.

By comparing F_{ST} frequency distributions between sequentially founded, geographically isolated Berthelot's pipit populations, we show how genomic differentiation accumulates across timeframes in the absence of gene flow (Fig. 2). Berthelot's pipit populations in the early stages of divergence (Canary Islands vs. Selvagens) after a recent colonisation event (<10 kya), have a positively skewed F_{ST} frequency distribution. This includes a handful of strongly divergent genomic regions, which may differ due to differential selection or bottleneck effects. Where populations have been geographically separated for longer timescales (i.e., Canary Islands vs. Madeiran archipelago 50 kya) leading to subspecies, drift continues to act causing overall genomic divergence (Nosil *et al.* 2009). Over much longer divergence timeframes, i.e., across the speciation event between the tawny pipit and Berthelot's pipit separated >2 Mya, divergence accumulates across the whole genome (mean F_{ST} = 0.41), approximating a normal distribution with only a few strongly conserved or divergent regions. These patterns are in concordance with speciation models from early to late stage divergence with geographical isolation (reviewed by Seehausen *et al.* 2014).

Across the Berthelot's pipit system, we can also compare genomic divergence between two independently bottlenecked populations, founded from the same source population (Canary

Islands) at different time points (50 kya and 8-10 kya) (Fig. 1). Here, the F_{ST} frequency distribution has a greater spread, likely due to the combined consequences of two founder events and subsequent drift. Unlike the comparisons of populations separated by one colonisation event, regions that have diverged due to natural selection are more difficult to identify due to high genome-wide divergence. Our patterns of genomic divergence among the different Berthelot's pipit archipelago populations concurs with comparisons of rapidly speciating island silvereye (*Zosterops lateralis*) populations separated by <200 years (Sendell-Price *et al.* 2021) where they only identified a few genomic regions under divergence. In contrast, between-species comparisons of *Ficedula* flycatchers with divergence times of several millions of years reflected the normal distribution we observed between the tawny pipit and Berthelot's pipit (Burri *et al.* 2015).

We identified weakly correlated patterns of F_{ST} across genomes from the different Berthelot's pipit populations, and between it and its sister species, the tawny pipit (Fig. 3). The fact that the same loci/regions are not divergent across population comparisons, suggests that long-term linked selection is not a major reason for divergence in the Berthelot's pipit. Instead such divergent loci may be due to independent evolutionary responses to selection pressures (see Munch *et al.* 2016). Our findings contrast with several studies of genomic landscapes across divergence timescales and through varied gene flow contexts (Renaut *et al.* 2013; Burri *et al.* 2015; Van Doren *et al.* 2017; Vijay *et al.* 2017; Stankowski *et al.* 2019), which show parallel patterns of divergence between geographically and morphologically distinct taxa. Instead, we find only weakly correlated F_{ST} scores and different putatively selected genomic regions diverging (see below) through time and speciation events across the range (Fig. 3 and 4). Why this is the case across the Berthelot's pipit system remains unknown, but it may be attributed to the stochastic nature of the founding process, as has been reported in island colonising silvereyes (*Zosterops lateralis*; Sendell-Price *et al.* 2021) and laboratory range expansion experiments of red flour beetle (*Tribolium castaneum*; Weiss-Lehman *et al*. 2019). Largely independent patterns of divergence between population pairs are reported in a range of other systems, for example through parallel speciation of stick insects (*Timema cristinae*; Soria-Carrasco *et al.* 2014), sympatric environmental adaptation of flatfish species (Le Moan *et al.* 2019) and adaptive radiation in Lake Victoria *Pundamilia* cichlid fishes (Meier *et al.* 2018).

Absence of secondary contact between the tawny and Berthelot's pipit for ~2 million years has resulted in considerable genome-wide divergence. We identified seven strongly conserved and seven strongly divergent autosomal genomic regions (Fig. 2 and 4). The

most divergent region formed a broad ~2 Mb peak on chromosome 1A. We also identified two peaks of broad divergence across the Z chromosome (Fig. S1). All other strongly conserved or divergent regions formed narrow sharp peaks. Through subsequent independent founder events from the Canary Islands to the Selvagens and Madeira, we identify 12 and five strongly divergent 50 kb windows, respectively, all mapped to autosomal chromosomes (Table 5). The majority of identified windows were unique to a specific population comparison. The only instance where we identify the same divergent genomic windows (two 50 kb windows), occurs between the tawny pipit and first colonised Berthelot's pipit archipelago, the Canary Islands, and between the Canary Islands and Madeiran archipelago, potentially due to strong divergence within the Canary Islands being detected within both population comparisons. We were unable to identify any strongly conserved genomic regions between Berthelot's pipit archipelago populations due to the lack of overall genomic divergence in these relatively recently separated populations (Fig. 2).

A clear relationship between reduced Tajima's D and reduced nucleotide diversity within genomic islands of divergence was found between the tawny and Berthelot's pipit, which could be indicative of selective sweeps on specific loci (Fay and Wu 2000) after the species split. However, this pattern was not observed between archipelago populations of Berthelot's pipit (Table 5). Highly divergent genomic regions between the Canary Islands and Selvagens consistently exhibited a strong decrease in Tajima's D within the Selvagens, commonly associated with large regions of low diversity or runs of homozygosity (ROH). Negative Tajima's D can result from either recent selective sweeps or population expansion following a bottleneck. In the case of the Selvagens pipit population we have strong evidence of a very small contemporary effective population size with considerable inbreeding (Chapter 3). Consequently, it seems likely that the genomic islands of divergence observed result from selective sweeps. Such patterns of selection have previously been reported in other wild populations despite small effective population size (e.g., de Jong *et al.* 2020) and over short time periods (e.g., Walsh *et al.* 2019). By contrast, genomic islands of divergence between the Canary Islands and Madeiran archipelago population were associated with increased Tajima's D values in Madeira which may result from balancing selection or sudden population contractions. Windows with corresponding moderate levels of genetic diversity may indicate balancing selection is driving divergence within these regions, as inbreeding effects due to a population contraction would be expected to result in low diversity regions. Similar patterns of balancing selection have been reported in many systems including in the maintenance of morphological crypsis in stick insects (*Timema*

cristinae; Lindtke *et al*. 2017) and malaria parasites (*Plasmodium falciparum*; Tetteh *et al.* 2009).

Our findings regarding the drivers of the islands of genomic divergence should be interpreted with caution: although negative Tajima's D scores indicate selective sweeps, they can also be associated with population expansion following a recent bottleneck (Tajima 1989), and estimates calculated from small size samples are prone to error (Galtier *et al.* 2000; Subramanian 2016). We cannot rule out the possibility that the Canary Island Berthelot's pipit populations may have experienced a contraction in the recent past; indeed, we provide some evidence for this possibility in Chapter 3. Nonetheless, this signature of reduced Tajima's D and nucleotide diversity is consistent across all the strongly divergent genomic regions across the speciation event. Future studies could undertake more detailed phenotyping at the individual level to conduct association studies to determine genotype effect on traits of interest. Such individual-based data can be exceptionally valuable to understand evolutionary processes shaping genetic variation (for examples see great tits (*Parus major*; Bosse *et al.* 2017, Seychelles warbler (*Acrocephalus sechellensis*; Davies *et al.* 2021), and the Berthelot's pipit would be a great system to apply association studies across multiple replicate populations.

We detected a potential ~2 Mb 'plateau' of divergence between the tawny and Berthelot's pipit, which is conserved across the Berthelot's pipit range (Fig. 6). Despite the absence of gene flow between the two sister species it is still surprising to detect such a broad peak of elevated divergence with low genetic diversity in both populations, as this suggests a long-term absence of recombination in this region. It is possible that this is due to strong divergent selection, occurring independently in both populations; however, it is much more likely that background selection combined with drift is driving high levels of divergence. Although the most probable explanation for suppressed recombination is an inversion (Rieseberg 2001; Kirkpatrick 2010), long read sequencing (Shao *et al.* 2018) and haplotype phasing (Alachiotis *et al.* 2012; Ferrer-Admetlla *et al.* 2014) need to be used to confirm this. We are aware of one other study - on European great tits (*Parus major*) - to have identified a moderately frequent (~5%) inversion overlapping with this region on chromosome 1A (Da Silva *et al.* 2019). The putative 1A inversion in Berthelot's pipit mapped to a gene dense region with 17 genes, many of which are associated with craniofacial/bone development or the immune response (Table 4). For example, *CMAS*, *TTLL12* and *YBX3* are all associated with viral defence (Carette *et al.* 2009; Ju *et al.* 2017; Qin *et al.* 2020); *TSPO*, *GEC1* and *STYK1* are associated with cellular autophagy or apoptosis (Veenman *et al.* 2007;

Chakrama *et al.* 2010; Zhou *et al.* 2020); *KCNJ8* increases wound healing and regeneration (Zhang and Bei 2015); *ABCC9* is associated with craniofacial defects in vertebrates (Czeschik *et al.* 2013; Harakalova *et al.* 2012); and *SCUBE1* modulates BMP signalling during craniofacial development (Xavier *et al.* 2009; Tu *et al.* 2008) and is linked to head morphology adaptation in Sticklebacks (Hohenlohe *et al.* 2010). Inversions have been found to harbour supergene complexes (Taylor and Campagna 2016) which have been linked to striking plumage and behavioural differences in the ruff (*Calidris pugnax*; Küpper *et al.* 2016) and white-throated sparrow (*Zonotrichia albicollis*; Tuttle *et al.* 2016), marine-freshwater divergence in threespine stickleback (*Gasterosteus aculeatus*; Roesti *et al.* 2015), disease susceptibility in non human primates (Porubsky *et al.* 2020) and highland adaptation in honeybees (*Apis mellifera*; Christmas *et al.* 2019). Such inversion haplotypes may facilitate local adaptation and speciation (Kirkpatrick and Barton 2006).

Pathogen prevalence varies substantially between different Berthelot's pipit populations, and between those and the tawny pipit, providing a strong environmental selection gradient (Illera *et al.* 2008; Spurgin *et al.* 2012; Armstrong *et al.* 2018). Studies of Berthelot's pipit have previously identified pathogen-mediated selection associated with key immune genes including toll-like receptor 4 (TLR4) (Gonzalez-Quevedo *et al.* 2015; Armstrong *et al.* 2019) and MHC class I (Spurgin *et al.* 2011; Gonzalez-Quevedo *et al.* 2016) within islands and across archipelagos (Armstrong *et al.* 2018). In the present study we identified various genes in highly diverged regions that could be involved with host-pathogen evolution including: *ITGA9*, *SENP7* and *NFRKB* – associated with the innate immune response (Sigurðarson 2020; Cui *et al.* 2017; Audard *et al.* 2012); and *WAC* and *GEC1* – which interact in the same pathway to regulate autophagy (Joachim *et al.* 2015). We did not, however, detect high levels of divergence in regions containing genes associated with pathogen response in comparisons involving the most recently founded Berthelot's pipit archipelago, the Selvagens. This is to be expected given the lack of pathogens in this very isolated archipelago.

Variation in bill morphology across the species range has been largely attributed to founder effects across the archipelgos (Spurgin *et al.* 2014; Armstrong *et al.* 2018). However, our findings suggest that selection may also play a role in shaping craniofacial development through each colonisation event (Table 4, Table 5), and areas of the genome putatively under selection included *CIT*, *NRXN1*, *PRDM10* and *WAC* - all associated with craniofacial defects in vertebrates (Shaheen *et al.* 2016; Zahir *et al.* 2008; Park and Kim 2010; De Santo *et al.* 2015), *ABI3BP* - associated with bone development (Zhang *et al.* 2014); and *SOCS7* –

which interacts with growth factors (Elliott and Johnston 2004). These genes may underly variation in bill morphology that enhances exploitation of differing food resources across the range.

Our findings provide evidence of repeated divergence, potentially as a result of selection, for genes previously found to be involved in head/bill and body size, eye development, metabolism, wound healing and immune defence, across different temporal scales in Berthelot's pipit. However, as genomic islands of divergence rarely occurred at the same location across inter-population comparisons, it is possible that different genes with similar phenotypic effects may be under selection. Parallel adaptation for traits through different candidate genes has been reported across vertebrates (Milner et al. 1999; Langin et al. 2015; Walsh et al. 2019). Overall, our findings add to growing evidence of immunity and head/bill and body size as ecologically important traits for adaptive divergence and speciation (cf. immune; Hughes and Yeager 1998; Jarvi *et al.* 2001; Davies *et al.* 2021, and bill, Grant 1968; Badyaev *et al.* 2008; Bosse *et al.* 2017; Lundregan *et al.* 2018; Dussex *et al.* 2021). Our approach to identifying genomic divergence is likely to be conservative. As such we recognise that we will not detect every gene under selection across the pipit system, instead we only identify genes in the most strongly divergent regions.

Between the tawny and Berthelot's pipit, we detected very few strongly conserved genomic regions despite speciation in allopatry similar to observations of diverging *Ficedula* flycatcher species by Burri *et al.* (2015). Genomic valleys of divergence may be associated with reduced F_{ST} and nucleotide diversity if they are due to parallel selection (Roesti *et al.* 2014). Alternatively they may be associated with reduced F_{ST} and moderate nucleotide diversity if due to balancing selection (Hohenlohe *et al.* 2010). In our case we have limited evidence to confirm either of these possibilities. If the genomic regions we have identified are under parallel/balancing selection in the tawny and Berthelot's pipit this highlights the importance of a further set of genes associated with body and head size and immune response (Table 6). Interestingly, these candidate genes include *NFX1*, a transcriptional repressor of MHC class II genes (Strominger *et al.* 1994; Gewin *et al.* 2004), a gene family known to be under balancing selection in many vertebrates (Meyer and Thomson 2001; Savage *et al.* 2020), including avian species (e.g., Alcaide *et al.* 2008; Brouwer *et al.* 2010; Ekblom *et al.* 2010). Evidence of long-term retention of MHC alleles is a well-known phenomenon (Klein 1987), and balancing selection over similar evolutionary timescales has been reported across populations of many taxa (e.g., Richardson and Westerdahl 2003; Bryja *et al.* 2007; Evans *et al.* 2010; Herdegen-Radwan *et al.* 2021). In Berthelot's pipit, island colonisation is initially

associated with reduced MHC diversity but there is evidence of the *in situ* generation of diversity via gene conversion (Spurgin *et al.* 2011).

6 Conclusions

Studying diverging populations along the speciation continuum in a system with a well-characterised demographic history should help develop an understanding of how the genomic landscape of divergence develops through differing drift and selection contexts. We identify different sets of highly differentiated loci through independent colonisation events across the Berthelot's pipit range, suggesting independent genetic responses to selection pressures. Candidate genes identified within these strongly divergent loci appear to be linked to phenotypic changes in body/bill size that we observe across the Berthelot's pipit colonisation range, adaptation to climate, pathogen defense and eye development. Our findings suggest that parallel adaptation through evolution of different loci may be occurring following population founding of new islands. Combining multi-population studies such as ours, with detailed phenotyping and association mapping could provide a powerful approach for understanding selection and evolution in the wild.

7 References

Alachiotis, N., Stamatakis, A. and Pavlidis, P. (2012). OmegaPlus: A scalable tool for rapid detection of selective sweeps in whole-genome datasets. *Bioinformatics* **28**:2274–2275.

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

Alef, T., Torres, S., Hausser, I., Metze, D., Türsen, Ü., Lestringant, G.G. and Hennies, H.C. (2009). Ichthyosis, follicular atrophoderma, and hypotrichosis caused by mutations in ST14 is associated with impaired profilaggrin processing. *Journal of Investigative Dermatology* **129**:862–869.

Alleaume-Benharira, M., Pen, I.R. and Ronce, O. (2006). Geographical patterns of adaptation within a species' range: Interactions between drift and gene flow. *Journal of Evolutionary Biology* **19**:203–215.

Armstrong, C., Davies, R.G., González-Quevedo, C., Dunne, M., Spurgin, L.G. and Richardson, D.S. (2019). Adaptive landscape genetics and malaria across divergent island bird populations. *Ecology and Evolution* **0**:1–21.

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.

Audard, V., Pawlak, A., Candelier, M., Lang, P. and Sahali, D. (2012). Upregulation of nuclear factor-related kappa b suggests a disorder of transcriptional regulation in minimal change nephrotic syndrome. *PLoS ONE* **7**:1–9.

Badyaev, A. V., Young, R.L., Oh, K.P. and Addison, C. (2008). Evolution on a local scale: Developmental, functional, and genetic bases of divergence in bill form and associated changes in song structure between adjacent habitats. *Evolution* **62**:1951–1964.

Barton, N.H. and Charlesworth, B. (1984). Genetic revolutions, founder effects, and speciation. *Annual Review of Ecology and Systematics* **15**:133–164.

Bay, R.A. and Ruegg, K. (2017). Genomic islands of divergence or opportunities for introgression? *Proceedings of the Royal Society B: Biological Sciences* **284**.

Berry, R.J. (1986). Genetics of insular populations of mammals, with particular reference to differentiation and founder effects in British small mammals. *Biological Journal of the Linnean Society* **28**:205–230.

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

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

Brown, T.C., Nicolson, N.G., Stenman, A., Juhlin, C.C., Gibson, C.E., Callender, G.G., Korah, R., *et al.* (2019). Insulin-like growth factor and SLC12A7 dysregulation: A novel signaling hallmark of non-functional adrenocortical carcinoma. *Journal of the American College of Surgeons* **229**:305–315.

Bryja, J., Charbonnel, N., Berthier, K., Galan, M. and Cosson, J.F. (2007). Density-related changes in selection pattern for major histocompatibility complex genes in fluctuating populations of voles. *Molecular Ecology* **16**:5084–5097.

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., *et al.* (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Research* **25**:1656–1665.

Calero-Riestra, M. and García, J.T. (2016). Sex-dependent differences in avian malaria prevalence and consequences of infections on nestling growth and adult condition in the Tawny pipit, *Anthus campestris*. *Malaria Journal* **15**:1–11.

Carette, J.E., Guimaraes, C.P., Varadarajan, M., Park, A.S., Wuethrich, I., Godarova, A., Kotecki, M., *et al.* (2009). Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326**:1231–1235.

Chakrama, F.Z., Seguin-Py, S., Le Grand, J.N., Fraichard, A., Delage-Mourroux, R., Despouy, G., Perez, V., *et al.* (2010). *GABARAPL1 (GEC1)* associates with autophagic vesicles. *Autophagy* **6**:495–505.

Cho, Y. Bin, Hong, S., Kang, K.W., Kang, J.H., Lee, S.M. and Seo, Y.J. (2020). Selective and ATP-competitive kinesin KIF18A inhibitor suppresses the replication of influenza A virus. *Journal of Cellular and Molecular Medicine* **24**:5463–5475.

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

Cousin, H., Gaultier, A., Bleux, C., Darribère, T. and Alfandari, D. (2000). *PACSIN2* is a regulator of the metalloprotease/disintegrin *ADAM13*. *Developmental Biology* **227**:197–210.

Cui, Y., Yu, H., Zheng, X., Peng, R., Wang, Q., Zhou, Y., Wang, R., *et al.* (2017). SENP7 Potentiates cGAS Activation by Relieving SUMO-Mediated Inhibition of Cytosolic DNA Sensing. *PLoS Pathogens* **13**:1–24.

Cvijović, I., Good, B.H. and Desai, M.M. (2018). The effect of strong purifying selection on genetic diversity. *Genetics* **209**:1235–1278.

Czeschik, J.C., Voigt, C., Goecke, T.O., Lüdecke, H.J., Wagner, N., Kuechler, A. and Wieczorek, D. (2013). Wide clinical variability in conditions with coarse facial features and hypertrichosis caused by mutations in *ABCC9*. *American Journal of Medical Genetics, Part A* **161**:295–300.

Dai, F., Lin, X., Chang, C. and Feng, X.H. (2009). Nuclear Export of Smad2 and Smad3 by RanBP3 Facilitates Termination of TGF-β Signaling. *Developmental Cell* **16**:345–357.

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

Delfín, D.A., DeAguero, J.L. and McKown, E.N. (2019). The Extracellular Matrix Protein ABI3BP in Cardiovascular Health and Disease. *Frontiers in Cardiovascular Medicine* **6**:22–26.

Delmore, K.E., Lugo Ramos, J.S., Van Doren, B.M., Lundberg, M., Bensch, S., Irwin, D.E. and Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters* **2**:76–87.

Van Doren, B.M., Campagna, L., Helm, B., Illera, J.C., Lovette, I.J. and Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology* **26**:3982–3997.

Dussex, N., Kutschera, V.E., Wiberg, R.A.W., Parker, D.J., Hunt, G.R., Gray, R.D., Rutherford, K., *et al.* (2021). A genome-wide investigation of adaptive signatures in protein-coding genes related to tool behaviour in New Caledonian and Hawaiian crows. *Molecular Ecology* **30**:973–986.

Ekblom, R., Sæther, S.A., Fiske, P., Kålås, J.A. and Höglund, J. (2010). Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. *Genetica* **138**:453–461.

El-Sayed, W., Parry, D.A., Shore, R.C., Ahmed, M., Jafri, H., Rashid, Y., Al-Bahlani, S., *et al.* (2009). Mutations in the beta propeller WDR72 cause autosomal-recessive hypomaturation amelogenesis imperfecta. *American Journal of Human Genetics* **85**:699–705.

Ellegren, H., Smeds, L., Burri, R., Olason, P.I., Backström, N., Kawakami, T., Künstner, A., *et al.* (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature* **491**:756–760.

Elliott, J. and Johnston, J.A. (2004). SOCS: Role in inflammation, allergy and homeostasis. *Trends in Immunology* **25**:434–440.

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

Fay, J.C. and Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. *Genetics* **155**:1405–1413.

Feder, J.L., Flaxman, S.M., Egan, S.P., Comeault, A.A. and Nosil, P. (2013). Geographic mode of speciation and genomic divergence. *Annual Review of Ecology, Evolution, and Systematics* **44**:73–97.

Feder, J.L. and Nosil, P. (2010). The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* **64**:1729–1747.

Ferrer-Admetlla, A., Liang, M., Korneliussen, T. and Nielsen, R. (2014). On detecting incomplete soft or hard selective sweeps using haplotype structure. *Molecular Biology and Evolution* **31**:1275–1291.

Fontenele EG, Moraes ME, d'Alva CB, Pinheiro DP, Landim SA, Barros FA, Trarbach EB, Mendonca BB, Jorge AA. (2015) Association study of GWAS-derived loci with height in Brazilian children: Importance of *MAP3K3, MMP24* and *IGF1R* polymorphisms for height variation. *Horm Res Paediatr.* **84**:248-53.

Galtier, N., Depaulis, F. and Barton, N.H. (2000). Detecting bottlenecks and selective sweeps from DNA sequence polymorphism. *Genetics* **155**:981–987.

Gao, J., Xu, G. and Xu, P. (2021). Full-length transcriptomic analysis reveals osmoregulatory mechanisms in *Coilia nasus* eyes reared under hypotonic and hyperosmotic stress. *Science of the Total Environment* **799**:149333.

Gewin, L., Myers, H., Kiyono, T. and Galloway, D.A. (2004). Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. *Genes and Development* **18**:2269–2282.

Gingras, R.M., Lwin, K.M., Miller, A.M. and Bretscher, A. (2020). Yeast Rgd3 is a phospho-regulated F-BAR–containing rhogap involved in the regulation of rho3 distribution and cell morphology. *Molecular Biology of the Cell* **31**:2570–2582.

Gonzalez-Quevedo, C., Davies, R.G. and 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.

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

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

Grabherr, M.G., Russell, P., Meyer, M., Mauceli, E., Alföldi, J., di Palma, F. and Lindblad-Toh, K. (2010). Genome-wide synteny through highly sensitive sequence alignment: Satsuma. *Bioinformatics* **26**:1145–1151.

Grant, P.R. (1968). Bill size, body size, and the ecological adaptations of bird species to competitive situations on islands. *Systematic Biology* **17**:319–333.

Gupta, K., Janardan, V., Banerjee, S., Chakrabarti, S., Srinivas, S., Mahishi, D., Raghu, P., *et al.* (2020). An evolutionarily conserved metallophosphodiesterase is a determinant of lifespan in Drosophila. *bioRxiv*.

Gupta, S.K. and Vlahakis, N.E. (2010). Integrin α9β1: Unique signaling pathways reveal diverse biological roles. *Cell Adhesion and Migration* **4**:194–198.

Han, P., Werber, J., Surana, M., Fleischer, N. and Michaeli, T. (1999). The calcium/calmodulin-dependent phosphodiesterase PDE1C down-regulates glucose-induced insulin secretion. *Journal of Biological Chemistry* **274**:22337–22344.

Harakalova, M., Van Harssel, J.J.T., Terhal, P.A., Van Lieshout, S., Duran, K., Renkens, I., Amor, D.J., *et al.* (2012). Dominant missense mutations in ABCC9 cause Cantú syndrome. *Nature Genetics* **44**:793–796.

Harris, Z.L. (2018). *Ceruloplasmin*. Elsevier Inc.

Harrison, R.G. (1991). Molecular changes at speciation. *Annual Review of Ecology and Systematics* **22**:281–308.

Herdegen-Radwan, M., Phillips, K.P., Babik, W., Mohammed, R.S. and Radwan, J. (2021). Balancing selection versus allele and supertype turnover in MHC class II genes in guppies. *Heredity* **126**:548–560.

Hofer, T., Foll, M. and Excoffier, L. (2012). Evolutionary forces shaping genomic islands of population differentiation in humans. *BMC Genomics* **13**:1–13.

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

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

Irwin, D.E., Milá, B., Toews, D.P.L., Brelsford, A., Kenyon, H.L., Porter, A.N., Grossen, C., *et al.* (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology* **27**:4839–4855.

Jang, S. and Javadov, S. (2018). Elucidating the contribution of ETC complexes I and II to the respirasome formation in cardiac mitochondria. *Scientific Reports* **8**:1–12.

Jarvi, S.I., Atkinson, C.T. and Fleischer, R.C. (2001). Immunogenetics and resistance to avian malaria in Hawaiian Honeycreepers (*Drepanidinae*). *Studies in Avian Biology No* **22**:254–263.

Jiang, L., Sun, Y., Wang, J., He, Q., Chen, X., Lan, X., Chen, J., *et al.* (2019). Proteasomal cysteine deubiquitinase inhibitor b-AP15 suppresses migration and induces apoptosis in diffuse large B cell lymphoma. *Journal of Experimental and Clinical Cancer Research* **38**:1–14.

Joachim, J., Jefferies, H.B.J., Razi, M., Frith, D., Snijders, A.P., Chakravarty, P., Judith, D., *et al.* (2015). Activation of ULK Kinase and autophagy by *GABARAP* trafficking from the centrosome is regulated by WAC and GM130. *Molecular Cell* **60**:899–913.

de Jong, M.J., Li, Z., Qin, Y., Quéméré, E., Baker, K., Wang, W. and Hoelzel, A.R. (2020). Demography and adaptation promoting evolutionary transitions in a mammalian genus that diversified during the Pleistocene. *Molecular Ecology* **29**:2777–2792.

Ju, L.-G., Zhu, Y., Lei, P.-J., Yan, D., Zhu, K., Wang, X., Li, Q.-L., *et al.* (2017). TTLL12 inhibits the activation of cellular antiviral signaling through interaction with VISA/MAVS. *The Journal of Immunology* **198**:1274–1284.

Kim, K., Perroud, B., Espinal, G., Kachinskas, D., Austrheim-Smith, I., Wolfe, B.M. and Warden, C.H. (2008). Genes and networks expressed in perioperative omental adipose tissue are correlated with weight loss from Roux-en-Y gastric bypass. *International Journal of Obesity* **32**:1395–1406.

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

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology* **8**.

Kirkpatrick, M. and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434.

Kirkpatrick, M. and Ravigné, V. (2002). Speciation by natural and sexual selection: Models and experiments. *American Naturalist* **159**.

Klein, J. (1987). Origin of major histocompatibility complex polymorphism: The trans-species hypothesis. *Human Immunology* **19**:155–162.

Küpper, C., Stocks, M., Risse, J.E., Dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., *et al.* (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics* **48**:79–83.

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

Langin, K.M., Sillett, T.S., Funk, W.C., Morrison, S.A., Desrosiers, M.A. and Ghalambor, C.K. (2015). Islands within an island: Repeated adaptive divergence in a single population. *Evolution* **69**:653–665.

Le Moan, A., Gaggiotti, O., Henriques, R., Martinez, P., Bekkevold, D. and Hemmer-Hansen, J. (2019). Beyond parallel evolution: When several species colonize the same environmental gradient. *bioRxiv*.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **0**:1–3.

Li, J.W., Yeung, C.K.L., Tsai, P.W., Lin, R.C., Yeh, C.F., Yao, C. Te, Han, L., *et al.* (2010). Rejecting strictly allopatric speciation on a continental island: Prolonged postdivergence gene flow between Taiwan (*Leucodioptron taewanus, Passeriformes Timaliidae*) and Chinese (*L. canorum canorum*) hwameis. *Molecular Ecology* **19**:494–507.

Liao, W., Atkinson, C.T., LaPointe, D.A. and Samuel, M.D. (2017). Mitigating future avian malaria threats to Hawaiian forest birds from climate change. *PLoS ONE* **12**:1–25.

Lindtke, D., Lucek, K., Soria-Carrasco, V., Villoutreix, R., Farkas, T.E., Riesch, R., Dennis, S.R., *et al.* (2017). Long-term balancing selection on chromosomal variants associated with crypsis in a stick insect. *Molecular Ecology* **26**:6189–6205.

Liu, C., Chen, G., Bentley, A.R., Doumatey, A., Zhou, J., Adeyemo, A., Yang, J., *et al.* (2019). Genome-wide association study for proliferative diabetic retinopathy in Africans. *npj Genomic Medicine* **4**:1–7.

Liu, T., Yuan, Y., Bai, D., Qi, Z., Yang, L., Zhang, T., Yang, X., *et al.* (2021). Genetic variants and mutational spectrum of Chinese Hermansky–Pudlak syndrome patients. *Pigment Cell and Melanoma Research* **34**:111–121.

Lu, J., Gu, Y., Feng, J., Zhou, W., Yang, X. and Shen, Y. (2014). Structural insight into the central element assembly of the synaptonemal complex. *Scientific Reports* **4**:1–6.

Lu, X., Fan, Y., Li, M., Chang, X. and Qian, J. (2021). HTR2B and SLC5A3 are specific markers in age-related osteoarthritis and involved in apoptosis and inflammation of osteoarthritis synovial cells. *Frontiers in Molecular Biosciences* **8**:1–13.

Lundregan, S.L., Hagen, I.J., Gohli, J., Niskanen, A.K., Kemppainen, P., Ringsby, T.H., Kvalnes, T., *et al.* (2018). Inferences of genetic architecture of bill morphology in house sparrow using a high-density SNP array point to a polygenic basis. *Molecular Ecology* **27**:3498–3514.

Mandard, S., Stienstra, R., Escher, P., Tan, N.S., Kim, I., Gonzalez, F.J., Wahli, W., *et al.* (2007). Glycogen synthase 2 is a novel target gene of peroxisome proliferator-activated receptors. *Cellular and Molecular Life Sciences* **64**:1145–1157.

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

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., *et al.* (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**:1817–1828.

Martin, A. and Lorenzo, J.A. (2001). Aves del Archipielago Canario.

Maynard-Smith, J. and Haigh, J. (1974). The hitch-hiking effect of a favourable gene. *Genetical Research* **23**:23–35.

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. *Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning* **20**:1297–1303.

Meier, J.I., Marques, D.A., Wagner, C.E., Excoffier, L. and Seehausen, O. (2018). Genomics of parallel ecological speciation in Lake Victoria cichlids. *Molecular Biology and Evolution* **35**:1489–1506.

Meyer, D. and Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: A review. *Annals of Human Genetics* **1996**:1–26.

Milner, J.M., Albon, S.D., Illius, A.W., Pemberton, J.M. and Clutton-Brock, T.H. (1999). Repeated selection of morphometric traits in the Soay sheep on St Kilda. *Journal of Animal Ecology* **68**:472–488.

Munch, K., Nam, K., Schierup, M.H. and Mailund, T. (2016). Selective sweeps across twenty millions years of primate evolution. *Molecular Biology and Evolution* **33**:3065–3074.

Nei, M. (2005). Selectionism and neutralism in molecular evolution. *Molecular Biology and Evolution* **22**:2318–2342.

Nielsen, R. (2005). Molecular Signatures of Natural Selection. *Annual Review of Genetics* **39**:197–218.

Noor, M.A.F. and Bennett, S.M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity* **103**:439–444.

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

O'day, E.M., Idos, G.E., Hill, C., Chen, J.W. and Wagner, G. (2018). Cytidine monophosphate N-acetylneuraminic acid synthetase enhances invasion of human triple-negative breast cancer cells. *OncoTargets and Therapy* **11**:6827–6838.

Oliveira, P., and D. Menezes. 2004. Birds of the archipelago of Madeira. Servico do Parque Natural deMadeira, Madeira.

Osmond, M.M. and Coop, G. (2020). Genetic signatures of evolutionary rescue by a selective sweep. *Genetics* **215**:813–829.

Park, J.A. and Kim, K.C. (2010). Expression patterns of *PRDM10* during mouse embryonic development. *BMB Reports* **43**:29–33.

Qin, Z., Qu, X., Lei, L., Xu, L. and Pan, Z. (2020). Y-box-binding protein 3 (*YBX3*) restricts influenza A virus by interacting with viral ribonucleoprotein complex and imparing its function. *Journal of General Virology* **101**:385–398.

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.

Ravinet, M., Yoshida, K., Shigenobu, S., Toyoda, A., Fujiyama, A. and Kitano, J. (2017). *The genomic landscape at a late stage of stickleback speciation: High genomic divergence interspersed by small localized regions of introgression*.

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

Rieseberg, L.H. (2001). Chromosomal rearrangements and speciation. *TRENDS in Ecology & Evolution* **16**:351–358.

Roesti, M., Gavrilets, S., Hendry, A.P., Salzburger, W. and Berner, D. (2014). The genomic signature of parallel adaptation from shared genetic variation. *Molecular Ecology* **23**:3944–3956.

Roesti, M., Hendry, A.P., Salzburger, W. and Berner, D. (2012). Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology* **21**:2852–2862.

Roesti, M., Kueng, B., Moser, D. and Berner, D. (2015). The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications* **6**.

Rönn, T., Poulsen, P., Tuomi, T., Isomaa, B., Groop, L., Vaag, A. and Ling, C. (2009). Genetic variation in ATP5O is associated with skeletal muscle ATP50 mRNA expression and glucose uptake in young twins. *PLoS ONE* **4**:1–6.

Ruegg, K., Anderson, E.C., Boone, J., Pouls, J. and Smith, T.B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology* **23**:4757–4769.

Sahoo, D., Swanson, L., Sayed, I.M., Katkar, G.D., Ibeawuchi, S.R., Mittal, Y., Pranadinata, R.F., *et al.* (2021). Artificial intelligence guided discovery of a barrier-protective therapy in inflammatory bowel disease. *Nature Communications* **12**:1–14.

De Santo, C., D'Aco, K., Araujo, G.C., Shannon, N., Study, D.D.D., Vernon, H., Rahrig, A., *et al.* (2015). *WAC* loss-of-function mutations cause a recognisable syndrome characterised by dysmorphic features, developmental delay and hypotonia and recapitulate 10p11.23 microdeletion syndrome. *Journal of Medical Genetics* **52**:754–761.

Savage, A.E., Gratwicke, B., Hope, K., Bronikowski, E. and Fleischer, R.C. (2020). Sustained immune activation is associated with susceptibility to the amphibian chytrid fungus. *Molecular Ecology*:2889–2903.

Schneider, K., White, T.J., Mitchell, S., Adams, C.E., Reeve, R. and Elmer, K.R. (2021). The pitfalls and virtues of population genetic summary statistics: Detecting selective sweeps in recent divergences. *Journal of Evolutionary Biology* **34**:893–909.

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., 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: Genes, Genomes, Genetics* **10**:3147–3163.

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

Shaheen, R., Hashem, A., Abdel-Salam, G.M.H., Al-Fadhli, F., Ewida, N. and Alkuraya, F.S. (2016). Mutations in CIT, encoding citron rho-interacting serine/threonine kinase, cause severe primary microcephaly in humans. *Human Genetics* **135**:1191–1197.

Shao, H., Ganesamoorthy, D., Duarte, T., Cao, M.D., Hoggart, C.J. and Coin, L.J.M. (2018). npInv: Accurate detection and genotyping of inversions using long read sub-alignment. *BMC Bioinformatics* **19**:1–13.

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

Sigurðarson, A. (2020). *Dual RNA-Seq Analysis of Host-Pathogen Interaction in Eimeria Infection of Chickens*.

Stajich, J.E. and Hahn, M.W. (2005). Disentangling the effects of demography and selection in human history. *Molecular Biology and Evolution* **22**:63–73.

Stankowski, S., Chase, M.A., Fuiten, A.M., Rodrigues, M.F., Ralph, P.L. and Streisfeld, M.A. (2019). Widespread selection and gene flow shape the genomic landscape during a radiation of Monkeyflowers. *PLoS Biology* **17**:1-31.

Da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., Van Oers, K., Dibbits, B., *et al.* (2019). The genomic complexity of a large inversion in Great Tits. *Genome Biology and Evolution* **11**:1870–1881.

Soria-Carrasco, V., Gompert, Z., Comeault, A., Farkas, T.E., Parchman, T.L., Johnston, J.S., Buerkle, A., *et al.* (2014). Stick Insect genomes reveal natural selection's role in parallel speciation. *Science* **344**:177–190.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. and 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. and 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. and Richardson, D.S. (2011). Gene conversion rapidly generates major histocompatibility complex diversity in recently founded bird populations. *Molecular Ecology* **20**:5213–5225.

Strominger, J.L., Johns, T. and Song, Z. (1994). A novel cysteine-rich sequence-specific DNA-binding protein interacts with the conserved X-box motif of the human major histocompatibility complex class II genes via a repeated Cys-His domain and functions as a transcriptional repressor. *Journal of Experimental Medicine* **180**.

Taylor, S. and Campagna, L. (2016). Avian supergenes. *Science*:10–11.

Tetteh, K.K.A., Stewart, L.B., Ochola, L.I., Amambua-Ngwa, A., Thomas, A.W., Marsh, K., Weedall, G.D., *et al.* (2009). Prospective identification of malaria parasite genes under balancing selection. *PLoS ONE* **4**.

Tomozawa, M., Nunome, M., Suzuki, H. and Ono, H. (2014). Effect of founding events on coat colour polymorphism of Apodemus speciosus (*Rodentia: Muridae*) on the Izu islands. *Biological Journal of the Linnean Society* **113**:522–535.

Tu, C.F., Yan, Y.T., Wu, S.Y., Djoko, B., Tsai, M.T., Cheng, C.J. and Yang, R.B. (2008). Domain and functional analysis of a novel platelet-endothelial cell surface protein, SCUBE. *Journal of Biological Chemistry* **283**:12478–12488.

Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., *et al.* (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology* **26**:344–350.

Urbanek, K., Sutherland, D.M., Orchard, R.C., Wilen, C.B., Knowlton, J.J., Aravamudhan, P., Taylor, G.M., *et al.* (2020). Cytidine Monophosphate N -Acetylneuraminic Acid Synthetase and Solute Carrier Family 35 Member A1 are required for reovirus binding and infection. *Journal of Virology* **95**.

Veenman, L., Papadopoulos, V. and Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (*TSPO*): regulation of apoptosis and steroidogenesis as part of the host-defense response. *Current pharmaceutical design* **13**:2385–405.

Via, S. and West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology* **17**:4334–4345.

Vijay, N., Ellegren, H. and Wolf, J.B.W. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. :4284–4295.

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

Walsh, J., Benham, P.M., Deane-Coe, P.E., Arcese, P., Butcher, B.G., Chan, Y.L., Cheviron, Z.A., *et al.* (2019). Genomics of rapid ecological divergence and parallel adaptation in four tidal marsh sparrows. *Evolution Letters* **3**:324–338.

Weiss-Lehman, C., Tittes, S., Kane, N.C., Hufbauer, R.A. and Melbourne, B.A. (2019). Stochastic processes drive rapid genomic divergence during experimental range expansions. *Proceedings of the Royal Society B: Biological Sciences* **286**.

Wolf, J.B.W. and Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics* **18**:87–100.

Xavier, G.M., Sharpe, P.T. and Cobourne, M.T. (2009). Scube1 is expressed during facial development in the mouse. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **312**:518–524.

Xu, H., Qu, C., Gan, L., Sun, K., Tan, J., Liu, X., Jiang, Z., *et al.* (2020). Deletion of the Impg2 gene causes the degeneration of rod and cone cells in mice. *Human Molecular Genetics* **29**:1624–1634.

Yan, J. and Zhu, H.B. (2007). Cyclic nucleotide phosphodiesterase IV expression, activity and targeting in cells of cardiovascular system. *Yaoxue Xuebao* **42**:571–575.

Yan, W., Long, P., Chen, T., Liu, W., Yao, L., Ren, Z., Li, X., *et al.* (2018). A natural occurring mouse model with adgrv1 mutation of usher syndrome 2C and characterization of its recombinant inbred strains. *Cellular Physiology and Biochemistry* **47**:1883–1897.

Ben Yaou, R., Hubert, A., Nelson, I., Dahlqvist, J.R., Gaist, D., Streichenberger, N., Beuvin, M., *et al.* (2017). Clinical heterogeneity and phenotype/genotype findings in 5 families with *GYG1* deficiency. *Neurology Genetics* **3**:e208.

Yousaf, R., Gu, C., Ahmed, Z.M., Khan, S.N., Friedman, T.B., Riazuddin, S., Shears, S.B., *et al.* (2018). Mutations in Diphosphoinositol-Pentakisphosphate Kinase *PPIP5K2* are associated with hearing loss in human and mouse. *PLoS Genetics* **14**:1–20.

Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.A., *et al.* (2008). A patient with vertebral, cognitive and behavioural abnormalities and a *de novo* deletion of NRXN1α. *Journal of Medical Genetics* **45**:239–243.

Zhang, F., Guo, X., Zhang, Y., Wen, Y., Wang, W., Wang, S., Yang, T., *et al.* (2014). Genome-wide copy number variation study and gene expression analysis identify *ABI3BP* as a susceptibility gene for Kashin-Beck disease. *Human Genetics* **133**:793–799.

Zhang, W. and Bei, M. (2015). Kcnh2 and Kcnj8 interactively regulate skin wound healing and regeneration. *Wound Repair and Regeneration* **23**:797–806.

Zhang, W., Das, P., Kelangi, S. and Bei, M. (2020). Potassium channels as potential drug targets for limb wound repair and regeneration. *Precision Clinical Medicine* **3**:22–33.

Zhou, C., Qian, X., Hu, M., Zhang, R., Liu, N., Huang, Y., Yang, J., *et al.* (2020). STYK1 promotes autophagy through enhancing the assembly of autophagy-specific class III phosphatidylinositol 3-kinase complex I. *Autophagy* **16**:1786–1806.

8 Supplementary material

Table S1. Prevalence of blood pathogen infection in 13 populations of the Berthelot's pipit across Macaronesia, with populations sampled in Chapters 3 and 4 in bold. Bracketed values are uncertain which of two islands they originate from.

Archipelago	Island	*Plasmodium*			*Leucocytozoon*			Pox			Sample size		
		2006	2009	2010 -	2006	2009	2010 -	2006	2009	2010 -	2006	2009	2010 -
Madeira	Deserta Grande	0	0	-	0	0	-	0	0	-	31	4	-
	Madeira	0	0	3	0	0	?	0	0	0	33	29	29
	Porto Santo	65	30	75	26	13	?	45	37	28	31	30	129
Selvagens	**Selvagem Grande**	0	0	-	0	0	-	0	0	-	34	42	-
Canary Islands	La Graciosa	4.2	0	-	0	0	-	0	0	-	24	26	-
	Lanzarote	23	48	72	0	0	?	54	16	17	13	31	196
	Fuerteventura	50	45	-	0	0	-	17	29	-	12	31	-
	Gran Canaria	45	15	-	7	0	-	16	27	-	31	33	-
	El Teide	7	0	0	0	0	0	0	4	0	30	25	31
	Tenerife	9	32	27/ 41	3	0	1/ 4	13	6	9/ 5	32	34	454
	La Gomera	53	35	-	0	0	-	3	10	-	30	20	-
	La Palma	4	0	-	0	0	-	0	0	-	28	22	-
	El Hierro	10	0	-	0	0	-	0	0	-	31	30	-

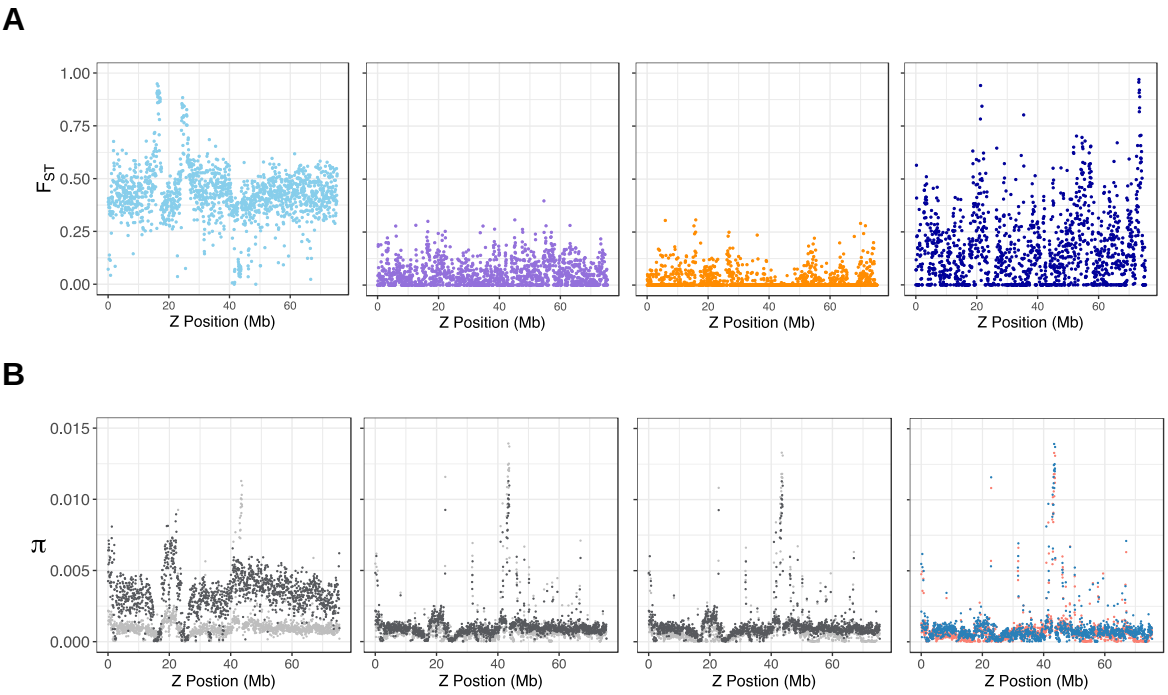


Figure S1. A) Pairwise F_{ST} and **B)** nucleotide diversity across the Z chromosome for each population comparison calculated in non-overlapping 50 kb windows, relative to the Zebra finch genome. Population comparison for F_{ST}: light blue = Tawny vs. Canary Islands, Canary Islands vs. Madeira = purple, Canary Islands vs. Selvagens = orange and Selvagens vs. Madeira = navy. Patterns of nucleotide diversity are plotted for both populations within each comparison; dark grey represents the source population and light grey, the newly founded population (excluding Selvagens vs. Madeira, separated by two independent colonisation events, coloured blue and pink, respectively).

5

General discussion and future directions

Berthelot’s pipit on the coast in northern Lanzarote, Canary Islands

1 General discussion

Evolutionary mechanisms shape patterns of genetic diversity and divergence following the establishment of new populations, which may lead to eventual speciation (Ravinet *et al.* 2017). Disentangling the relative role of natural selection and genetic drift in shaping genetic variation is fundamental to understanding how wild populations evolve and new species are formed, which may be particularly important in a rapidly changing natural world. In this thesis, I initially set out to explore genomic signatures of adaptation across spatial scales in Berthelot's pipit (*Anthus berthelotii*), an island endemic that speciated within the last 2.5 million years and then spread across its current range (Voelker 1999). To develop an understanding of which factors shape genetic diversity across populations, we must first have in-depth knowledge of their population history, including colonisation events and past and present demography. Then, using large genomic marker sets, it may be possible to disentangle the relative roles of drift and selection acting over evolutionary time and across spatial scales. Here, I have used complementary genomic datasets - restriction-site associated DNA sequencing (RAD-seq) and whole genome resequencing - to maximise the sampling density (number of individuals and populations) and the genomic marker density (from fewer individuals and populations) respectively. In this final chapter, I discuss my collective findings and what I believe to be the important conclusions, and outline avenues for further research.

1.1 Colonisation history, demography and drift

Establishing the demographic history of Berthelot's pipit, including colonisation events, subsequent gene flow, contemporary and historic fluctuations in population size and structure within and between populations, was an important first step for determining the evolutionary dynamics at play across the species' populations. In Chapter 2, I used reduced representation RAD-seq to characterise population-level genetic diversity and the pattern of colonisation across the range of Berthelot's pipit. Using demographic trees, this chapter documented the sequence of colonisation, initially from the African mainland tawny pipit (*Anthus campestris*) to the Canary Islands, and the subsequent independent colonisation of both the Madeiran and Selvagens archipelagos separately from the Canary Islands. These findings are in line with past inferences of Berthelot's pipit colonisation history (Illera *et al.* 2007; Spurgin *et al.* 2014; Armstrong *et al.* 2018) and are supported by population-level

models of fluctuating effective population size (N_e) through time (Chapter 3). Consistent with other studies of sequentially founded archipelago populations across a range of taxa (Goodman *et al.* 2001; Jordan and Snell 2008; Sendell-Price *et al.* 2021), my results show strong drift has occurred between the three archipelago populations of the Berthelot's pipit, and limited evidence of gene flow among these populations. Population-level linkage disequilibrium analyses confirmed the contemporary genetic signatures of severe bottlenecks in the latter colonised archipelagos, with similar patterns between separate populations within archipelagos (Spurgin *et al.* 2014). Indeed, the distribution of genomic divergence between population pairs of Berthelot's pipit across archipelagos and between the tawny pipit presented in Chapter 4, provides further support for lack of gene flow between archipelagos. Such patterns are similar to those reported in other animals evolving in allopatry (Martin *et al.* 2013; Han *et al.* 2017; Sendell-Price *et al.* 2020). In contrast to expectations, and to other studies of Macaronesian birds (Trumpeter finch (*Bucanetes githagineus*; Barrientos *et al.* 2009), common chaffinch (*Fringilla coelebs*; Rodrigues *et al.* 2014) and common kestrels (*Falco tinnunculus*; Kangas *et al.* 2018)), results from Chapter 2 provide evidence of very weak gene flow across populations within archipelagos after colonisation, irrespective of the geographic distance between islands. However, during the early stages of divergence it may be difficult to distinguish low genome-wide divergence due to short timescales since dispersal from gene flow countering divergence (reviewed in Ravinet *et al.* 2017).

One approach to investigating the role of contemporary and ancient population processes in shaping patterns of genetic diversity is to study the landscape of diversity across the genome. In Chapter 3, I used whole genome resequencing across the three archipelago level populations of Berthelot's pipit and its sister species the tawny pipit, to investigate genomic variation and model ancient population contractions and expansions, and inbreeding. The results confirm loss of genetic diversity from the African mainland tawny pipit to the island Berthelot's populations, with PSMC modelling of N_e supporting previous inferences of initial colonisation ~2 Mya (Chapter 3). These models further suggest colonisation of the Madeiran and Selvagens islands may have occurred at different time points, firstly to the Madeiran archipelago ~50 kya, with a more recent colonisation and bottleneck in Selvagens <10 kya. Although recent estimates of N_e based on few individuals may be inaccurate, these results are supported by genome-wide patterns of structure (PCA) and divergence (F_{ST}) between archipelagos (Chapter 3).

Colonisation bottlenecks and subsequent constraints on population size appear to have played an important role in shaping genetic diversity across Berthelot's pipit genomes.

Extensive homozygous genome segments (runs of homozygosity, ROH) typically result from drastic population contractions and inbreeding (Escoda and Castresana 2021; Robinson *et al.* 2021), with longer more recently originating ROH likely to harbour more deleterious variation (Szpiech *et al.* 2013; Stoffel *et al.* 2021) and cause inbreeding depression (Kardos *et al.* 2017). By analysing whole genome patterns of diversity (Chapter 3) ROH were revealed extending >1 Mb across the genomes of individuals within the sequentially bottlenecked populations on islands across the Madeiran and Selvagens archipelagos. The length and frequency of ROH is consistent with recent strong inbreeding and small N_e in Selvagens, and more distant or less extreme inbreeding in Madeira. This is in concordance with our understanding of colonisation history, with strong bottlenecks and constrained population size due to limited geographic range and isolation of the Selvagens. I provide some evidence in Chapter 4 that several ROH are associated with selective sweeps in Selvagens, but the extent of ROH coverage (up to 38% of the autosomal genome), mainly reflects population history. Such extensive coverage of the genome with ROH has been reported in a Scottish population of killer whale (*Orcinus orca*) as a result of bottlenecks during range expansion followed by extreme long-term inbreeding due to small population size and isolation (Foote *et al.* 2021). These signatures, combined with known colonisation history of Berthelot's pipit across archipelagos, suggests long-term loss of genetic diversity is due to founder effects and high inbreeding within sequentially colonised, small Berthelot's pipit populations (Chapter 3). Thus, genome-wide diversity in Berthelot's pipit is predominantly shaped founder effects and inbreeding, and populations evolve largely independently following founding.

1.2 Selection, adaptation and spatial scales

In addition to drift, natural selection may also play an important role in shaping genetic and phenotypic divergence across the Berthelot's pipit range. By analysing patterns of divergence among populations, genes potentially important for adaptive evolution may be revealed. In the second part of Chapter 2 I utilise the early-stage divergence between recently separated (and subsequently isolated) populations within archipelagos (i.e., between populations within the Madeiran archipelago and within the Canary Islands) to identify divergent loci. If signatures of selection are eroded over time due to subsequent evolutionary forces (Lenormand 2002; Tigano and Friesen, 2016), then such loci may have gone undetected in previous studies of divergent selection between archipelagos in this system (Armstrong *et al.* 2018). I build on previous genome scan analyses at broad scales

and identify divergent loci putatively under differential selection among these recently separated populations. In the RAD-seq dataset two such outlier loci map to a gene, *ADAM12*, associated with growth and bone development (Tokumasu *et al.* 2016; Yoshida and Yáñez 2021). Furthermore, genotypes at these loci correlate with phenotypic variation in head length across the Berthelot's pipit range (Chapter 2). We detect a relatively small number of outlier loci between populations within archipelagos, compared to the greater number detected using genome scans at broad scales (Armstrong *et al.* 2018) using an identical methodological approach. If populations within Berthelot's pipit archipelagos are separated by short timescales with more recent gene flow, compared to between archipelagos, then divergence due to selection accumulates over time (cf. Meier *et al.* 2011; Safran *et al.* 2016). Using the same RAD-seq dataset allows direct comparison of loci detected across different spatial scales, but loci coverage represents only a fraction of the genome. It is therefore likely that this approach does not detect many of the loci under selection in this system, which may be improved by using whole genomes.

Having established that signatures of selection are strongest between archipelagos (Armstrong *et al.* 2018; Chapter 2), the focus of Chapter 4 was to use whole genomes to determine the genomic landscape of divergence at broad spatial scales between Berthelot's pipit archipelagos. I also utilised the tawny pipit genome to characterise the nature of divergence through the initial island colonisation from the mainland by the ancestor, and subsequent speciation. In support of findings in Chapter 2, it was clear that genomic differentiation accumulates over time mainly due to founder effects and subsequent drift. However, between population comparisons a small proportion of the genome exhibited more rapid divergence, with the distribution of divergence shifting towards higher F_{ST} values from early- (recently founded) to late-stage divergence.

Divergence comparisons between populations identified between 9 and 22 strongly divergent genomic regions, the majority of which included potential candidate genes. Between the tawny and Berthelot's pipit, the most strongly divergent region was a ~2 Mb 'plateau' found on chromosome 1A, which may have resulted from an inversion through speciation. This region, which is conserved across the Berthelot's pipit range, contained a high density of candidate genes associated with craniofacial development (Cousin *et al.* 2000; Xavier *et al.* 2009; Harakalova *et al.* 2012) and immune response (Veenman *et al.* 2007; Chakrama *et al.* 2010; Zhang and Bei 2015; Ju *et al.* 2017; Gupta *et al.* 2020; Urbanek *et al.* 2020; Qin *et al.* 2020; Zhou *et al.* 2020). Inversions may result in evolution of co-adapted alleles or 'supergenes' (Kirkpatrick and Barton 2006) and have been found to

result in major morphological and behavioural polymorphisms in avian studies (Zinzow-Kramer *et al.* 2015; Küpper *et al.* 2016).

Other strongly divergent regions between Berthelot's pipit population comparisons mapped to small genomic regions with sharp peaks of differentiation (Chapter 4). Despite identifying mostly unique genomic islands of divergence through speciation and across the Berthelot's pipit colonisation range, these results indicate repeated divergence for the same traits across different geographic scales and timeframes (ca. ~2 Mya – 8 kya), following population founding. Parallel evolution is common across wild systems, although repeated divergence for traits sometimes occurs via the same mutations at specific loci (cf. Protas *et al.* 2006; Baxter *et al.* 2010; Yassin *et al.* 2016; McCulloch *et al.* 2021; Zhang *et al.* 2021) or via differentiation across many different loci controlling the same trait (Elmer *et al.* 2014; Fang *et al.* 2020; Chen and Chiang 2021).

Natural selection has been found to be a primary driver of bill shape variation across avian populations, linking genetic and phenotypic variation (Lamichhaney *et al.* 2015; Bosse *et al.* 2017). In Chapter 4, I repeatedly detect strongly divergent loci associated with craniofacial development through archipelago colonisation (i.e., tawny vs. Canary Islands (Park and Kim 2010), Canary Islands vs. Madeira (Zahir *et al.* 2008) and Canary Islands vs. Selvagens (De Santo *et al.* 2015; Shaheen *et al.* 2016)). A further set of strongly divergent loci are associated with bone development and/or growth (Kim *et al.* 2008; Elliott and Johnston 2004; Zhang *et al.* 2014), which may more broadly impact body size (as well as craniofacial growth), a key trait known to be divergent between mainland and island populations (Case and Schwaner 1993; White and Searle 2007). On islands, dispersal may be biased for increased body size and larger more robust bills than their mainland conspecifics (Clegg and Owens 2002; Leisler and Winkler 2015), and it is possible this occurs repeatedly through sequential population founding in Berthelot's pipit.

Pathogens also exert strong selective pressure and fitness costs on wild populations (Jarvi *et al.* 2001; Acevedo-Whitehouse and Cunningham 2006; Wang *et al.* 2017) including Berthelot's pipit populations (Gonzalez-Quevedo *et al.* 2014; Armstrong *et al.* 2018). I detect different sets of strongly divergent loci associated with immune response between Berthelot's pipit and the tawny pipit, and between the Canary Islands and Madeiran archipelago populations of Berthelot's pipit (Chapter 4). The identified loci may be useful novel candidates for the study of host-pathogen response. Interestingly, no loci associated with immune response were detected within strongly divergent genomic regions in

comparisons involving the isolated Selvagens archipelago, where there is no avian pox or malaria (Illera *et al.* 2008; Spurgin *et al.* 2012).

My analysis of genomic islands of divergence across the colonisation range (Chapter 4) also builds on previous evidence (Armstrong *et al.* 2018) for selection of lipid and carbohydrate metabolism (Han *et al.* 1999; Elliott and Johnston 2004; Kim *et al.* 2008). Divergent selection for genes related to metabolic processes have been related to latitudinal gradients, potentially due to climate adaptation, across a range of species (Pujolar *et al.* 2014; Zhang *et al.* 2019; Wilder *et al.* 2020). Further sets of strongly divergent loci map to genes related to wound healing and stress response repeatedly following population founding across the range (Alef *et al.* 2009; Gupta and Vlahakis 2010; Zhang and Bei 2015). Through each founder event loci in genes associated with eye development are also identified (Liu *et al.* 2019; Xu *et al.* 2020). Eye development has not previously been studied in the Berthelot's pipit, but is known to be under adaptive evolution across a wide range of taxa (Yokoyama and Yokoyama 1996; Frentiu *et al.* 2007). Environmental drivers of divergence across these loci are unknown in the pipit, but it is possible lesions caused by *Avipoxvirus* (pox virus) may be linked to selection related to healing, while climatic and habitat variation as well as pathogens may be associated with stress response.

Through combining a range of population genetic analysis, I have also been able to make inferences about the nature of selection within and between populations. In Chapter 4, I investigate patterns of nucleotide diversity and Tajima's D within strongly divergent regions. When combined with an understanding of population history, these suggest that differentiation occurs primarily due to selective sweeps in the Canary Islands and Selvagens and balancing selection and/or recent inbreeding in Madeira, although more detailed sequencing and phenotyping are needed to confirm this. My results also suggest that a few regions of the genome exhibit exceptionally low differentiation over long-evolutionary timescales through late-stage divergence between species, potentially due to parallel or balancing selection (Chapter 4). These strongly conserved regions included loci in a gene known to regulate the major histocompatibility complex (MHC) class II (Gewin *et al.* 2004), a key immune family involved in host-pathogen coevolution, and genes known to be associated with craniofacial and bill development including bone morphogenic proteins (BMP) (Tu *et al.* 2008; Dai *et al.* 2009; Xavier *et al.* 2009). Although we cannot rule out the possibility that population founder-induced drift has aided shifts in evolutionary important traits, these findings suggest a role for selective processes across population comparisons within this species.

2 Future directions

The findings of this thesis have emphasised many avenues for future research. An obvious first avenue is to extend whole genome sequencing across more individuals and populations across the Berthelot's pipit range (Szarmach *et al.* 2021). Rapidly decreasing costs of whole genome sequencing, paired with dramatically improved base call accuracy of long read sequencing such as PacBio HiFi, make it financially viable now to resequence genomes for hundreds of individuals across the range. There is also an opportunity for further phenotyping across Berthelot's pipit populations, to identify genotype-phenotype associations and their impact on individual fitness. Studies using such expansive sequencing and phenotyping are able to more accurately apply allele frequency approaches to i) characterise genome-wide variation including population history and gene flow, and ii) conduct genome scans for loci under selection and genome-wide association studies for traits of interest (Lamichhaney *et al.* 2015; Zhou *et al.* 2016; Varshney *et al.* 2017; Van Belleghem *et al.* 2021).

In Chapter 4 I detail novel candidate genes that show divergence across populations potentially because of different immunological and morphological adaptive evolution across these populations. Designing primers to enable the amplification of the entire region around these candidate regions would allow divergent SNPs within the region to be characterised. It is then possible to very cheaply conduct targeted genotyping at these loci across hundreds of individuals per population to assess population-level allele frequencies and to undertake direct association studies (reviewed in Hirschhorn and Daly 2005). Intense population sampling has already been undertaken across many Berthelot's pipit island populations with high disease prevalence (during my PhD and previous studies), for the purpose of further investigating novel disease candidates revealed in this thesis. For association studies, detailed environmental gradients/phenotyping need to be undertaken at the individual level, such as, for example, disease screening to associate pox or malaria strains with genotype at specific loci.

Long-read and linked-read genomic technologies provide the opportunity to explore structural variants and recombination landscape, which may enable further investigation of the potential inversion haplotype identified between the tawny and Berthelot's pipit speciation in Chapter 4. Generating phased haplotypes in heterozygous individuals may be aided by technologies such as Hi-C to generate chromosome-level scaffolds (Korbel and

Lee 2013). Studies on inversion haplotypes in natural populations are beginning to reveal the mechanisms driving and maintaining them in various taxa including birds (great tit (*Parus major*; Da Silva *et al.* 2019), willow warbler (*Phylloscopus trochilus*; Lundberg *et al.* 2017) and white-throated sparrow (*Zonotrichia albicollis*; Zinzow-Kramer *et al.* 2015)), nonhuman primates (Porubsky *et al.* 2020), fish, (Japanese grenadier anchovy (*Coilia nasus*; Zong *et al.* 2021) and insects (honeybees (*Apis mellifera*; Christmas *et al.* 2019)). Aside from several inversions found to correlate strongly with behavioural and mating phenotypes within species (Küpper *et al.* 2016; Tuttle *et al.* 2016), the adaptive value of intrachromosomal rearrangements is largely unknown.

In Chapter 4, a candidate gene potentially under balancing selection was identified within a genomic valley of divergence. This gene interacts with the expression of MHC class II; which has been reported in other studies (e.g., Ekblom *et al.* 2010; Herdegen-Radwan *et al.* 2021). The high levels of MHC gene duplication and conversion (among other gene families) in passerines means MHC genes within individuals are difficult to assemble with short-read sequencing technology. Together this means some functional MHC alleles may not be assigned in the genome, hence selection potentially acting directly on these loci may be missed unless a targeted sequencing approach is applied (reviewed in O'Connor *et al.* 2019). Long-read and linked-read sequencing technologies such as PacBio and Oxford Nanopore (see Chapter 1) would allow full length MHC alleles to be sequenced and mapped (e.g., Westbrook *et al.* 2015; He *et al.* 2020). Such chromosome-level genome assemblies may soon make it possible to position these highly replicated gene families in Berthelot's pipit, and for patterns of divergence within the MHC to be studied in genome scans.

Although this thesis has focussed on spatial scales of evolution in Berthelot's pipit, these findings, across populations that have been separate for a range of relatively short evolutionary timescales, show how different timeframes are important in shaping population structure, selection and morphological divergence. Studies on museum samples across the Berthelot's pipit range are now underway (personal communication, Sheppard *et al.*) and will add an additional temporal dimension to understanding evolution within, as well as across, populations. Using historical DNA (hDNA) it may be possible to track the fate of alleles within genes through time in this species, to better understand the effects of drift and selection, as has been used across a range of species (cf. Hoeck *et al.* 2010; Bi *et al.* 2019; Parejo *et al.* 2020; O'Toole *et al.* 2021). This hDNA approach has been used to gain further understanding of pathogen-mediated selection (Mikheyev *et al.* 2015; Gilroy *et al.* 2016; Alves *et al.* 2019). Museum samples could easily be used in future studies of traits of interest, such as immune function or bill morphology, in Berthelot's pipits where they may aid

not only in detecting the genes under selection but also the relative role of the different mechanisms at play (see Spurgin and Richardson 2010).

Genome sequencing of the tawny pipit, the Berthelot's pipit sister species, has provided a mainland comparison and context for diversity and divergence across the Berthelot's pipit range. However, this does not allow for assessment of broad evolution of key traits across phylogenetic trees. It may therefore be helpful to contextualise genomic findings across the Berthelot's pipit range using other avian taxa as an outgroup, for which sequence data is often freely available. Studies of diversity and divergence between taxa can reveal distinct or shared evolutionary trajectories across a range of taxonomic scales, which may also be used to understand evolution of specific genes or traits (Rodrigues *et al.* 2014; Van Doren *et al.* 2017; Minias *et al.* 2021).

3 General conclusions

By making use of the environmental, morphological and genetic variation that exists across island Berthelot's pipit populations, this thesis has provided several key insights into how population history, genetic drift and natural selection shape genetic diversity and divergence in the wild. The research presented in this thesis provides some general conclusions that may be useful for similar studies. Firstly, the importance of considering multiple evolutionary explanations for patterns of genetic diversity and divergence are highlighted. Findings show both demography and selection can result in very similar genomic signatures, or indeed signatures can be confounded by the effects of strong drift. Secondly, conclusions drawn from reduced representation and whole genome sequencing on Berthelot's pipit suggest similar patterns of colonisation history, drift and selection pressures. That said, the ability to visualise SNP-by-SNP patterns of diversity and divergence with whole genome sequencing has aided our deeper understanding of contemporary population processes, as well as more clearly identified divergence between populations. Sequencing limitations are mostly alleviated now that it is realistic for studies to generate population-scale whole genome sampling which is neither limited by individual or marker sampling. Finally, using a genome scan approach to understanding divergence across these fragmented populations identified i) new candidate genes associated with regulation of assumed adaptive traits previously investigated in the Berthelot's pipit and ii) candidate genes associated with traits not previously investigated. These findings emphasise the importance of not only focussing on

traits and genes of known interest if we are to gain a more complete understanding of what drives divergence, adaptation and speciation.

Berthelot's pipit provides an excellent model for investigating evolution of genetic diversity, divergence and speciation through successive colonisation events and across varied landscapes. It is my hope that the research presented in this thesis may be useful in understanding the ability of wild populations to adapt to changing or novel selection pressures. These findings emphasise the key roles population history, genetic drift and selection play in shaping genetic diversity and divergence following the establishment of new populations, and how this may lead to eventual speciation.

4 References

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

Alef, T., Torres, S., Hausser, I., Metze, D., Türsen, Ü., Lestringant, G.G. and Hennies, H.C. (2009). Ichthyosis, follicular atrophoderma, and hypotrichosis caused by mutations in ST14 is associated with impaired profilaggrin processing. *Journal of Investigative Dermatology* **129**:862–869.

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., 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.

Barrientos, R., Kvist, L., Barbosa, A., Valera, F., López-Iborra, G.M. and 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.

Baxter, S.W., Nadeau, N., Maroja, L., Wilkinson, P., Counterman, B.A., Dawson, A., Beltran, M., *et al.* (2010). Genomic hotspots for adaptation: The population genetics of Müllerian mimicry in Heliconius erato. *PLoS Genetics* **6**.

Van Belleghem, S.M., Cole, J.M., Montejo-Kovacevich, G., Bacquet, C.N., McMillan, W.O., Papa, R. and Counterman, B.A. (2021). Selection and isolation define a heterogeneous divergence landscape between hybridizing Heliconius butterflies. *Evolution* **75**:2251–2268.

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.

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

Case, T.J. and Schwaner, T.D. (1993). Island/mainland body size differences in Australian varanid lizards. *Oecologia* **94**:102–109.

Chakrama, F.Z., Seguin-Py, S., Le Grand, J.N., Fraichard, A., Delage-Mourroux, R., Despouy, G., Perez, V., *et al.* (2010). *GABARAPL1* (*GEC1*) associates with autophagic vesicles. *Autophagy* **6**:495–505.

Chen, M. and Chiang, C.W.K. (2021). Allele frequency differentiation at height-associated SNPs among continental human populations. *European Journal of Human Genetics* **29**:1542–1548.

Christmas, M.J., Wallberg, A., Bunikis, I., Olsson, A., Wallerman, O. and Webster, M.T. (2019). Chromosomal inversions associated with environmental adaptation in honeybees. *Molecular Ecology* **28**:1358–1374.

Clegg, S.M. and Owens, P.F. (2002). The 'island rule' in birds: medium body size and its ecological explanation. *Proceedings of the Royal Society B: Biological Sciences* **269**:1359–1365.

Cousin, H., Gaultier, A., Bleux, C., Darribère, T. and Alfandari, D. (2000). PACSIN2 is a regulator of the metalloprotease/disintegrin *ADAM13*. *Developmental Biology* **227**:197–210.

Dai, F., Lin, X., Chang, C. and Feng, X.H. (2009). Nuclear Export of Smad2 and Smad3 by RanBP3 Facilitates Termination of TGF-β Signaling. *Developmental Cell* **16**:345–357.

Ekblom, R., Sæther, S.A., Fiske, P., Kålås, J.A. and Höglund, J. (2010). Balancing selection, sexual selection and geographic structure in MHC genes of Great Snipe. *Genetica* **138**:453–461.

Elliott, J. and Johnston, J.A. (2004). SOCS: Role in inflammation, allergy and homeostasis. *Trends in Immunology* **25**:434–440.

Escoda, L. and Castresana, J. (2021). The genome of the Pyrenean desman and the effects of bottlenecks and inbreeding on the genomic landscape of an endangered species. *Evolutionary Applications* **14**:1898–1913.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X. and Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology and Evolution* **4**:1105–1115.

Foote, A.D., Hooper, R., Alexander, A., Baird, R.W., Baker, C.S., Ballance, L., Barlow, J., *et al.* (2021). Runs of homozygosity in killer whale genomes provide a global record of demographic histories. *Molecular Ecology*:1–16.

Frentiu, F.D., Bernard, G.D., Cuevas, C.I., Sison-Mangus, M.P., Prudic, K.L. and Briscoe, A.D. (2007). Adaptive evolution of color vision as seen through the eyes of butterflies. *In the Light of Evolution* **1**:187–204.

Gewin, L., Myers, H., Kiyono, T. and Galloway, D.A. (2004). Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. *Genes and Development* **18**:2269–2282.

Gilroy, D., van Oosterhout, C., Komdeur, J. and Richardson, D.S. (2016). Avian β-defensin

variation in bottlenecked populations: the Seychelles warbler and other congeners. *Conservation Genetics* **17**:661–674.

Gonzalez-Quevedo, C., Davies, R.G. and 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.

Goodman, S.J., Tamate, H.B., Wilson, R., Nagata, J., Tatsuzawa, S., Swanson, G.M., Pemberton, J.M., *et al.* (2001). Bottlenecks, drift and differentiation: The population structure and demographic history of sika deer (*Cervus nippon*) in the Japanese archipelago. *Molecular Ecology* **10**:1357–1370.

Gupta, K., Janardan, V., Banerjee, S., Chakrabarti, S., Srinivas, S., Mahishi, D., Raghu, P., *et al.* (2020). An evolutionarily conserved metallophosphodiesterase is a determinant of lifespan in Drosophila. *bioRxiv*.

Gupta, S.K. and Vlahakis, N.E. (2010). Integrin α9β1: Unique signaling pathways reveal diverse biological roles. *Cell Adhesion and Migration* **4**:194–198.

Han, F., Lamichhaney, S., Rosemary Grant, B., Grant, P.R., Andersson, L. and Webster, M.T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research* **27**:1004–1015.

Han, P., Werber, J., Surana, M., Fleischer, N. and Michaeli, T. (1999). The calcium/calmodulin-dependent phosphodiesterase PDE1C down-regulates glucose-induced insulin secretion. *Journal of Biological Chemistry* **274**:22337–22344.

Harakalova, M., Van Harssel, J.J.T., Terhal, P.A., Van Lieshout, S., Duran, K., Renkens, I., Amor, D.J., *et al.* (2012). Dominant missense mutations in *ABCC9* cause Cantú syndrome. *Nature Genetics* **44**:793–796.

He, K., Minias, P. and Dunn, P.O. (2020). Long-Read Genome Assemblies Reveal Extraordinary Variation in the Number and Structure of MHC Loci in Birds. *Genome biology and evolution* **13**:1–13.

Herdegen-Radwan, M., Phillips, K.P., Babik, W., Mohammed, R.S. and Radwan, J. (2021). Balancing selection versus allele and supertype turnover in MHC class II genes in guppies. *Heredity* **126**:548–560.

Hirschhorn, J. and Daly, M. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* **6**.

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

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

Jarvi, S.I., Atkinson, C.T. and Fleischer, R.C. (2001). Immunogenetics and Resistance To Avian Malaria in Hawaiian Honeycreepers (*Drepanidinae*). *Studies in Avian Biology No* **22**:254–263.

Jordan, M.A. and Snell, H.L. (2008). Historical fragmentation of islands and genetic drift in populations of Galápagos lava lizards (*Microlophus albemarlensis* complex). *Molecular Ecology* **17**:1224–1237.

Ju, L.-G., Zhu, Y., Lei, P.-J., Yan, D., Zhu, K., Wang, X., Li, Q.-L., *et al.* (2017). TTLL12 Inhibits the Activation of Cellular Antiviral Signaling through Interaction with VISA/MAVS. *The Journal of Immunology* **198**:1274–1284.

Kangas, V.M., Carrillo, J., Debray, P. and 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**:1–13.

Kardos, M., Åkesson, M., Fountain, T., Flagstad, Ø., Liberg, O., Olason, P., Sand, H., *et al.* (2017). Genomic consequences of intensive inbreeding in an isolated wolf population. *Nature Ecology & Evolution* **2**:124–131.

Kim, K., Perroud, B., Espinal, G., Kachinskas, D., Austrheim-Smith, I., Wolfe, B.M. and Warden, C.H. (2008). Genes and networks expressed in perioperative omental adipose tissue are correlated with weight loss from Roux-en-Y gastric bypass. *International Journal of Obesity* **32**:1395–1406.

Kirkpatrick, M. and Barton, N. (2006). Chromosome inversions, local adaptation and speciation. *Genetics* **173**:419–434.

Korbel, J.O. and Lee, C. (2013). Genome assembly and haplotyping with Hi-C. *Nature Biotechnology* **31**:1099–1101.

Küpper, C., Stocks, M., Risse, J.E., Dos Remedios, N., Farrell, L.L., McRae, S.B., Morgan, T.C., *et al.* (2016). A supergene determines highly divergent male reproductive morphs in the ruff. *Nature Genetics* **48**:79–83.

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

Leisler, B. and Winkler, H. (2015). Evolution of island warblers: Beyond bills and masses. *Journal of Avian Biology* **46**:236–244.

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

Liu, C., Chen, G., Bentley, A.R., Doumatey, A., Zhou, J., Adeyemo, A., Yang, J., *et al.* (2019). Genome-wide association study for proliferative diabetic retinopathy in Africans. *npj Genomic Medicine* **4**:1–7.

Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., Åkesson, S., *et al.* (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters* **1**:155–168.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Slazar, C., Walters, J.R., Simpson, F., Blaxter, M., *et al.* (2013). Genone-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research* **23**:1817–1828.

McCulloch, G.A., Guhlin, J., Dutoit, L., Harrop, T.W.R., Dearden, P.K. and Waters, J.M. (2021). Genomic signatures of parallel alpine adaptation in recently evolved flightless insects. *Molecular Ecology*:1–10.

Meier, K., Hansen, M.M., Bekkevold, D., Skaala and Mensberg, K.L.D. (2011). An assessment of the spatial scale of local adaptation in brown trout (*Salmo trutta L.*): Footprints of selection at microsatellite DNA loci. *Heredity* **106**:488–499.

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

Minias, P., Włodarczyk, R., Remisiewicz, M., Cobzaru, I. and Janiszewski, T. (2021). Distinct evolutionary trajectories of MHC class I and class II genes in Old World finches and buntings. *Heredity* **126**:974–990.

O'Connor, E.A., Westerdahl, H., Burri, R. and Edwards, S. V. (2019). Avian MHC Evolution in the Era of Genomics: Phase 1.0. *Cells* **8**:1–21.

O'Toole, B., Simmons, N.B. and Hekkala, E. (2021). Reconstructing the Genomic Diversity of a Widespread Sub-Saharan Bat (*Pteropodidae: Eidolon helvum*) Using Archival Museum Collections. *Acta Chiropterologica* **22**:227–241.

Parejo, M., Wragg, D., Henriques, D., Charrière, J.D. and 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.

Park, J.A. and Kim, K.C. (2010). Expression patterns of *PRDM10* during mouse embryonic development. *BMB Reports* **43**:29–33.

Porubsky, D., Sanders, A.D., Höps, W., Hsieh, P.H., Sulovari, A., Li, R., Mercuri, L., *et al.* (2020). Recurrent inversion toggling and great ape genome evolution. *Nature Genetics* **52**:849–858.

Protas, M.E., Hersey, C., Kochanek, D., Zhou, Y., Wilkens, H., Jeffery, W.R., Zon, L.I., *et al.* (2006). Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* **38**:107–111.

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

Qin, Z., Qu, X., Lei, L., Xu, L. and Pan, Z. (2020). Y-box-binding protein 3 (*YBX3*) restricts influenza A virus by interacting with viral ribonucleoprotein complex and imparing its function. *Journal of General Virology* **101**:385–398.

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.

Robinson, J.A., Bowie, R.C.K., Dudchenko, O., Aiden, E.L., Hendrickson, S.L., Steiner, C.C., Ryder, O.A., *et al.* (2021). Genome-wide diversity in the California condor tracks its prehistoric abundance and decline. *Current Biology* **31**:2939–2946.e5.

Rodrigues, P., Lopes, R.J., Reis, S., Resendes, R., Ramos, J.A. and Tristão da Cunha, R. (2014). Genetic diversity and morphological variation of the common chaffinch *Fringilla coelebs* in the Azores. *Journal of Avian Biology* **45**:167–178.

Safran, R.J., Scordato, E.S.C., Wilkins, M.R., Hubbard, J.K., Jenkins, B.R., Albrecht, T., Flaxman, S.M., *et al.* (2016). Genome-wide differentiation in closely related populations: The roles of selection and geographic isolation. *Molecular ecology* **25**:3865–3883.

De Santo, C., D'Aco, K., Araujo, G.C., Shannon, N., Study, D.D.D., Vernon, H., Rahrig, A., *et al.* (2015). WAC loss-of-function mutations cause a recognisable syndrome characterised by dysmorphic features, developmental delay and hypotonia and recapitulate 10p11.23 microdeletion syndrome. *Journal of Medical Genetics* **52**:754–761.

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: Genes, Genomes, Genetics* **10**:3147–3163.

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

Shaheen, R., Hashem, A., Abdel-Salam, G.M.H., Al-Fadhli, F., Ewida, N. and Alkuraya, F.S. (2016). Mutations in CIT, encoding citron rho-interacting serine/threonine kinase, cause severe primary microcephaly in humans. *Human Genetics* **135**:1191–1197.

Da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., Van Oers, K., Dibbits, B., *et al.* (2019). The Genomic Complexity of a Large Inversion in Great Tits. *Genome Biology and Evolution* **11**:1870–1881.

Spurgin, L.G., Illera, J.C., Jorgensen, T.H., Dawson, D.A. and 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. and 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. and Richardson, D.S. (2010). How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. *Proceedings of the Royal Society B: Biological Sciences* **277**:979–988.

Stoffel, M.A., Johnston, S.E., Pilkington, J.G. and Pemberton, J.M. (2021). Mutation load decreases with haplotype age in wild Soay sheep. *Evolution Letters* **5**:187–195.

Szarmach, S.J., Brelsford, A., Witt, C. and Toews, D.P.L. (2021). Comparing divergence landscapes from reduced-representation and whole-genome re-sequencing in the yellow-rumped warbler (*Setophaga coronata*) species complex . *Molecular Ecology*:1–12.

Szpiech, Z.A., Xu, J., Pemberton, T.J., Peng, W., Zöllner, S., Rosenberg, N.A. and Li, J.Z. (2013). Long runs of homozygosity are enriched for deleterious variation. *American Journal of Human Genetics* **93**:90–102.

Tokumasu, Y., Iida, A., Wang, Z., Ansai, S., Kinoshita, M. and Sehara-Fujisawa, A. (2016). *ADAM12*-deficient zebrafish exhibit retardation in body growth at the juvenile stage without developmental defects. *Development Growth and Differentiation* **58**:409–421.

Tu, C.F., Yan, Y.T., Wu, S.Y., Djoko, B., Tsai, M.T., Cheng, C.J. and Yang, R.B. (2008). Domain and functional analysis of a novel platelet-endothelial cell surface protein, *SCUBE*. *Journal of Biological Chemistry* **283**:12478–12488.

Tuttle, E.M., Bergland, A.O., Korody, M.L., Brewer, M.S., Newhouse, D.J., Minx, P., Stager, M., *et al.* (2016). Divergence and functional degradation of a sex chromosome-like supergene. *Current Biology* **26**:344–350.

Urbanek, K., Sutherland, D.M., Orchard, R.C., Wilen, C.B., Knowlton, J.J., Aravamudhan, P., Taylor, G.M., *et al.* (2020). Cytidine Monophosphate N -Acetylneuraminic Acid

Synthetase and Solute Carrier Family 35 Member A1 Are Required for Reovirus Binding and Infection . *Journal of Virology* **95**.

Van Doren, B.M., Campagna, L., Helm, B., Illera, J.C., Lovette, I.J. and Liedvogel, M. (2017). Correlated patterns of genetic diversity and differentiation across an avian family. *Molecular Ecology* **26**:3982–3997.

Varshney, R.K., Saxena, R.K., Upadhyaya, H.D., Khan, A.W., Yu, Y., Kim, C., Rathore, A., *et al.* (2017). Whole-genome resequencing of 292 pigeonpea accessions identifies genomic regions associated with domestication and agronomic traits. *Nature Publishing Group* **49**:1082–1088.

Veenman, L., Papadopoulos, V. and Gavish, M. (2007). Channel-like functions of the 18-kDa translocator protein (TSPO): regulation of apoptosis and steroidogenesis as part of the host-defense response. *Current pharmaceutical design* **13**:2385–405.

Voelker, G. (1999). Dispersal, Vicariance, and Clocks: Historical Biogeography and Speciation in a Cosmopolitan Passerine Genus (*Anthus: Motacillidae*). *Evolution* **53**:1536.

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

Westbrook, C.J., Karl, J.A., Wiseman, R.W., Mate, S., Koroleva, G., Garcia, K., Sanchez-Lockhart, M., *et al.* (2015). No assembly required: Full-length MHC class I allele discovery by PacBio circular consensus sequencing. *Human Immunology* **76**:891–896.

White, T.A. and Searle, J.B. (2007). Factors explaining increased body size in common shrews (*Sorex araneus*) on Scottish islands. *Journal of Biogeography* **34**:356–363.

Wilder, A.P., Palumbi, S.R., Conover, D.O. and Therkildsen, N.O. (2020). Footprints of local adaptation span hundreds of linked genes in the Atlantic silverside genome. *Evolution Letters* **4**:430–443.

Xavier, G.M., Sharpe, P.T. and Cobourne, M.T. (2009). Scube1 is expressed during facial development in the mouse. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution* **312**:518–524.

Xu, H., Qu, C., Gan, L., Sun, K., Tan, J., Liu, X., Jiang, Z., *et al.* (2020). Deletion of the Impg2 gene causes the degeneration of rod and cone cells in mice. *Human Molecular Genetics* **29**:1624–1634.

Yassin, A., Delaney, E.K., Reddiex, A.J., Seher, T.D., Bastide, H., Appleton, N.C., Lack, J.B., *et al.* (2016). The pdm3 Locus Is a Hotspot for Recurrent Evolution of Female-Limited Color Dimorphism in Drosophila. *Current Biology* **26**:2412–2422.

Yokoyama, S. and Yokoyama, R. (1996). Adaptive evolution of photoreceptors and visual pigments in vertebrates. *Annual Review of Ecology and Systematics* **27**:543–567.

Yoshida, G.M. and Yáñez, J.M. (2021). Multi-trait GWAS using imputed high-density genotypes from whole-genome sequencing identifies genes associated with body traits in Nile tilapia. *BMC Genomics* **22**:1–13.

Zahir, F.R., Baross, A., Delaney, A.D., Eydoux, P., Fernandes, N.D., Pugh, T., Marra, M.A., *et al.* (2008). A patient with vertebral, cognitive and behavioural abnormalities and a de novo deletion of NRXN1α. *Journal of Medical Genetics* **45**:239–243.

Zhang, F., Guo, X., Zhang, Y., Wen, Y., Wang, W., Wang, S., Yang, T., *et al.* (2014). Genome-wide copy number variation study and gene expression analysis identify *ABI3BP* as a susceptibility gene for Kashin-Beck disease. *Human Genetics* **133**:793–799.

Zhang, M., Suren, H., Holliday, J.A. and Gaut, B. (2019). Phenotypic and Genomic Local Adaptation across Latitude and Altitude in Populus trichocarpa. *Genome Biology and Evolution* **11**:2256–2272.

Zhang, X., Rayner, J.G., Blaxter, M. and Bailey, N.W. (2021). Rapid parallel adaptation despite gene flow in silent crickets. *Nature Communications* **12**.

Zhou, C., Qian, X., Hu, M., Zhang, R., Liu, N., Huang, Y., Yang, J., *et al.* (2020). *STYK1* promotes autophagy through enhancing the assembly of autophagy-specific class III phosphatidylinositol 3-kinase complex I. *Autophagy* **16**:1786–1806.

Zhou, X., Meng, X., Liu, Z., Chang, J., Wang, B., Li, M., Wengel, P.O. Ter, *et al.* (2016). Population Genomics Reveals Low Genetic Diversity and Adaptation to Hypoxia in Snub-Nosed Monkeys. *Molecular Biology and Evolution* **33**:2670–2681.

Zinzow-Kramer, W.M., Horton, B.M., Mckee, C.D., Michaud, J.M., Tharp, G.K., Thomas, J.W., Tuttle, E.M., *et al.* (2015). Genes located in a chromosomal inversion are correlated with territorial song in white-throated sparrows. *Genes, Brain and Behavior* **14**:641–654.

Zong, S.B., Li, Y.L. and Liu, J.X. (2021). Genomic Architecture of Rapid Parallel Adaptation to Fresh Water in a Wild Fish. *Molecular biology and evolution* **38**:1317–1329.