J.C. Stonehouse

The genomics of adaptation to climate in European great tit (*Parus major*) populations

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Submission 2024

The genomics of adaptation to climate in European great tit (*Parus major*) populations

Joanne Clare Stonehouse

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Ecology and Evolutionary Biology
School of Biosciences
The University of Sheffield

Submission April 2024

Acknowledgements

I would like to massively thank my PhD supervisor Prof Jon Slate for all the effort and patience required for supporting this project, especially as much of it was remotely after the pandemic.

I appreciate the collaborators and researchers for their previous work and data collection on the genomics of great tit populations, that went into providing a starting point for this project.

I appreciated the chat with my lab group and office mates in Sheffield, and being around like-minded people in APS. I enjoyed the views from Stanage Edge and ice cream from Our Cow Molly in Sheffield. I also had the opportunity to meet fellow ACCE PhD students at the University of York and University of Liverpool.

I would like to appreciate my brother Robert for a huge amount of childcare that gave me time to think without distractions. I also give thanks to the nursery and school teachers that we have met on the way.

I started my PhD with my border collie Inca, walking through the magical Rivelin Valley. My twins toddled through Hillsborough Park with Inca. A time that won't be forgotten.

I was inspired to love nature and science from my wonderful Mum. I spent many an afternoon at the Pinewoods in Harrogate with my Nan. I hope to inspire my 6 year olds Zachary and Owen in their own playful learning.

Abstract

Understanding the genomics of adaptation to past climate changes is important to consider in evolutionary ecology. Using the great tit (*Parus major*), as a model species, I sought to gain insight into the processes of adaptation, both within and between populations, in a wild system. I performed two different types of tests to detect regions under positive selection. In Chapter 2, I performed a multiple-population genome-environment association (GEA) analysis broadly spanning the species range. I identified 36 candidate genes associated with adaptation to climate. The data suggested that climate adaptation appears to be polygenic and genetically complex, involving many different genes and biological process pathways. In Chapter 3, I performed extended haplotype homozygosity testing to detect regions under selective sweeps. I identified 17 potential climate adaptation regions, that had long haplotypes consistent with positive selection in single or geographically close populations. I found little evidence of signatures of parallel evolution in different populations.

I then further investigated the temporal trends of the 11 sweep regions associated with climate adaptation in the UK populations, using whole-genome sequenced museum specimens (Chapter 4). Three candidate climate adaptation SNPs showed significant allele frequency changes over the last century. The closest genes to these three potential climate adaptation regions were identified as the best supported candidate genes. Evidence supporting them included stringent tests to detect selection from genomic data and an unusually strong temporal trend in the UK.

The GEA analysis in chapter 2 suggests that substantial climate-associated genetic variation remains in the pan-European populations, which is likely to be important for adaptation in response to future predicted climate changes. Models of future climate changes particularly relevant to wild great tit populations include more frequent and intense heat waves predicted in the Mediterranean region, which could cause population crashes due to heat stress, especially in the nestling stage where parental care is required.

Table of contents

Acknowledgements ... *3*

Abstract ... *4*

Table of contents ... *5*

Declaration ... *9*

1. General Introduction ... 11

1.1. Population responses to climate change ... *11*

1.2. Genetic adaptation to climate changes ... *12*
1.2.1. Adaptive traits: few loci of large effect ... 13
1.2.2. Polygenic adaptation: many loci of small effect ... 15
1.2.3. Other evolutionary processes of long-term allele frequency changes ... 15
1.2.4. Parallel evolution of local adaptation ... 16

1.3. Examining spatial evolutionary changes of adaptation ... *16*
1.3.1. Genomic detection of signatures of local adaptation ... 16
1.3.2. Cross-population EHH statistics to detect selective sweeps ... 17
1.3.3. Genome scans of population differentiation to detect regions under selection ... 18
1.3.4. Genome-environment association to infer candidate adaptive loci ... 20
1.3.5. Common considerations for using methods to detect regions under selection ... 22

1.4. Comparative candidate climate adaptation genes across vertebrates ... *22*
1.4.1. Experimental validation of candidate genes ... 23

1.5. Examining temporal evolutionary changes of adaptation ... *24*
1.5.1. Museum time-series analysis of evolutionary change ... 24
1.5.2. Experimental evolution over generations ... 25
1.5.3. Local adaptation and genomic vulnerability for population conservation ... 26

1.6. Great tit study species ... *26*
1.6.1. Great tit species complex ... 27
1.6.2. Population distribution of great tits ... 27
1.6.3. Great tit dispersal and gene flow ... 30
1.6.4. Demographic trends in contemporary European great tit populations ... 32
1.6.5. Great tit ecological model system ... 32

1.7. Local adaptation and climate change ... *34*
1.7.1. Climate change and advancement of egg lay date ... 34
1.7.2. Long-term field data to ask evolutionary questions over time ... 36
1.7.3. Local environmental conditions ... 36
1.7.4. Local adaptation to urbanisation ... 37
1.7.5. Local adaptation to past climate changes ... 38
1.7.6. Forecast climate change of the future ... 38

1.8. Great tit Genome ... *40*
1.8.1. Great tit reference genome ... 40

1.8.2. 10K SNP chip genotyping 42
1.8.3. Whole-genome sequencing 43
1.8.4. 500K SNP Chip genotyping 43
1.8.5. The Great Tit HapMap project – genotyped populations 44
1.8.6. Low population differentiation 46
1.8.7. Signatures of selection in the great tit genome 46

1.9. Summary of Introduction Chapter 1 47
1.9.1. Identifying adaptive genetic variation 47
1.9.2. Population genomics of climate adaptation in a model system 47

1.10. Outline summary of thesis chapters 48

1.11. References 49

2. The genomics of adaptation to climate in European great tit (Parus major) populations. 68

2.1. Abstract 70

2.2. Impact Summary 71

2.3. Introduction 72

2.4. Materials and Methods 77

2.5. Results 83

2.6. Discussion 88

2.7. Acknowledgements 93

2.8. Author Contributions 93

2.9. Data Accessibility 94

2.10. References 95

2.11. Supplementary Material for Chapter 2 digital file 104
2.11.1. Appendix 2 (supplementaryChapter2.pdf) 104
2.11.2. GOtoGene_PC1.txt 104
2.11.3. GOtoGene _PC2.txt 104
2.11.4. GOtoGene _PC3.txt 104
2.11.5. GOtoGene _PC4.txt 104

3. Haplotype-based tests for selective sweeps within and between populations 105

3.1. Abstract 105

3.2. Introduction 106
3.2.1. Spatial Patterns of Selection and Local Adaptation 106
3.2.2. Parallel adaptive evolution in geographically separated populations 107
3.2.3. Detecting selection 109
3.2.4. Haplotype-based tests for selective sweeps 109
3.2.5. Integrated haplotype homozygosity score 110
3.2.6. Complementary Haplotype-based statistics 111
3.2.7. Signatures of selection in the great tit HapMap dataset 112

3.3. Methods *115*
3.3.1. Study system and genomic data 115
3.3.2. Cross-population XP-EHH 116
3.3.3. Pairwise Rsb 117
3.3.4. Extended haplotype homozygosity tests to identify selective sweeps 118

3.4. Results *120*

3.5. Discussion *128*
3.5.1. Genes within sweep regions and their effects in other species 128
3.5.2. Selective sweeps in non-UK outlier populations 130
3.5.3. Do selective sweeps provide support for parallel evolution? 131
3.5.4. Are candidate genes in regions of low recombination? 131
3.5.5. Limitation of methods to detect sweeps 132
3.5.6. Combined methods to detect sweeps 133

3.6. References *136*

3.7. Supplementary Material Chapter 3 digital files *146*
3.7.1. Appendix 3 (supplementaryChapter3.pdf) 146
3.7.2. S.Tables S3.2-S3.9 146

4. Temporal trends in allele frequencies of potential climate adaptation genes in the UK great tit museum record 147

4.1. Abstract *147*

4.2. Introduction *148*

4.3. Methods *151*
4.3.1. DNA sample collection 152
4.3.2. Paired-end sequencing, trimming and alignment 152
4.3.3. Genotype Calling 153
4.3.4. Investigate Z-linked markers erroneously mapped to an autosome 157

4.4. Results *160*

4.5. Discussion *165*
4.5.1. Further Investigation of temporal trends at candidate adaptation genes 167
4.5.2. Sequencing from museum samples 167
4.5.3. Future directions 168

4.6. References *170*

4.7. Appendix Chapter 4 digital files *173*
4.7.1. Appendix 4 (supplementaryChapter4.pdf) 173
Including Figures S4.1-S4.6 showing signatures of selective sweeps in the UK museum record at SNPs: AX-100681217, AX-100879265 and AX-100459612. 173
4.7.2. Appendix 4.2 (sampleInfo.txt) 173
4.7.3. Appendix 4.3 (NEOF_application.docx) 173
4.7.4. Appendix 4.4 (sequencing report.pdf) 173

5. Discussion - Genomics of climate adaptation 174

5.1. Methods to validate candidate adaptive loci for climate adaptation ... 175

5.2. Gene editing for functional validation of candidate adaptation genes ... 176

5.3. Investigating future genomic responses for climate adaptation ... 176

5.3.1. Genomic selection experiments to investigate adaptation to future climate changes ... 177

5.3.2. Genomics of climate adaptation ... 179

5.4. Future research directions ... 180

5.4.1. Monitoring Technology ... 180

5.4.2. Predictive ecology ... 182

5.5. References ... 185

Declaration

Chapter 2: The genomics of adaptation to climate in European great tit (*Parus major*) populations

The following people were co-authors and involved in turning Thesis Chapter 2 into the publication of article 'The genomics of adaptation to climate in European great tit *(Parus major*) populations', in the Journal *Evolution Letters*, **8**(1), pages 18-28 (2024), doi: https://www.doi.org.uk/10.1093/evlett/qrad043.

Joanne C Stonehouse[1], Lewis G Spurgin[2], Veronika N Laine[3,6], Mirte Bosse[4,7], The Great Tit HapMap Consortium, Martien AM Groenen[4], Kees van Oers[3], Ben C Sheldon[5], Marcel E Visser[3], Jon Slate[1].

1. School of Biosciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK
2. School of Biological Sciences, University of East Anglia, Norwich Research Park, UK
3. Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, NL
4. Animal Breeding and Genomics, Wageningen University & Research, Wageningen, NL
5. Edward Grey Institute, Department of Biology, University of Oxford, UK
6. Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland
7. Amsterdam Institute for Life and Environment (A-LIFE), Section Ecology and Evolution, Vrije Universiteit Amsterdam, Amsterdam, NL

Author Contributions: JS conceived the study from data collected by the Great Tit HapMap Consortium project. JS, LGS, MB and VNL performed quality control and downstream filtering of the SNP data. JS, KvO, MAMG and MEV supervised genomic data collection. BCS, JS, KvO, B and MEV organised sample collection, with members of the GTHC providing samples and ecological information relevant to their study populations. JCS performed the computation and analysed the data, with credit to JS for the GO term gene scoring script addressing review's concerns with the default methods. JS and JCS drafted the initial version of the manuscript and all authors contributed to later versions of the manuscript. JCS produced all the figures. In addition, the manuscript was reviewed by two anonymous reviewers and a Dryad data repository reviewer.

Chapter 3: Haplotype-based tests for selective sweeps within and between populations

The following people were involved in this research project: Joanne C. Stonehouse, Prof. Jon Slate, working on phased chromosomes of the genotyped data produced by the great tit Hap Map Consortium (described in Spurgin et al. 2024).

JS phased each chromosome of the SNP genotype data into haplotypes;

JS conceived the idea for the study and provided guidance on methodology;

JCS performed the bioinformatics data analysis and produced the output figures;

JS supplied the data for the great tit recombination rate map and scripted a test to see if sweeps are found in low recombination regions;

JCS drafted the Chapter, guided by the comments provided by JS.

Chapter 4: Temporal trends in allele frequencies of potential climate adaptation genes in the UK great tit museum record

The following people were involved in this research project: Joanne Stonehouse, Prof. Jon Slate (University of Sheffield) and Dr De Chen (Beijing Normal University). The sequencing was performed by Claudia Wierzbicki, Charlotte Nelson and Richard Gregory at the NERC Environmental Omics Facility (NEOF), housed within the Centre for Genomic Research (CGR), University of Liverpool.

JS and DC collected the UK museum samples and prepared the DNA for sequencing;

CW prepared libraries, CN sequenced the samples, and RG trimmed and filtered the data;

JS provided guidance on methodology, and JCS and JS ran quality control on the samples;

JCS performed the bioinformatics data analysis, with credit to JS for code to remove Z-linked markers and to investigate binned regions of Chr 7 for a possible inversion.

JCS drafted the chapter and JS provided comments.

The contribution of the other collaborators to work acknowledged within this thesis has been indicated.

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university.

1. General Introduction

1.1. Population responses to climate change

All species of plants and animals worldwide are rapidly experiencing human-driven climate changes, the pace of which might preclude an adequate response in many populations, with worse-case scenarios of population crashes and extinctions threatening biodiversity (IPCC 2023). Global warming of the terrestrial and aquatic ecosystem environments may challenge physiological limits or alter phenology of breeding or life-history traits (Parmesan and Yohe 2003). The indirect effects of climate change possibly include alterations to community dynamics, with the introduction of new species, changes in community composition, or the introduction of pandemics to a region (Walther et al. 2002). Monitoring how populations and ecosystem functioning will respond to a changing environment is a major priority for conservation (Malhi et al. 2020). Populations that are unable to respond appropriately may suffer mortality of individuals and population declines (IPCC 2023).

Populations of animals are able to respond to a changing environment in three ways: 1) Through latitudinal or altitudinal range shifts to more favoured conditions (Rubenstein et al. 2023). For example, some dietary specialist butterfly species have colonized higher latitudes, that were previously unsuitable in in Sweden and Finland (Sunde et al. 2023). 2) By phenotypic plasticity, which is the ability of a genotype to express different phenotypes in different conditions (Price et al. 2003). Shifts in phenology due to climate change are reported for earlier egg laying and migration dates in birds, emergence timing in insects and date of tadpole metamorphosis in aquatic systems (Visser and Both 2005). Phenotypic plasticity may be heritable and partially controlled by genetic mechanisms (Oostra et al. 2018). 3) Organisms may respond by genetic adaptation to new environmental conditions i.e. evolutionary change (Parmesan and Yohe 2003; Atkins and Travis 2010). None of these three potential responses to climate change are mutually exclusive processes (Parmesan and Yohe 2003).

This Introduction Chapter will now focus on genetic adaptation to climate, as this is to be detected in the main theme of this thesis. Investigating the genomic basis underlying evolutionary responses to climate change is important to conduct, as genomic approaches will help gain insight into how vertebrate populations adapt /or have the potential to adapt to a changing climate.

1.2. Genetic adaptation to climate changes

Local adaptation is the process whereby individuals in one population have evolved phenotypes that give them higher relative fitness in their local environment, than contrasting to populations that are adapted to heterogeneous environments. For example, locally adapted genotypes outperform transplanted genotypes (Kawecki & Ebert 2004). Spatially varying natural selection acting across populations from different environments is a key driver of local adaptation e.g. driving local adaptation in reef coral from lagoon or reef slope habitats in Western Australia (Thomas et al. 2022) and driving local adaptation of the UK bank vole to regional differences in climate (Marková et al. 2023). Similarly, many desert-adapted mammals show adaptations in fat metabolism, insulin signaling, water retention and the arachidonic acid metabolism pathway, that are not seen in populations found in less dry and arid conditions (Rocha et al. 2021).

An adaptive allele may arise in a population by being: 1) Selected from the pre-existing standing genetic variation. 2) Selected as a newly arising independent mutation. 3) Introduced into the population via gene flow from a different locally adapted population (Pritchard et al. 2010). The genomic architecture of an adaptive trait influences how populations locally adapt, with produced allele frequency changes happening at a few genes of large effect or many genes of small effect (polygenic adaptation) or a combination of the two. The rate of genetic change will also be influenced by linkage disequilibrium,

recombination rate variation and the influence of gene interactions (see Wadgymar et al. 2017 and Pardo-Diaz et al. 2015 for reviews on identifying the loci of genetic adaptation).

1.2.1. Adaptive traits: few loci of large effect

A sequence of selective sweeps at a single locus or a few loci underlies the genetic basis of a range of adaptive traits (see **Figure 1.1** for an illustration of hard vs soft sweeps). In a hard selective sweep, a new mutation and long-range hitchhiking alleles are favoured, rapidly carrying a long haplotype to high frequency in the population i.e. fixation (Pritchard et al. 2010). Haplotype diversity is influenced by demography, marker spacing, between-haplotype recombination and rare mutation events (Stumpf et al. 2004).

For adaptive traits controlled by few loci of major effect, distinct genomic sweeps will be observed on few genes, characterised by large allele frequency changes in adaptive loci and those nearby, to form lengths of identical haplotypes (Reid et al. 2023). Classic examples exist for adaptive traits controlled by few major effect loci, namely adaptive melanism in the peppered moth (Van't Hof et al. 2016) and sickle cell anaemia in humans (Elguero et al. 2015). In addition, lactose tolerance in humans primarily occurs by evolutionary changes at only one gene, with multiple selective sweeps on different alleles being identified at the same gene (Tishkoff et al. 2007).

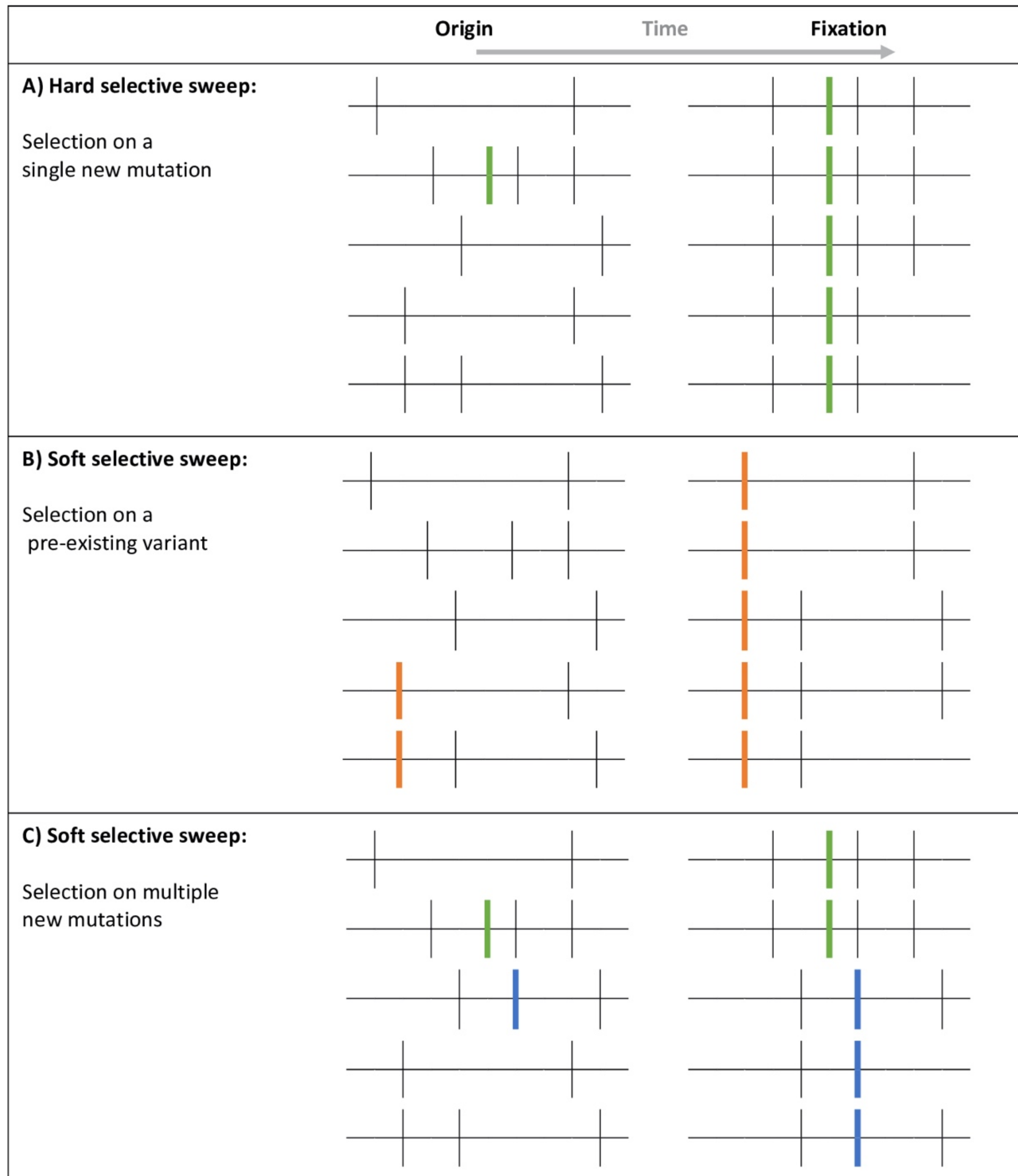


Figure 1.1: Hard and soft selective sweep processes. The green, blue and orange vertical lines represent alleles under selection, while the black lines are adjacent SNPs. Signatures of recently completed hard selective sweeps (panel A) include reduced genetic variation along the haplotype, although rare new mutations may arise. Soft sweeps include selection on a pre-existing mutation (panel B) or selection on multiple new mutations (panel C). Soft sweeps and partial (ongoing sweeps) are more difficult to detect than hard sweeps using standard tests of selection because the haplotypes under selection stand out less relative to haplotypes in the rest of the genome. Figure redrawn from Pritchard et al. 2010, with the original figure also published in Jenson 2014.

1.2.2. Polygenic adaptation: many loci of small effect

In polygenic adaptation changes to lots of genes are involved, whereby a phenotype evolves as a result of selection driving small effect allele frequency changes across hundreds or more sites i.e. many loci of small effect are contributing (Fisher 1919) – effectively by a series of selective sweeps. Most life-history adaptive traits are described as polygenic in plants, animals and humans (Visscher et al. 2017), or are expected to be. Highly polygenic traits may have a greater potential for adaptation to novel environments because of the larger number of loci that can change in frequency (Jain et al. 2017). An example of polygenic climate adaptation is observed between two common species of tabletop coral (*Acropora hyacinthus*), where evolved differences in microhabitat and thermal stress tolerance traits have resulted in variation in each species' resilience to climate change (Rose et al. 2018).

1.2.3. Other evolutionary processes of long-term allele frequency changes

The process of natural selection on alleles drives genetic adaptation to the local environment, but this positive selection is often opposed by allele frequency changes due to genetic drift and gene flow in natural populations (Andrews 2010). In addition, balancing selection can also produce short term directional selection (i.e. maintaining polymorphisms above what is expected under neutrality), resulting in stable allele polymorphisms (Charlesworth 2006). A case-study of balancing selection regimes on HLA (adaptive immune response) genes across human populations, showed different sets of alleles were maintained in different populations. The expectation from theory is that balancing selection reduces population differentiation. However, the authors of the case-study found balancing selection regimes could increase population differentiation (Brandt et al. 2018).

1.2.4. Parallel evolution of local adaptation

Adaptive traits can evolve in parallel in different populations or in different species from selection of the same contributing loci (Reid et al. 2023). The adaptive alleles may be repeatedly selected in parallel from shared ancestral standing genetic variation, independently arising mutations, or spread across populations via gene flow (Savolainen et al. 2013). An example of parallel local adaptation to altitude has been demonstrated for two tropical butterfly species found on mountain habitats on either side of the Andes, but not in five close relatives. This example of parallel adaptation in mountain butterflies found support for the role of standing genetic variation and gene flow from independently adapted species in the process of local adaptation (Montejo-Kovacevich et al. 2022). Furthermore, different combinations of allele changes may produce the same adaptive phenotype in different populations. This has been termed genotypic redundancy (Yeaman 2022) and is considered as the non-parallel evolution of the adaptive trait (Reid et al. 2023).

1.3. Examining spatial evolutionary changes of adaptation

1.3.1. Genomic detection of signatures of local adaptation

Signatures of local adaptation can be investigated across the genome using 1) extended haplotype homozygosity (EHH) statistics to detect selective sweeps; 2) genome scans of differentiation to detect signatures of selection; and 3) genome scans to perform genome-environment association (GEA) analyses. Selection signatures underlying adaptation of large effect loci are revealed more clearly with all of the above methods, than when trait architecture is highly polygenic and involves the detection of many small effect loci (Novembre and Barton 2018).

1.3.2. Cross-population EHH statistics to detect selective sweeps

Signatures of selective sweeps can be detected in the genome from a distinctive signal of low genetic variation across large regions of the extended linkage disequilibrium haplotype, with these regions departing from neutrality (Fay et al. 2002). A faster sweep carries a longer haplotype with it because recombination does not have time to whittle down the region. Similarly, the sweep is larger in regions with low recombination rates (Pritchard et al. 2010).

Several haplotype-based statistics have been developed for comparing EHH profiles between populations (reviewed by Abondio et al. 2022), to identify hard selective sweeps and infer when positive selection is recent and ongoing (Sabeti et al. 2007). Five widely used cross-population EHH statistics include the integrated haplotype homozygosity score iHS (Voight et al. 2006), cross population XP-EHH (Sabeti et al. 2007), the ratio of EHHS between populations Rsb (Tang et al. 2007), cross population composite likelihood ratio XP-CLR (Chen et al. 2010) and the more recently developed cross-population number of segregating sites by length XP-nSL (Szpiech et al. 2021).

Cross-population EHH statistics are often used in combination with each other and can be combined with the Fst statistic measure of genetic differentiation (see **Section 1.3.3**) to identify regions under selection, e.g. XP-EHH and Fst were combined in a study of selection across European and Middle East sheep breeds (Eydivandi et al. 2021). The significant loci identified in the intersection of the combined EHH statistics, are potentially the strongest candidate genes for adaptation traits. For example, using iHS, XP-EHH, Rsb and Fst methods, candidate genes for environmental adaptation traits have been identified in north African cattle. These genes are implicated in adaptive immune responses and for adaptation to extreme temperatures and/or chronic hypoxia in the mountain regions (Ben-Jemaa et al. 2020). Further studies using combined EHH statistics have identified selection signatures of selection for heat tolerance in Zebu (Li et al. 2020), and candidate genes for hot climate adaptation in sheep (Álvarez et al. 2020; Zhang et al. 2021).

1.3.3. Genome scans of population differentiation to detect regions under selection

Population differentiation between pairs of populations can be detected by fixation index Fst (Wright 1949) outlier approaches, comparing allele frequency changes because of divergent selection patterns, to within-population allele frequencies (Beaumont and Nichols 1996). The output Fst statistic can have a value of zero representing identical populations, ranging up to one, to representing fully differentiated populations (Wright 1949). Lewontin and Krakauer (1973) proposed that loci likely under selection could be identified from those Fst values that diverged from more or less than expected by chance, as all of the loci experience the same background processes e.g. drift and population structure. Beaumont and Nichols (1996) identified genomic regions significantly differentiated from neutral model expectations as ‘outliers’ (e.g. in the top 1% of Fst values), implementing a null model approach to identify regions likely under selection, from which candidate adaptive loci and their closest candidate genes can be identified (**Figure 1.2**). For example, Fst methods have been used to identify candidate genes for adaptation to salinity in Atlantic Herring (Larsson et al. 2007). Some extended population differentiation statistics have now been developed to account for genetic drift and hierarchical population subdivisions caused by variation in population size, structure and kinship in models when modelling the neutral genetic divergence between populations e.g. the FLK value (Bonhomme et al. 2010). However, in general Fst is robust to demography and is still a widely used first step for identifying candidate genes under selection in studies of evolutionary ecology (Beaumont 2005).

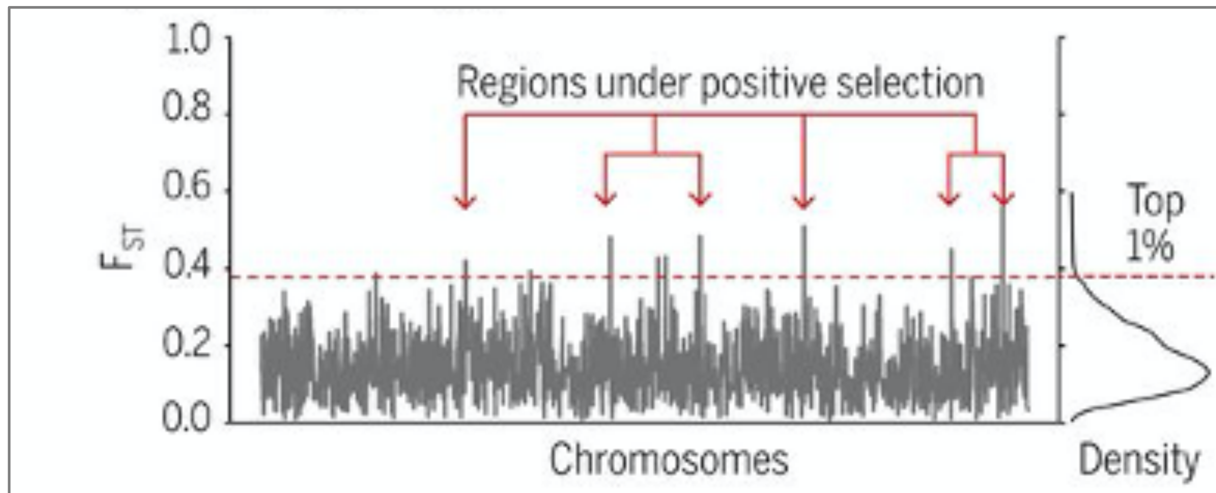


Figure 1.2: Genome scan of population differentiation, shown for the Fst outlier approach. Genomic regions under positive selection are identified as 'outliers' for high differentiation (e.g. top 1% Fst). SNP variants identified in regions under positive selection are candidate adaptive loci. Signatures of many candidate loci under positive selection are identified for polygenic traits. Figure reproduced from Fan et al. 2016.

Using Fst to measure population differentiation has often been included in suites of tests to detect genetic signals of adaptation. For example, Fst and XP-CLR testing co-identified candidate genes *EPAS1* and *EGLN1* for adaptation of Tibetan people to living at altitude (Xu et al. 2011). Ghanatsaman et al. (2023) performed pairwise Fst and nucleotide diversity (pi) testing of indigenous goats from west and south Asia vs European goat breeds, to detect selection signatures. They identified three candidate genes related to adaptation to hot, dry climates in Asian breeds from the top 1% Fst threshold as a cut-off. Among them candidate genes *CDH9* and *HERC6* had already been linked to environmental adaptation in cattle (Serranito et al. 2021) and *MSRB3* was linked to heat stress in pigs (Edea and Kim 2014). A further two candidate genes *RYR1* and *LRFN5,* relating to the immune response, met both the Fst and nucleotide diversity top 1% significance threshold.

1.3.4. Genome-environment association to infer candidate adaptive loci

Genome-environment association (GEA) analyses are powerful approaches that can be applied to the situation of having genotyped individuals from multiple populations and those populations varying for an environmental variable(s) of interest (Rellstab et al. 2015). These GEA approaches measure genetic differentiation between all of the populations at each locus, and then test to see which loci are most strongly associated with the environmental variable(s) (Rellstab et al. 2015; Hoban et al. 2016; Forester et al. 2018). GEA essentially extends Fst differentiation tests of population genomic data, by including environmental data, and these GEA tests can detect subtle allele frequency shifts across multiple loci due to weak selection; a characteristic of polygenic adaptation (Pritchard and Di Rienzo 2010; Berg and Coop 2014).

The environmental data can be collected in the field or sourced from climate data sets for specific time periods. See Schoeman et al. (2023) for a review of global climate models for use in life sciences. The underlying assumptions or limitations of a chosen climate data set are important to consider in the interpretation of the results of a GEA (Rellstab et al. 2015; Bernatchez et al. 2023). Single environmental variables may be run independently, but it has been demonstrated that constrained signatures of multi-dimensional environmental data sets, e.g. by first performing a principal component analysis (PCA) or particularly for using redundancy analysis RDA (Capblancq and Forester 2021), performs better at identifying true-positive loci under selection and reducing false-positive rates (Forester et al. 2018). For example, temperature, rainfall and wind were found to be the main environmental variables driving adaptation in natural populations of fruitfly *Drosophila melanogaster*, extrapolated from 59 environmental variables and a GEA approach (Bogaerts-Márquez et al. 2020).

Several GEA methods have been developed including BayEnv2 (Günther and Coop 2013), LFMM (Frichot et al. 2013), BayeScEnv (de Villemereuil and Gaggiotti 2015) and BayPass

(Gautier 2015), with Bay-/Baye- denoting a Bayesian framework. When a GEA is performed in any of the packages, a value of environmental association is produced for each considered locus, calculated by the chosen method (Rellstab et al. 2015). Some of the methods incorporate corrections to reduce false-positive reporting in different spatial scenarios, e.g. BayPass accounts for any autocorrelation of spatial processes or joint demography with environmental gradients (Gautier 2015). When the GEA is run in BayPass to identify environmental association, the output of Bayes Factor (BF) values presented on a Manhattan plot would look like **Figure 1.3**. The significance threshold for outlier identification in BayPass is >10 BF for 'good' evidence of association to the environmental variable and >20 BF for 'decisive' evidence (Gautier 2015). See Morey et al. 2016 for a review of how BF can quantify statistical evidence.

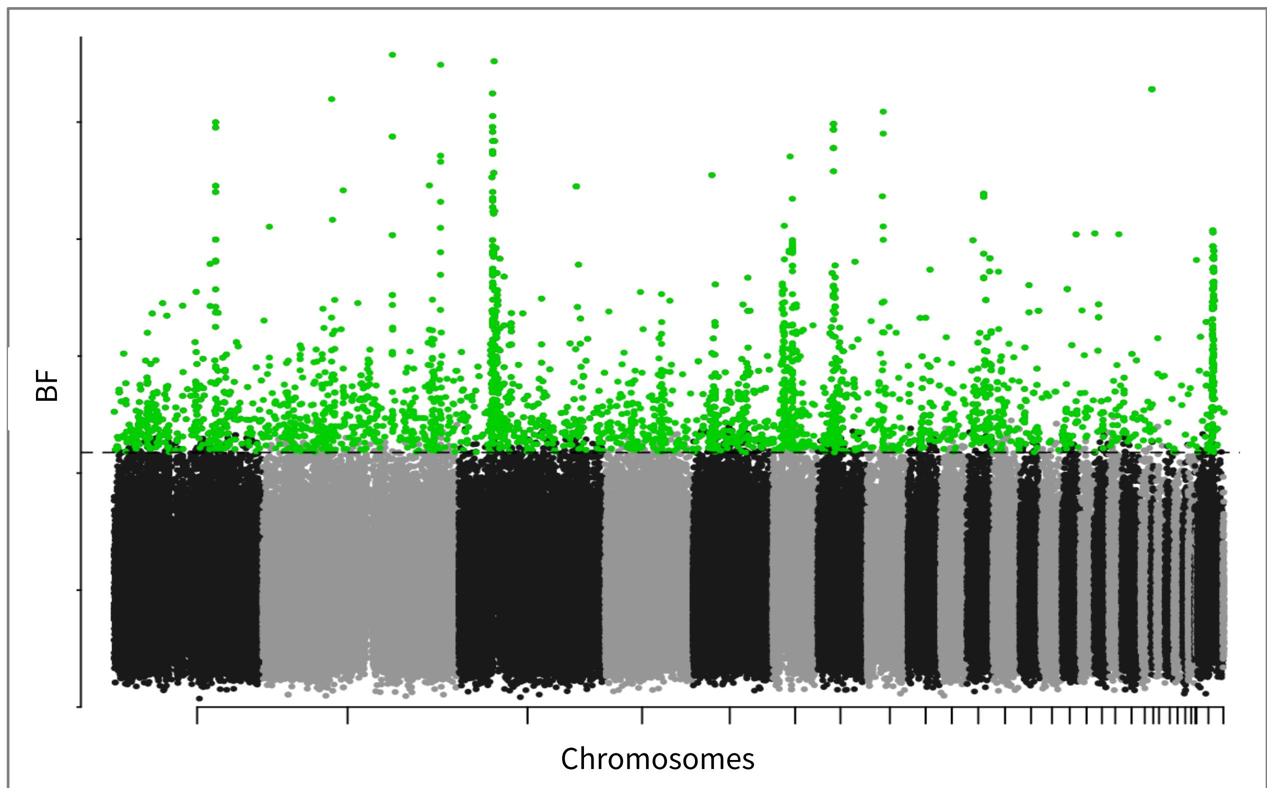


Figure 1.3: The genome-wide output of a genome-environment association analysis presented as a Manhattan plot. Outlier variants identified above the significance threshold for evidence of association to the environmental covariate of interest are shown in green. Data are from the GEA data analysis using BayPass described in Chapter 2.

When candidate adaptation loci are identified, a reference genome for the species can then be used to identify the closest gene(s) to each (Tiffin and Ross-Ibarra 2014). The produced set of candidate environmental adaptation genes can then be statistically tested for functional enrichment of the species' most up to date gene ontology (GO) terms, for biological process, molecular function or cellular component (Tomczak et al. 2018). An example of a published GEA study using BayPass methods, identified candidate climate adaptation genes in cattle, with candidate genes enriched for GO terms including UV protection, thermoregulation, heat stress and desiccation (Flori et al. 2018). Overall, in relation to climate change research, GEAs have a broad use for identifying candidate climate adaptation genes across species, and climate adaptations in common can be compared for functional enrichment (Waldvogel et al. 2020).

1.3.5. Common considerations for using methods to detect regions under selection

When interpreting any of the above methods of genome scans for detecting regions under selection or associated with environmental adaptation, it is necessary to consider that: 1) a small number of false-positives may be statistically generated e.g. in GEA methods (Gautier 2015), and a small number of true-positives could be missed. 2) Any identified candidate adaptation SNP may simply be in close linkage with the functional site under selection (Wang et al. 2010). 3) Candidate adaptation SNPs are only potential candidates because none of the genome scan methods directly determine the causal drivers of selection (Tiffin and Ross-Ibarra 2014; Wadgymar et al. 2022; Bernatchez et al. 2023).

1.4. Comparative candidate climate adaptation genes across vertebrates

The genome scan approaches to identify regions under selection, can be used as one method alone (e.g. GEA) or in a suite of combined cross-population EHH and Fst or GEA methods, to help narrow down sets of candidate genes underlying adaptation (Ahrens et al. 2018). A combination of population differentiation testing, Fst and pcadapt (Privé et al. 2020), and GEA using partial redundancy analysis (Capblancq and Forester 2021), detected multilocus climatic adaptation of bank voles in the UK to regional environmental differences, with GO term enrichment for many functions that link to heat or oxidative stress (Marková et al. 2023).

In published genome scans for selection, few candidate climate adaptation genes have been co-identified across vertebrate species (Franks and Hoffman 2012; Tian et al. 2020). Studies of domestic species have tended to identify candidate genes that are known to be associated with food production traits (Passamonti et al. 2021; Flori et al. 2018; Saravanan et al. 2021), presumably because they are detecting signatures of artificial selection/animal breeding. The few shared candidate climate adaptation genes appear to be related to adaptation to hot or hypoxic environments e.g. detected under selection in cattle breeds from warm regions (Flori et al. 2018; Ben-Jemaa et al. 2020; Saravanan et al. 2021) and as candidate genes for adaptation to heat-stress in chickens (Tian et al. 2020; Rachman et al. 2024).

1.4.1. Experimental validation of candidate genes

Experimental validation of the involvement of candidate adaptation loci with climate/environmental adaptation is desirable (Hoffmann et al. 2021; Bernatchez et al. 2023). Common garden experiments involve rearing individuals from different environments in a common, controlled setting e.g. an aviary, a field or a laboratory, which can be used to study, but not test, local adaptation processes (de Villemereuil et al. 2016). However, it may be difficult to rear some larger vertebrate species in laboratory controlled settings, especially for those with long generation times (Bernatchez et al. 2023).

In contrast, reciprocal transplants of individuals from populations experiencing different conditions can capture the fitness benefit for adapted individuals to test/prove local adaptation (with an overview of studies of this type reviewed in Wadgymar et al. 2022). Reciprocal transplant experiments might be difficult to execute in mobile or long-lived species due to logistical and safety constraints of conducting field experiments in the wild (Wadgymar et al. 2022). Despite the validation of candidate adaptation genes being an important next step, very few of the published candidate climate adaptation genes for vertebrates have had their fitness effects experimentally quantified (Pardo-Diaz et al. 2015; Hoffmann et al. 2021; Wadgymar et al. 2022).

1.5. Examining temporal evolutionary changes of adaptation

The genetic adaptation of a population to its experienced environment can be investigated through examination of adaptive allele frequency changes over time, utilising a time-series of sequenced museum specimens or sequencing individuals across multiple generations in an experimental evolution setting.

1.5.1. Museum time-series analysis of evolutionary change

Natural history collections can provide a genomic resource of species/population history, whereby animal specimens collected from wild populations across multiple time points can be genotyped or sequenced, to investigate temporal genomic changes (Clark et al. 2023). Museum specimens are typically from populations in the last 100 years (Holmes et al. 2016) and changes in allele frequency can also be compared to present day contemporary populations or even compared back to ancestral populations (Dehasque et al. 2020). A time-series of museum specimens can be used to investigate adaptive genetic changes to local environmental conditions, with adaptation indicated by a relationship between a change in

an environment variable and changes in loci under selection (Holmes et al. 2016; Clark et al. 2023). However, interpretation of allele frequency trajectories over time must reflect that changes in the environment may not necessarily be driving selection (Loog et al. 2017; Clark et al. 2023).

The availability of museum specimens for a given time-period is reliant upon what specimens were originally collected in the field, and how often populations were re-sampled (Holmes et al. 2016). Storage methods of museum specimens also influence the quality of DNA extracted for sequencing e.g. short read, and sometimes only low coverage, sequence data is often the only option for museum samples (Raxworthy and Smith 2021). Museums may also be wary of allowing DNA extraction from their collections. Potential biases in allele frequency change may be observed from museum time-series for a population of interest (Raxworthy and Smith 2021; Clark et al. 2023), which may be caused by demographic history and the confounding effects of gene flow e.g. historical bottlenecks, migration events or admixture when previously isolated populations interbreed (Rellstab et al. 2015). See Clark et al. (2023) for a recent comprehensive review of the use of temporal genomics for measuring evolutionary responses to climate change. Signatures of balancing, purifying and positive selection can be investigated using a time-series of museum or ancient paleo-specimens (Loog et al. 2017).

1.5.2. Experimental evolution over generations

In the setup of experimental evolution, species with very fast generation times, such as microbes, *Daphnia* or *Drosophila*, are reared in controlled laboratory conditions, with replicate populations facing different environmental conditions (Kawecki et al. 2012). Evolution in action for local adaptation can be observed in experiments of this type, with genome-wide responses to single and multi-dimensional selection and novel mutations trackable across hundreds of generations (reviewed by White et al. 2022). However, the genetic responses to environmental agents seen in population lines from model species in

simplified, highly controlled experimental setting may not always transpire in natural populations with interacting evolutionary pressures and environmental heterogeneity (Kawecki et al. 2012).

Candidate loci identified under selection in an experimental evolution setting need to be functionally validated, by either experimental confirmed phenotypes which can be time-consuming to conduct, or by a second evolve and resequence experiment - where crossing an evolved population with ancestral founders is expected to see a repeatable frequency increase of the candidate allele when exposed to the same environmental conditions (Burny et al. 2020).

1.5.3. Local adaptation and genomic vulnerability for population conservation

Genomic investigations of how past selection has affected spatial and temporal allele frequency changes in response to a changing climate, are providing new insights into local adaptation (Wadgymar et al. 2022; Bernatchez et al. 2023; Clark et al. 2023). The identification of environmental adaptation loci may help inform population genomic vulnerability assessments (Hoffmann et al. 2021). If populations are assessed to be exhibiting low genetic variation, such as for fragmented or range edge populations, conservation genetic practices may be recommended to conserve the adaptive potential of these populations to be able to respond to future climate changes (Nosil et al. 2019).

1.6. Great tit study species

The remainder of this introduction chapter will now focus on the great tit study system and the investigation of their possible genomic adaptation to climate.

1.6.1. Great tit species complex

The great tit *Parus major* (order passerine) is a terrestrial hole-nesting songbird, distributed across Eurasia (van Balen 1973; BirdLife International 2024). The *Birds of the World* database (Gosler et al. 2020) describes a one species complex for *Parus major*, as there is insufficient genetic evidence to regard the great tit (*P. major)*, Japanese tit (*P. minor)* and their hybrid crosses in far East Russia and China, along with the cinereous tit (*P. cinereus)* as separate species. See Johansson et al. (2013) for a published phylogeny of *Parus* species based on nuclear genes *Myoglobin* and *ODC*, and the mitochondrial gene *ND2*. In addition, investigation of mitochondrial differences in *Parus major* by Song et al. (2020) identified one mitochondrial group across Europe and four mitochondrial groups across Asia. They discussed that the greater diversity across Asia might reflect the greater range of habitats provided by the Himalayas and may suggest the origination of *Parus major* in Southeast Asia (Song et al. 2020).

1.6.2. Population distribution of great tits

The range of the great tit subspecies groups are displayed across Eurasia (**Figure 1.4**). The *Parus major major* group has the widest distribution area of all the groups, with range edges into Russia, North Africa, and Iran. Different populations of *Parus major* are often classified by island or geographic distribution e.g. *Parus major corsus* (Iberian peninsula and Corsica), -*mallorcae* (Balearic Islands), -*ecki* (Sardinia), -*niethammeri* (Crete), -*aphrodite* (Mediterranean), -*excelsus* (Northwest Africa), -t*erraesanctae* (Northwest Syria, Lebanon, Israel and Jordan), -*blanfordi* (North Iraq and Iran), -*kapustini* (Northwest China), -*tibetanus* (Southwest China to Tibet), and -*intermedius* (Northeast Iran and adjacent to SW Turkmenistan) (Clements et al. 2022). *Parus major newtoni* is only found in the UK (**Figure 1.5**), distinguished from the morphologically similar *-major* that is also present in the UK, by longer bill lengths in *newtoni* (Kvist et al. 1999; Kvist et al. 2003).

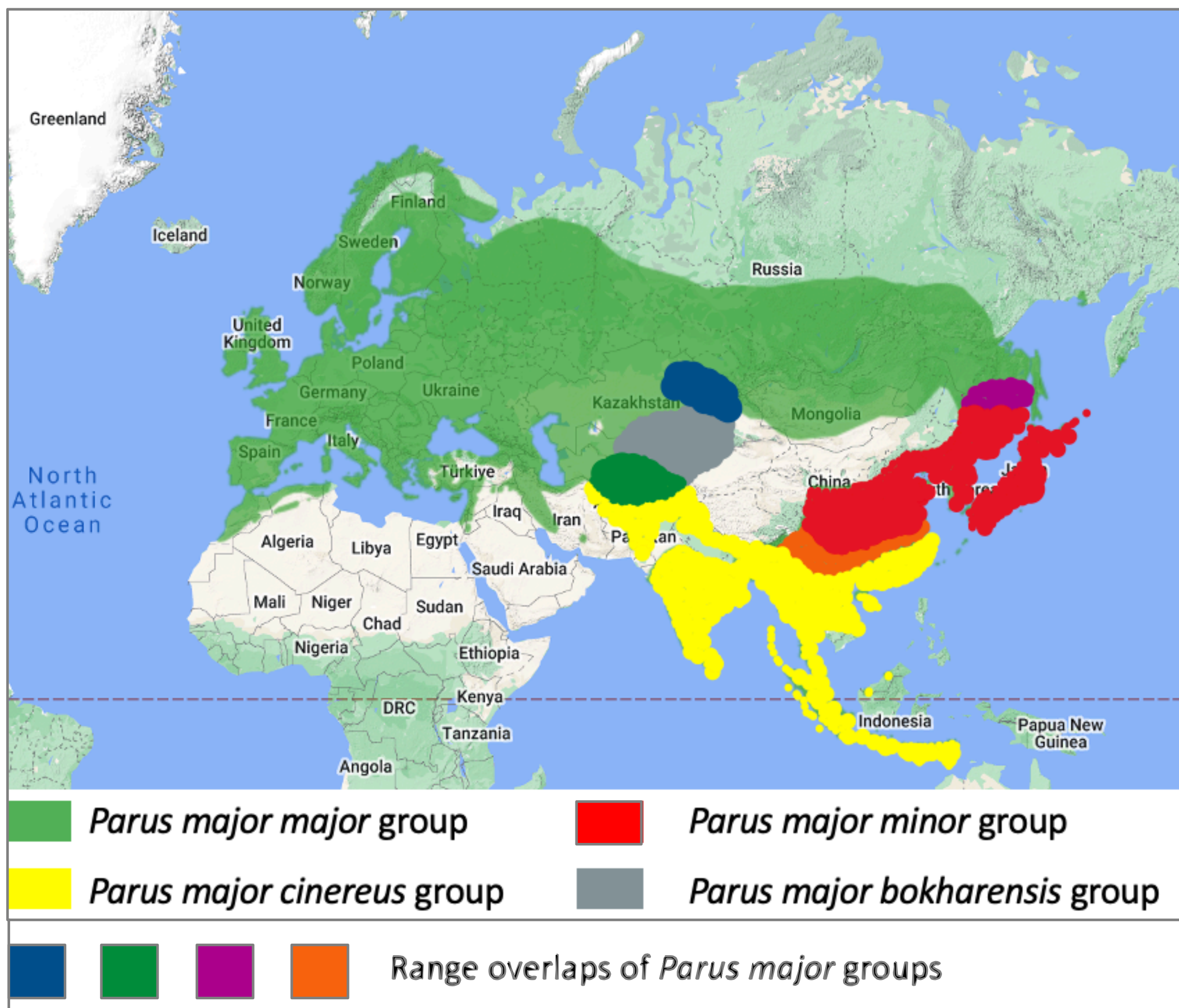


Figure 1.4: *Parus major major*, *-minor*, *-cinereus* and *-bokharensis* group distributions across Eurasia (interpreted from Kvist et al. 2006). Base map figure reproduced from BirdLife International (2024) Species factsheet: *Parus major*. Downloaded from http://datazone.birdlife.org/species/factsheet/great-tit-parus-major.

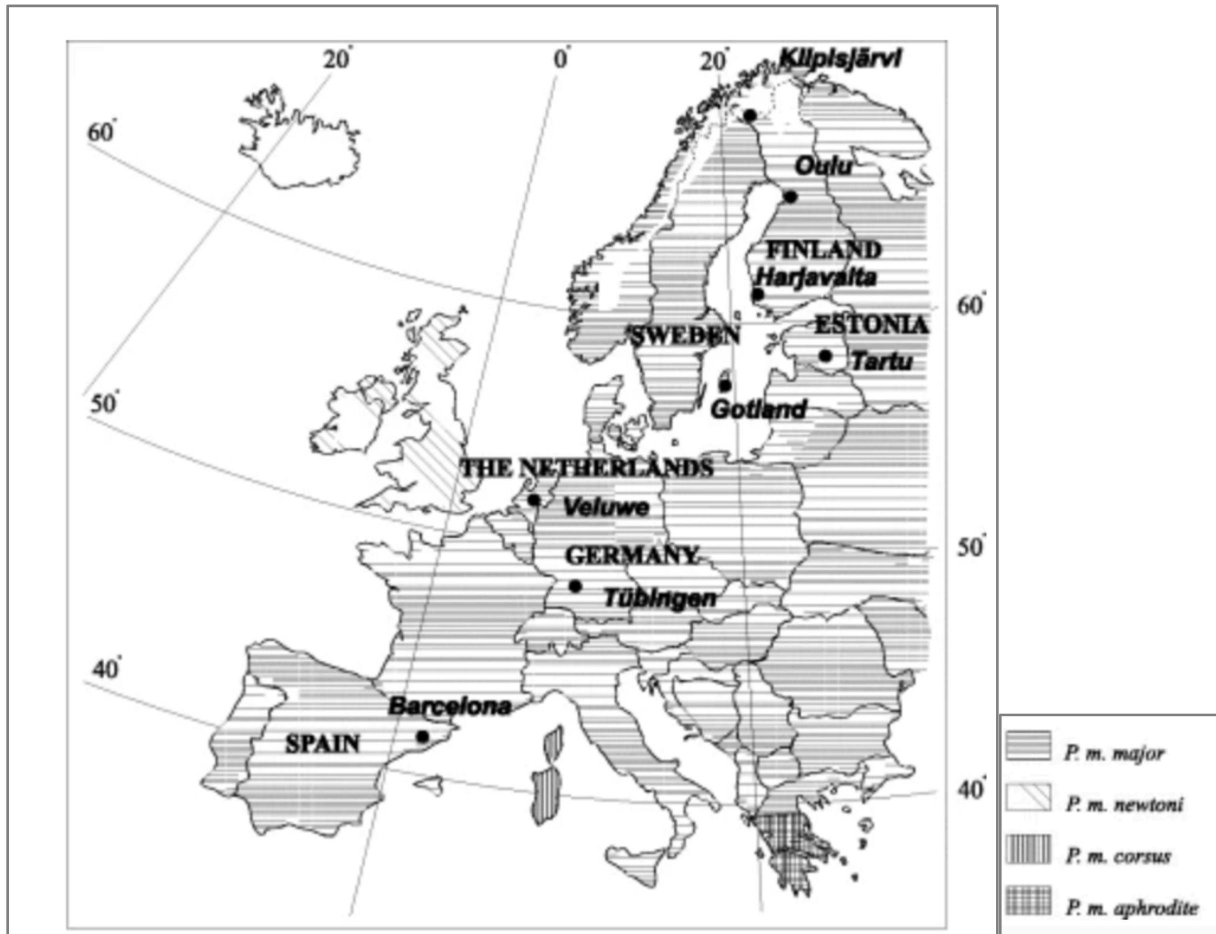


Figure 1.5: Distribution of Parus major subspecies for sampled European great tits, reproduced from Kvist et al. 1999.

A reconstruction of great tit European colonisation using the HapMap samples (see section 1.8.5) suggests that the current European population distribution has likely spread from a single restricted refugium in South-Eastern Europe after the last ice age (Spurgin et al. 2024). Contemporary wild populations of great tit experience a wide range of climate conditions across Eurasia (BirdLife International 2024), which suggests that they have locally adapted to them since the last glacial period (Spurgin et al. 2024). The potential for adaptation to many different climate conditions justifies the choice of the great tit as a study system to investigate the genomic signatures of climate adaptation, rather than an alternative vertebrate species with a narrower range of experienced climate conditions.

1.6.3. Great tit dispersal and gene flow

In Europe, and especially the south of Europe, the great tit is resident i.e. they do not migrate, after establishing territories in or near their natal site (Gosler 1993). Migration is considered rare, although it is sometimes recorded between Northern European populations (Kvist et al. 1999). The recorded migration routes of great tits over Europe have been mapped on an online migration atlas from bird ringing records: https://migrationatlas.org/node/1850 (**Figure 1.6**). Many of the populations across Europe contain mostly resident and native of origin individuals (Gosler 1993), but longer distance migration has been observed in the UK population in response to extremely harsh conditions (Kvist et al. 1999). The migration map illustrates that dispersal between great tit populations does occur, with some birds migrating long-distances from the Baltic region, suggesting some facilitation of gene-flow between populations via highly mobile individuals. However, some individual birds may migrate over large distances, but then return to native natal grounds to breed, such that gene-flow does not occur (Prof. Ben Sheldon and Prof. Marcel Visser pers. com.).

Clearly though, rare migration, has been enough to ensure Fst is very low between geographically distant populations e.g. low differentiation (Fst <0.05) between UK and NL populations (Van Bers et al. 2012; Bosse et al. 2017; Spurgin et al. 2024). The large effective population sizes of great tits found in continental Europe (Van Bers et al. 2012) and low levels of differentiation is suggestive of a panmictic population across Europe (Spurgin et al. 2024).

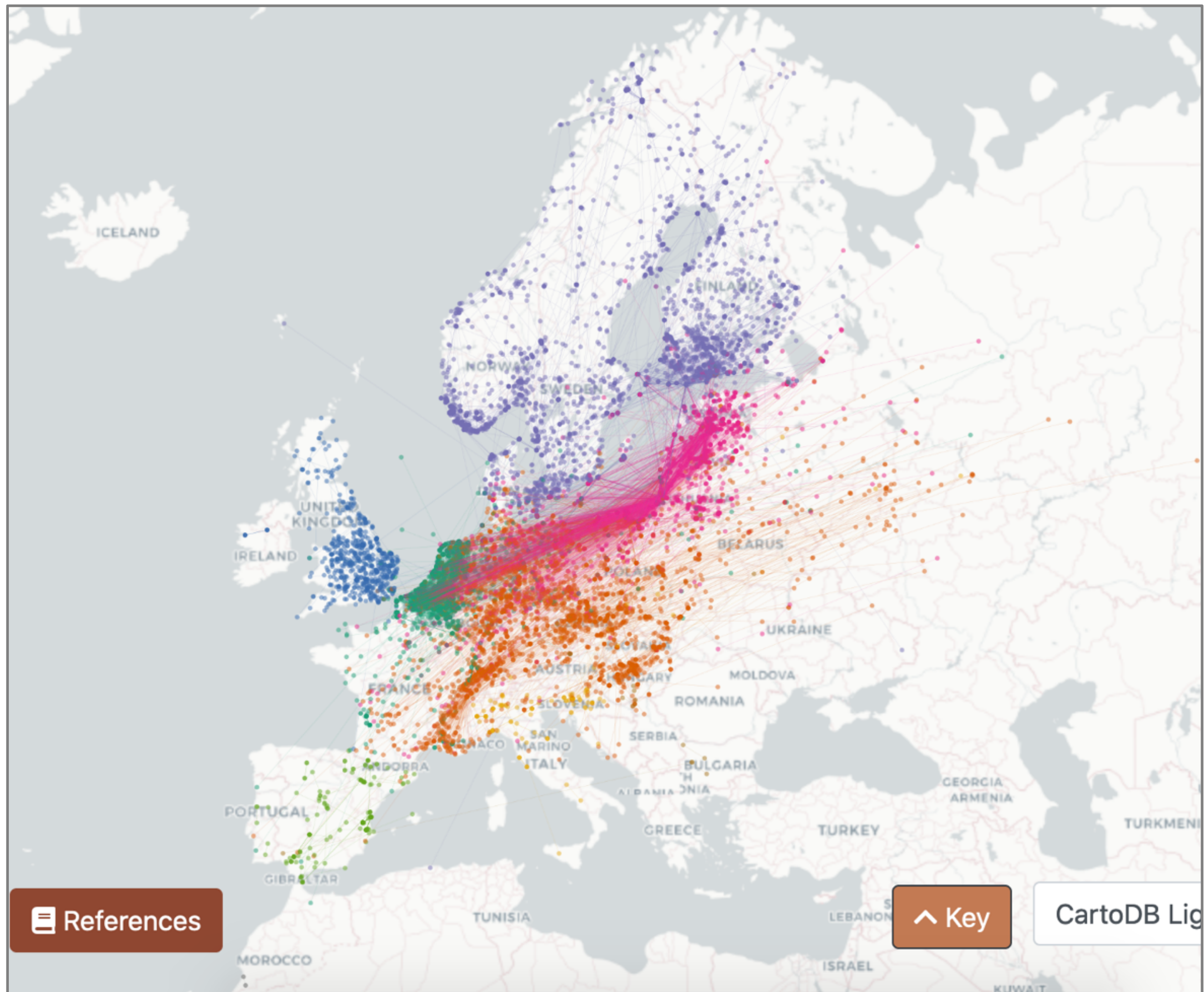


Figure 1.6: Great tit migration flights lines recorded across continental Europe. Screenshot from The Eurasian-African Bird Migration Atlas for the great tit: https://migrationatlas.org/node/1850 (downloaded in Dec 2023). The data is based on the European Union for Bird Ringing (EURING) records (ring recoveries and tracks) - https://euring.org. The Atlas was developed in collaboration with the Bonn Convention (Convention on Migratory Species) by the European Environment Agency and funded by the Italian Government.

1.6.4. Demographic trends in contemporary European great tit populations

The number of breeding pairs of *P. major* is stable in the European Breeding Bird Atlas 2 (EBBA2 2022). Although, in the future this may follow a general trend of moderate bird declines in woodland, farmland and urban birds across Europe since 1980, attributed to raised average temperatures and the use of pesticides and fertilizers (Rigal et al. 2023). Halupka et al. (2023) produced a global meta-analysis of climate change effects on avian offspring production (1970-2019) for 200 populations of 104 bird species. Over the study period, significantly fewer offspring were produced in 17.4% of populations, and significantly more offspring in 10.4% of populations. Climate change was suggested to influence offspring production, either positively or negatively, through its compounded effects on ecological and life history traits of species. For example, smaller sedentary species tended to produce more offspring under global warming because of warmer chick-rearing periods. Specifically for eight great tit nest-box populations (29,590 clutches, studied over 2-5 decades) that were included in the described meta-analysis, the four populations from protected areas were estimated to produce fewer offspring due to climate change, while the other four populations produced more offspring than expected (Halupka et al. 2023).

1.6.5. Great tit ecological model system

Wild populations of great tits that readily breed in provided woodland nest boxes are an excellent ecological model system to record phenotypes in the wild and monitoring population dynamics. Three significant long-term great tit field study sites in Europe are located in: 1) Wytham Woods, Oxford, UK (Lack 1964), which has grown from 100 to 1000 nest boxes, recorded for 75 years (Sheldon et al. 2022). 2) Hoge Veluwe, Netherlands (NL), on the Dutch mainland and 3) Vlieland, NL, on a North Sea island. Both sites in the NL, have monitored ~400-450 nest boxes since 1955 (with both study areas described in Tinbergen and Boerlijst 1990). Local environmental and climate conditions of the field study sites are measured each year.

Field observation of egg lay date and chick development in nest boxes, and sighting and re-capture of ringed great tits over their ~4 year lifespan (van Balen 1973), can provide life history trait information on an individual's health and reproductive success/fitness (McCleery et al. 2004; Santure et al. 2013; Sheldon et al. 2022). Social pedigrees may be constructed for monitored populations of great tits, but these would likely contain errors due to extra pair paternity e.g. ~15% in the Wytham population (Patrick et al. 2012). Blood can be sampled for DNA sequencing, which can be used to investigate relatedness, population differentiation and signatures of selection (see Bello et al. 2023 for a review of studying selection signatures and identifying candidate genes using whole-genome data in chickens).

Behaviour of individuals in a field study population can be directly observed. For example, field studies researchers at Wytham Woods, Oxford have radio-tagged populations of tit species to monitor the use of feeding stations to investigate foraging behaviour and social dynamics (Farine and Lang 2013; Aplin et al. 2015; Firth et al. 2016). Further investigation of foraging in mixed species flocks has revealed dominance of great tits towards blue tits, and avoidance of marsh tits if great tits are present at feeding stations (Maziarz et al. 2023).

Long-term monitoring of great tit study populations may reveal trends across different populations over time; for example any changes in life-history traits or any influencing changes in community interactions e.g. disease/parasite infection, competition with other species, and predation levels (Sheldon et al. 2022). A wide range of diseases and parasites are known to infect wild great tit populations e.g. bird flu (Wade et al. 2022), malaria (Isaksson et al. 2013), *Salmonella* in nest boxes and bird feeders (Krawiec et al. 2015; Boonyarittichaikij et al. 2017), and Trichomonosis disease caused by a protozoan parasite *Trichomonas gallinae* (Robinson et al. 2010), which affects fecundity and mortality in infected bird populations (Lachish et al. 2011). These long-term field studies make great tits an excellent ecological model system to investigate population ecology, evolution and behaviour, and community dynamics over multiple decades of time.

1.7. Local adaptation and climate change

1.7.1. Climate change and advancement of egg lay date

Great tits typically lay their first brood of 7-9 eggs in April-early May, with second broods possible later in the summer (van Balen 1973). The optimal timing (phenology) of egg lay date and hatching date of great tit chicks is associated with spring synchrony of the tri-trophic (oak tree-winter moth-great tit) food web, to raise great tit chicks with plentiful caterpillars (Perrins 1970; Visser et al. 1998; Hinks et al. 2015). Earlier egg laying in great tits has been associated with warmer recorded spring temperatures over the last six decades in the UK (see Figure 1 of Cole et al. 2021). An advanced egg lay date has also been observed in the Hoge Veluwe, NL study population, where a phenological mismatch in peak caterpillar prey has led to directional selection for earlier seasonal breeding (Visser et al. 2021). The advancement of egg lay date in great tits may phenologically match or mismatch the tracking of advancement in peak caterpillar abundance (**Figure 1.7**).

In recent years in the Hoge Veluwe great tit study population, the caterpillar peak abundance date has advanced no further, while the date of the mean great tit egg laying date has got earlier, thereby reducing the current phenological mismatch (Visser et al. 2021). It remains to be seen if the earlier egg lay date in great tits can continue to match peak caterpillar resources for chicks with future climate changes; as mis-matched timings could inhibit the growth, development and survival of clutches. Future climate projections of the Hoge Veluwe great tit study population suggest that phenological mismatches will increase, as advancements in the caterpillar abundance peaks will likely outpace earlier egg laying dates (Visser et al. 2021).

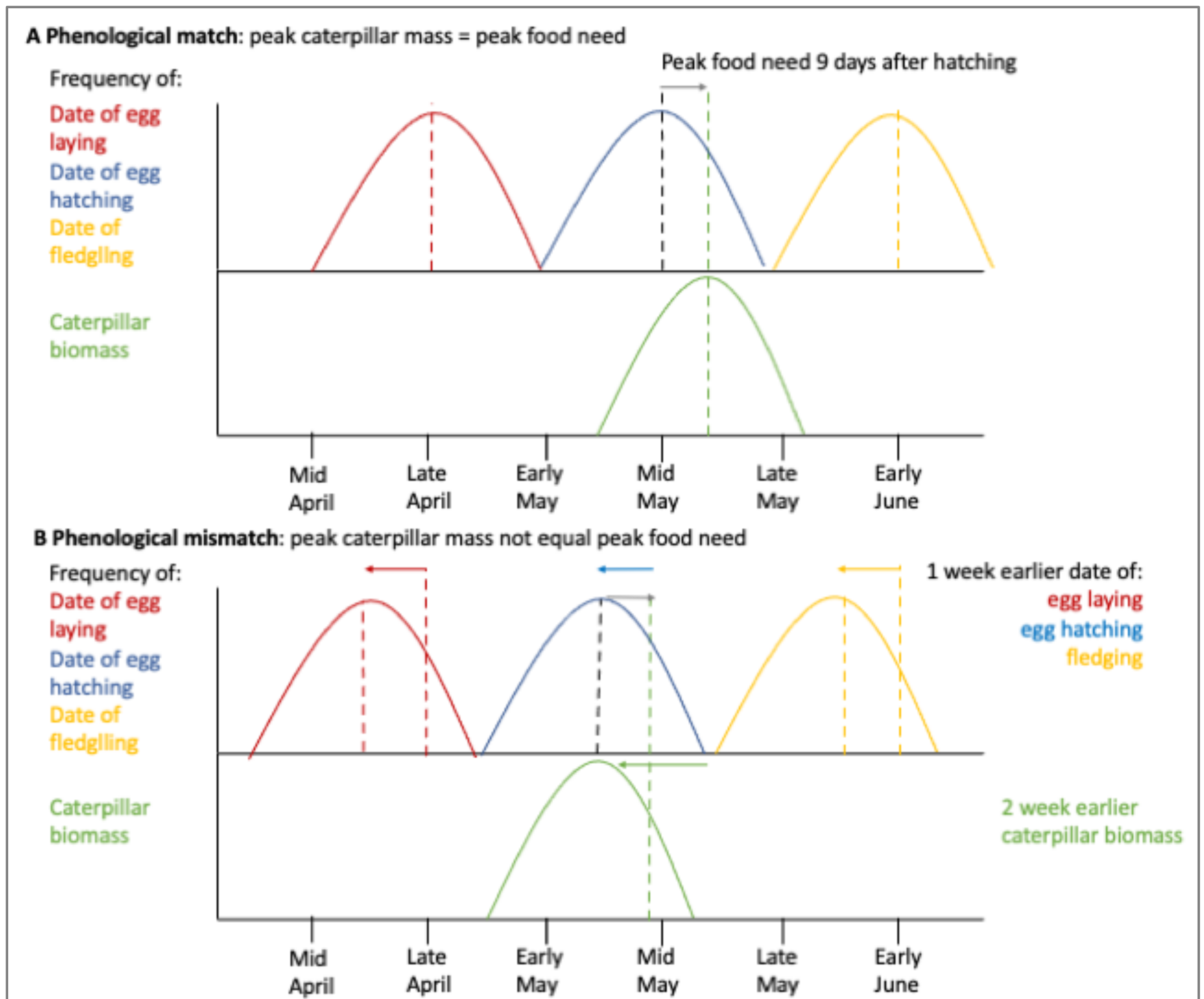


Figure 1.7: A scenario of phenological match and mismatch of egg lay date and caterpillar abundance for rearing great tit chicks. In panel A, the timing of egg lay date (red), egg hatching (blue) and fledgling (yellow) is matched for the peaks of great tit chick's food demands and caterpillar abundance. In panel B, a scenario of climate change is presented, where warmer Spring temperatures have influenced one-week earlier egg lay and hatching dates, with twice the advancement in peak caterpillar biomass. The timing of egg lay date is now mismatched for the peaks of chick's food demands and caterpillar availability. Figure interpreted from Charmantier and Gienapp (2014) and Mizumo Tomotani (2017).

Climate change may also contribute to egg hatching failure (Assersohn et al. 2021) and first year mortality of great tits. Unfortunately, the majority of great tits do not survive their first year, due to environmental stress factors, including heat or drought stress, food shortages, competition, predation, disease or injury (Clobert et al. 1988; Lindner et al. 2023), and these factors will likely be negatively influenced by climate changes.

1.7.2. Long-term field data to ask evolutionary questions over time

The observation of natural populations of great tits, including at the three long-term study sites in the UK and NL, requires yearly intensive fieldwork in the breeding season, and ongoing funding for project and staff costs (Sheldon et al. 2022). However, long-term field study sites are described as having added value as they age (Clutton-Brock and Sheldon 2010), and they lend themselves to investigations of temporal evolutionary questions for changes in a focal population. The genomic data set collected from long-term field studies of great tits (see **Section 1.8**) may help researchers to understand the plastic and microevolutionary processes involved in earlier egg laying and climate change - (with these processes reviewed by Charmantier and Gienapp 2014). Investigating long-term population genomic data sets may give insight into the genomic basis of adaptation in birds (Campagna and Toews 2022), towards understanding the genomic processes involved in local adaptation of populations to different environmental/climate conditions experienced over time.

1.7.3. Local environmental conditions

Local environmental conditions are unique to each population of great tit, with a range of considerable variation in a) Temperate habitat types e.g. boreal forest, woodland, grassland, arable, urban, rural gardens, shrubland (BirdLife International 2024), except for areas of tundra, steppe and desert (Eguchi 1980). b) Altitudes e.g. recorded at 0-4420m (BirdLife

International 2024). c) Urbanisation e.g. cities on hot days are reported to reach temperatures 10-15 °C higher than surrounding rural areas (Mentaschi et al. 2022). City living great tits are more likely to experience heat stress, higher pollution levels and reduced insects in their diet than rural counterparts (Charmantier et al. 2017; Corsini et al. 2020). d) Experienced climate conditions e.g. spatially-separated populations will experience different local climatic variables in the present day or they may have experienced different climatic conditions temporally over their population evolutionary history (Jensen and Leigh 2022).

1.7.4. Local adaptation to urbanisation

In urban environments, passerine birds tend to have smaller clutches, lighter and smaller offspring (Chamberlain et al. 2009), paler feathers (Salmón et al. 2023), and a more generalist diet because of fewer insects (both in terms of species-types and abundance) that can be supported by non-native trees (Jensen et al. 2023). Urban infrastructure is associated with air pollution and/or light pollution that can inhibit moths and the biomass of caterpillars as a food source for birds (Boyes et al. 2021). Exposure to light of different colours at night has been observed to affect daily activity levels in blue tits (de Jong et al. 2017).

Genomic adaptation to urbanisation in great tits is reported in Salmón et al. 2021, with the epigenetics of different pairs of compared urban vs forest great tit populations showing different methylome responses (Caizergues et al. 2021). Great tit nestlings at two forest sites in Hungary were more vulnerable to extreme hot weather than conspecifics in the city (Pipoly et al. 2022). The authors of this forest vs urban comparison suggest that urban individuals might be better adapted to lose heat (e.g. altered plumage, small body size and brood reduction) and/or better at parental provision in the heat. This is similar to the finding that across 54 species of tropical butterflies in Panama, species can be good at physiologically tolerating high temperature or at adopting behaviours to avoid high

temperatures; these strategies for dealing with high temperature can be at the expense of each other, impacting how climate change effects them (Ashe-Jepson et al. 2023).

1.7.5. Local adaptation to past climate changes

Understanding the genetic processes of local adaptation to past climate changes is an important focus in Evolutionary Biology. Here in this thesis in **Chapters 2-4**, genome-wide adaptation to the locally experienced climate across contemporary populations of European great tit is to be investigated and discussed in detail. The aim is to identify candidate adaptive loci for climate adaptation.

1.7.6. Forecast climate change of the future

The sixth assessment synthesis report on anthropogenic climate change predicts rising terrestrial rising air surface temperatures (+1.5°C by early 2030). Along with increasing extreme weather events, including more intense heatwaves in the summer (see **Figure 1.8** for forecast heatwaves in continental Europe projected 2040-2060). This latest IPCC synthesis report also predicts for the mid-century increased drought stress, floods caused by heavy rain episodes or hurricanes, and more wildfire (IPCC 2023). Each of these predicted climate changes could threaten populations of wild species across the globe (Parmesan and Yohe 2003).

The Met Office recorded 2023 as the warmest year on record for global average temperatures (https://tinyurl.com/53fyz8ed). Furthermore, an example of a climate change induced population crash was observed in June 2019, where ~200 great tit nestlings died following an extreme 43.5°C heatwave in Montpellier, France (reported in the Independent newspaper https://tinyurl.com/48zdadcf). The chicks may have died from heat-stressed physiology or because the parents were unable to provide enough food in such heat-stressed conditions.

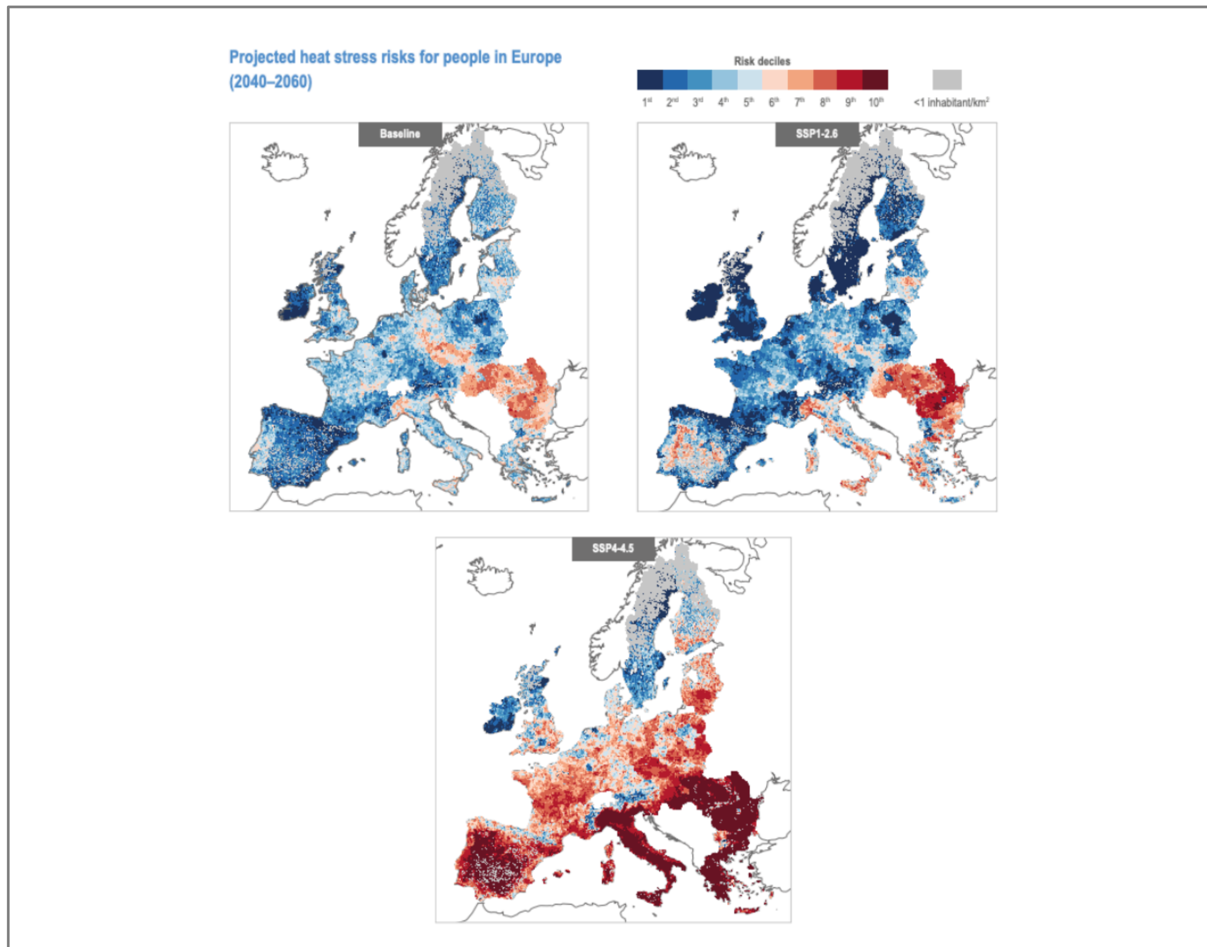


Figure 1.8: The risk of heat stress for humans is projected to increase in the eastern Mediterranean and southern Europe from baseline 1986-2005 (left hand figure) for two socioeconomic scenarios (SSP1.2.6 and SSP4.4.5) incorporated in the climate change models for 2040-2060 (right and lower figures, respectively). The heat stress risk value is calculated by aggregation of heatwave days, population vulnerability and exposure (Bednar-Friedl et al. 2023). Thus, regions predicted with more frequent and intense heatwaves like the Mediterranean likely pose heat-stress risks to the health of both people and wild animals in the mid-21st Century (Wedler et al. 2023). However, the realised impact of climate change will depend on strategies for large-scale reduction in greenhouse gas emissions to limit global warming (IPCC 2023). Figure reproduced from Bednar-Friedl et al. (2023).

1.8. Great tit Genome

1.8.1. Great tit reference genome

The great tit reference genome (Laine et al. 2016) was produced from very high coverage (95x) sequencing and de novo assembly of an individual captive male great tit from the Netherlands; published with 32 pairs of chromosomes in the nuclear genome (including the homogametic male sex chromosome in birds ZZ), 1 mitochondrial organelle and the DNA methylome made available (the current version of the reference genome GCF_001522545.3 is downloadable from https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001522545.3/. Chromosomes in passerine genomes are numbered consistently with the chicken genome, rather than indexed by chromosome size (Ellegren 2013). The avian W sex chromosome (heterogametic in females i.e. ZW) was not sequenced in the great tit reference genome, to provide greater resolution of the Z chromosome (Laine et al. 2016). The karyotype of the great tit genome is highly conserved with other avian species (**Figure 1.9**), with only a small number of chromosome inversion rearrangements (Ellegren 2013; Santure 2016) relative to the zebra finch genome (Warren et al. 2010). Notably, a large (~64Mb) inversion polymorphism has been identified across 90% of Chromosome 1A in the great tit genome, encompassing almost 1000 genes (da Silva et al. 2019).

The increasing number of assembled and annotated bird genomes increases the power of comparative genomics to ask questions across species. For example, the Bird 10,000 Genomes (B10K) Project (http://b10k.genomics.cn/) (Zhang et al. 2014), has reached its second milestone of releasing 363 bird genomes, representing 92% of all bird families (Feng et al. 2020). While the Darwin Tree of Life project is sequencing UK species including birds, providing reference genomes of key species within animal and plant communities (https://www.sanger.ac.uk/programme/tree-of-life/).

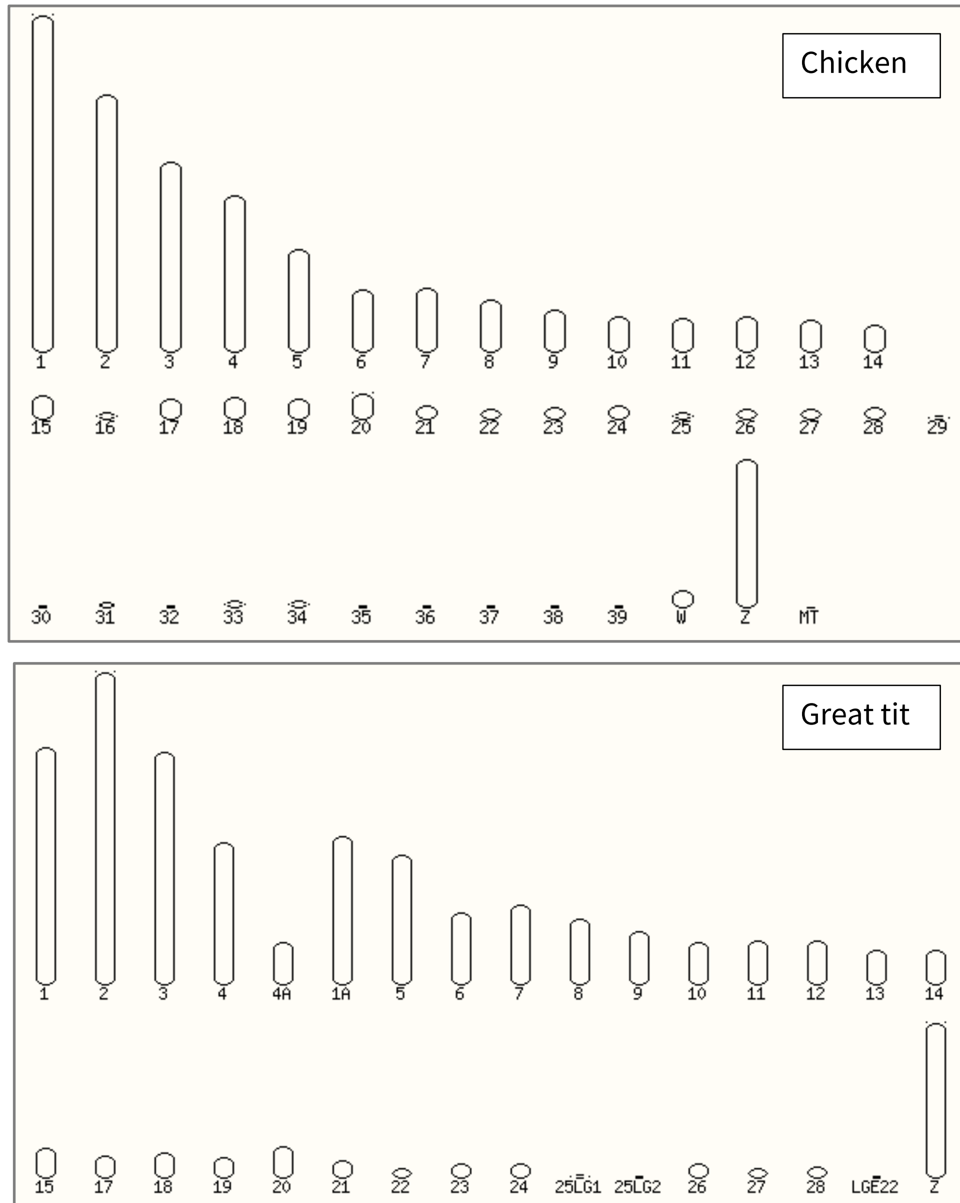


Figure 1.9: Avian Karyotypes are highly conserved. The top panel represents the 39 chromosome pairs of the karyotype from the GRCg7b broiler chicken reference genome (https://www.ensembl.org/Gallus_gallus/Location/Genome; Huang et al. 2023), with 10 macrochromosomes and 10 tiny dot chromosomes included in the 29 microchromosomes (<20 Mb). The lower panel displays the 32 chromosome pairs of the karyotype from the great tit reference genome (https://www.ensembl.org/Parus_major/Location/Genome; Laine et al. 2016), with 15 macrochromosomes (chromosomes 1-12, 1A, 4A and Z) and 17 microchromosomes identified.

1.8.2. 10K SNP chip genotyping

In 2012, a 10K SNP chip was developed for the great tit genome using an Illumina BeadArray technology (Van Bers et al. 2012), with 9193 SNPs evenly spaced across the nuclear genome. The identified markers were used to genotype 2700 great tits from the UK population (Wytham Woods, Oxford) and 2000 from the NL population (Hoge Veluwe and Westerheide), with high genotyping call rates of 0.98 and 0.93 respectively. Approximately 1100 loci showed differentiation between the populations, measured by weak but significant Fst (Fixation Index) values. Gene ontology (GO) enrichment analysis revealed these loci were associated with an overrepresentation of cardiovascular development genes (van Bers et al. 2012).

High-density linkage maps for the UK and NL were later produced by van Oers et al. (2014) from the genotyped UK and NL great tit populations. The two independently produced maps contained over 6,500 SNPs in 32 linkage groups, with similar marker order and distances in each population. The genome-wide male and female map lengths for each population were similar, interpreted as having little difference in genome-wide recombination rates between the sexes. However, some subtle sex differences in recombination rates were observed locally within and between chromosomes i.e. subtle heterochiasmy (van Oers et al. 2014).

The development of the great tit 10K SNP chip (Van Bers et al. 2012) and the great tit linkage maps (van Oers et al. 2014) allowed Santure et al. (2015) to investigate the genetic architecture of quantitative traits. They used replicate UK and NL wild populations of great tit, investigated using four approaches: chromosome partitioning, quantitative trait locus (QTL) mapping, genome-wide association study (GWAS), and by estimating the number of SNPs contributing to variation. All four approaches gave support that eight quantitative traits appeared to be polygenic and hence influenced by many genes of small effect distributed throughout the genome, with no large effect genes detected. These phenotypic traits included clutch size, egg mass, fledgling weight (of offspring), adult weight, fledgling weight (of individual), tarsus length, wing length and exploratory behaviour (Santure et al.

2015). The authors discussed that the great tit has a very large effective population size and correspondingly low levels of linkage disequilibrium (LD). They suggested that because of low LD, describing the genetic architecture of polygenic traits would likely require genotyping far more loci e.g. 500K, and for methods that need phenotypic data, far more sampled individuals would be needed than might be feasible to conduct in the field.

1.8.3. Whole-genome sequencing

In addition to describing the great tit reference genome and methylome, Laine et al. (2016) resequenced at ~10x coverage the whole genomes of 29 wild great tits from 25 European populations (including ten birds from Wytham Woods, Oxford). This greater resolution of sequencing relative to a SNP chip, was used to investigate population differentiation, genome-wide diversity, regions under selection, substitution rates and GO enrichment analyses of genes under selection. The authors found evolutionary signals of selection on cognition, interpreted from an overrepresentation of genes related to neuronal functions of learning and cognition in regions under positive selection (Laine et al. 2016).

1.8.4. 500K SNP Chip genotyping

In 2018, a higher density SNP chip was developed using Affymetrix Axiom technology (Kim et al. 2018), providing a rapid and accurate genotyping tool of great tit population differentiation. The 500K nuclear-wide SNP markers were identified from the 29 whole-genome sequencing data and one extra great tit from near Zurich with lower ~5x coverage. The intersection of pipelines for SNP discovery using GATK, SAMtools and ANGSD methods in the multiple populations identified 1.74M SNPs before filtering, with 610,970 SNPs selected in total. The pipeline of SNP variant discovery found: i) SNPs discovered on other platforms were more likely to be converted into working assays. ii) those with higher call rates had less genotyping error rates. iii) SNP discovery using multiple populations minimized the allele frequency ascertainment bias (Kim et al. 2018).

Bosse et al. (2017) applied population and quantitative genomics approaches using the 500K SNP chip, to identify loci under divergent selection for variation in length of bills, between the great tit population in Wytham (UK) compared with two Dutch populations (NL). A genome-wide association study (GWAS) was performed with the first principal component (PC1) of the genotype bill length data run as the phenotype. The highly significant outlier regions were potentially involved in the genomic architecture of bill length, and contained 28 annotated candidate genes involved in palate development, and skeletal development and morphogenesis. The COL4A5-C "long-billed" haplotype was found to be under selection in the UK, consistent with the history of supplementary bird feeding in UK gardens. However, Perrier and Charmantier (2019) drew caution in response to the paper's interpretation that bird feeders are driving the selection for longer bill length, because the candidates may be more likely to be detected by the statistical methods in regions of low recombination.

In addition to genotyping SNP markers, the 500K SNP chip was applied to identify structural or copy number variation (CNV) in great tit populations. For example, Barton and Zeng (2019) quantified purifying and positive selection acting on short insertion and deletion (INDEL) variation in coding regions (using ten high coverage resequenced European great tits from Corcoran et al. 2017).

1.8.5. The Great Tit HapMap project – genotyped populations

The great tit HapMap project brought together ~2000 field samples of great tits spanning much of the species range, with 20-50 individuals per population, and used the 500K SNP chip (Kim et al. 2018) to produce a genotyped data set (see Spurgin et al. 2024 for full details of the collaborators, field sampling and sequencing methods). The SNP variants were quality control filtered by Spurgin et al. (2024) in PLINK (Purcell et al. 2007) and retained 647 individual great tits from 29 populations (**Figure 1.10**).

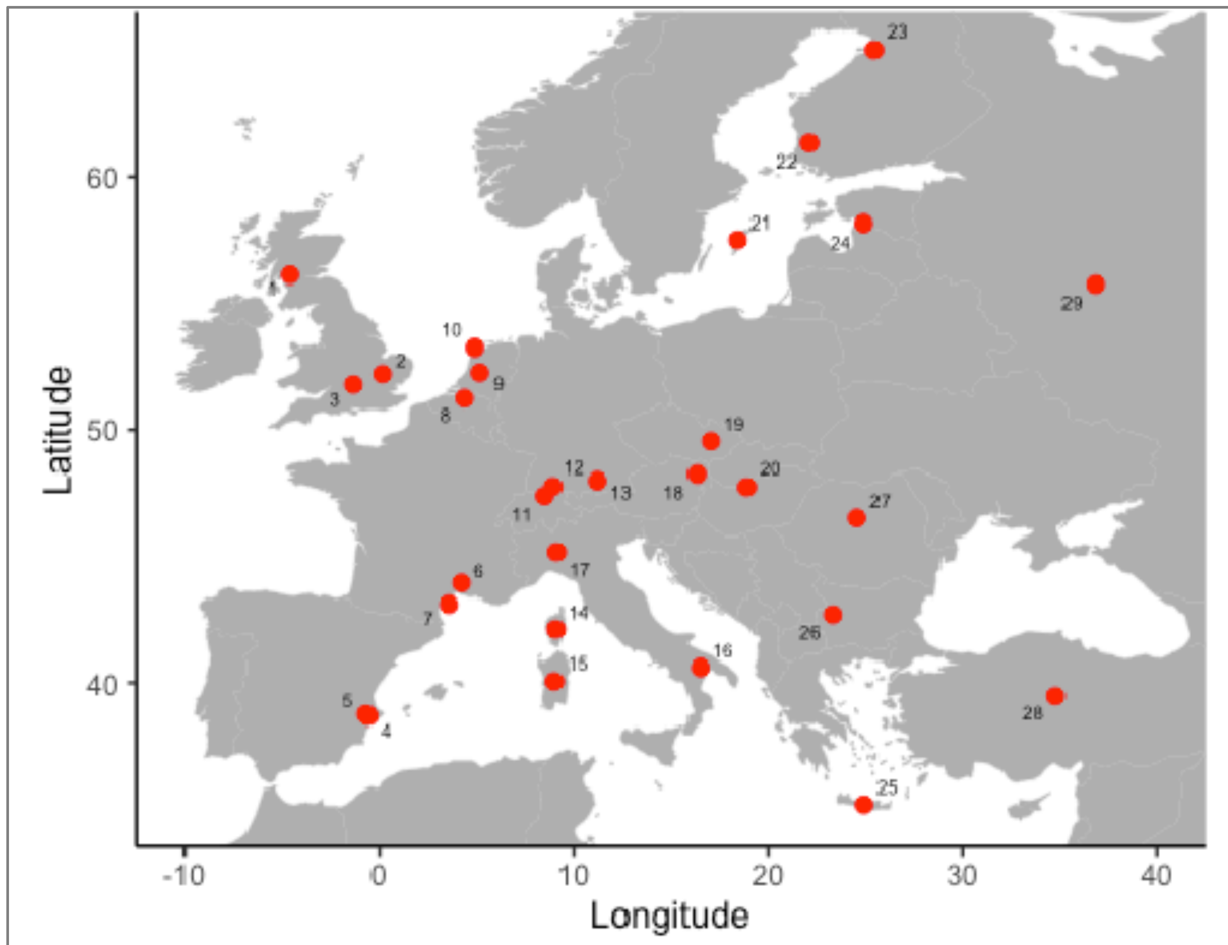


Figure 1.10: The sample locations of 29 great tit populations genotyped for the great tit HapMap project. Map reproduced from Spurgin et al. 2024. Sample location co-ordinates and sample size (N) are shown in Spurgin et al. 2024 S.Table 1.

The typing included samples from the Iberian Peninsula, Italy and Southeastern Europe / Balkans to investigate common glacial refugia and migration routes in European plants and animals (Taberlet et al. 1998; Hewitt 1999; Schmitt 2007), from which the great tit could have spread across Europe after the last glacial period. Colonisation reconstruction using TreeMix (Pickrell and Pritchard 2012) provided support that the populations have likely spread from a single restricted glacial refugium in either the Balkans and/or Turkey (Spurgin et al. 2024).

1.8.6. Low population differentiation

Spurgin et al. (2024) investigated population differentiation in compared great tit HapMap populations via Fst-based methods, to identify regions of the genome that had undergone selective sweeps. The authors identified low overall genomic divergence between European great tit populations, with similar cross-population low-differentiation patterns shown in Bosse et al. (2017), and via earlier microsatellite methods in van Bers et al. (2012). Although, some genomic differentiation was observed between Mediterranean Island populations and peripheral populations at the range edges of the species, despite the low differentiation between European populations. The authors detected reduced genetic diversity and bottlenecks in the Mediterranean Island populations (Crete, Corsica and Sardinia) relative to mainland European populations, which were suggested to result from genetic drift in small island populations. Unique genomic outlier regions were observed in peripheral populations e.g. Scotland, England, Spain and Finland, with recent directional (positive) selection most likely acting on those peripheral populations. These unique outlier regions contained candidate genes for adaptation to the environmental conditions experienced by the range edge populations. The candidate loci are putatively involved in morphology, thermal adaptation, and colouration (Spurgin et al. 2024).

1.8.7. Signatures of selection in the great tit genome

A set of published studies have identified genome-wide signatures of selection in the great tit genome (e.g. Spurgin et al. 2024, Bosse et al. 2017, Laine et al. 2016), however they do not reveal which agent of selection is responsible. This thesis will specifically investigate climate as the driver of selection and will look to see in which population(s) selection has happened. The filtered great tit HapMap genotype data set from Spurgin et al. (2024) was provided as the starting data files for this PhD project.

1.9. Summary of Introduction Chapter 1

1.9.1. Identifying adaptive genetic variation

In summary, SNP and whole-genome genotyping technology has enabled the spatial and temporal study of population genetic adaptation across different environmental conditions (Franks and Hoffmann 2012; Wadgymar et al. 2022; Bernatchez et al. 2023; Clark et al. 2023). Genome scans for identifying candidate environmental adaptation candidate loci need cautious interpretation because: (i) it is difficult to distinguish between false-positive candidate loci (Gautier 2015), causal loci or candidate loci simply in close linkage with the functional site under selection (Wang et al. 2010). (ii) the environmental variable may be correlated with the trait selection, but not necessarily causal (Tiffin and Ross-Ibarra 2014; Wadgymar et al. 2022). (iii) the candidate adaptation loci need experimental confirmation for the effect on fitness to be confirmed (Hoffmann et al. 2021; Wadgymar et al. 2022). Few of the published candidate climate adaptation genes identified in vertebrates have been validated for biological function in adaptation, e.g. by showing that they have fitness effects (Pardo-Diaz et al. 2015; Marková et al. 2023).

1.9.2. Population genomics of climate adaptation in a model system

In summary, great tits are a model species to understand the genomics of climate adaptation because: (i) Genomic resources for the great tit have been developed (van Oers et al. 2014; Laine et al. 2016; Kim et al. 2018). (ii) Contemporary wild populations experience a wide range of environments and climate conditions across Eurasia, which they have presumably locally adapted to since colonisation following the last ice age (Spurgin et al. 2024). (iii) There exists long-term monitoring data of wild populations across Europe (Visser et al. 2021; Cole et al. 2021; Sheldon et al. 2022). (iv) Many individuals in each population are native and resident (van Balen 1973; Gosler 1993), so it is fair to assume that they are well adapted to their local environment.

1.10. Outline summary of thesis chapters

Over the thesis Chapters 2-4 I developed an investigation of the genomics of adaptation to climate in European great tit (*Parus major*) populations. In Chapter 2, I performed a genome-environment analysis (GEA), by combining multiple local climate variables with single nucleotide polymorphisms (SNP) genotype data from multiple populations, broadly spanned the species range. This process identified genes putatively linked to adaptation to climate. In Chapter 3, I used extended hapotype homozygosity tests (informed by Chapter 2) to identify sweep regions potentially linked to climate adaptation. This analysis targeted detecting which of these populations sweeps occurred in and if parallel selection had occured. In Chapter 4 I focused on the candidate climate adaptation regions identified in the UK populations and investigated the temporal trends of allele frequency changes over the last century, using 95 whole-genome sequenced *Parus major* specimens originally collected from the UK and Isle of Man. Evidence suppporting these candidate climate adaptation genes included the stringent tests to detect selection from genomic data and strong temporal trends in the UK.

1.11. References

Abondio, P., Cilli, E., & Luiselli, D. (2022). Inferring Signatures of Positive Selection in Whole-Genome Sequencing Data: An Overview of Haplotype-Based Methods. *Genes*, **13**(5):e926. doi:10.3390/genes13050926

Ahrens, C.W., Rymer, P.D., Stow, A., Bragg, J., Dillon, S., Umbers, K.D.L., Dudaniec, R.Y. (2018). The search for loci under selection: trends, biases and progress. *Mol. Ecol.*, **27**(6):1342–1356. doi:10.1111/mec.14549

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N.A., & Goyache, F. (2020). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Sci. Rep.*, **10**:e2824. doi:10.1038/s41598-020-59839-x

Amiri Ghanatsaman, Z., Ayatolahi Mehrgardi, A., Asadollahpour Nanaei, H., Esmailizadeh, A. (2023). Comparative genomic analysis uncovers candidate genes related with milk production and adaptive traits in goat breeds. *Sci. Rep.*, **13**:e8722. doi:10.1038/s41598-023-35973-0

Andrews, C.A. (2010) Natural Selection, Genetic Drift, and Gene Flow Do Not Act in Isolation in Natural Populations. *Nature Education Knowledge*, **3**(10):5. https://www.nature.com/scitable/knowledge/library/natural-selection-genetic-drift-and-gene-flow-15186648/

Aplin, L.M., Farine, D.R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B.C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, **518**(7540):538-41. doi:10.1038/nature13998

Ashe-Jepson, E., Arizala Cobo, S., Basset, Y., Bladon, A.J., Kleckova, I., Laird-Hopkins, B.C., Mcfarlane, A., Sam, K., Savage, A.F., Zamora, A.C., Turner, E.C., & Lamarre, G.P.A. (2023). Tropical butterflies use thermal buffering and thermal tolerance as alternative strategies to cope with temperature increase. *J. Anim. Ecol.*, **92**:1759–1770. doi:10.1111/1365-2656.13970

Assersohn, K., Marshall, A.F., Morland, F., Brekke, P., & Hemmings, N. (2021). Why do eggs fail? Causes of hatching failure in threatened populations and consequences for conservation. *Anim. Conserv.*, **24**:540-551. doi:10.1111/acv.12674

Atkins, K.E., & Travis, J.M.J. (2010). Local Adaptation and the Evolution of Species' Ranges under Climate Change. *J. Theor. Biol.*, **266**(3):449–457. doi: 10.1016/j.jtbi.2010.07.014

Barton, H.J., & Zeng, K. (2019). The Impact of Natural Selection on Short Insertion and Deletion Variation in the Great Tit Genome. *Genome Biol. Evol.*, **11**(6):1514–1524. doi:10.1093/gbe/evz068

Beaumont M.A., & Nichols, R.A. (1996). Evaluating loci for use in the genetic analysis of population structure. *Proc. R. Soc. Lond. B*, **263**:1619–1626. doi:10.1098/rspb.1996.0237

Beaumont, M.A. (2005). Adaptation and speciation: what can Fst tell us? *Trends Ecol. Evol*., **20**(8):435-440. doi:10.1016/j.tree.2005.05.017.

Bednar-Friedl, B., R. Biesbroek, D.N. Schmidt, P. Alexander, K.Y. Børsheim, J. Carnicer, E. Georgopoulou, M. Haasnoot, G. Le Cozannet, P. Lionello, O. Lipka, C. Möllmann, V. Muccione, T. Mustonen, D. Piepenburg, and L. Whitmarsh, 2022: Europe. In: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [H.-O. Pörtner, D.C. Roberts, M. Tignor, E.S. Poloczanska, K. Mintenbeck, A. Alegría, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, B. Rama (eds.)]. Cambridge University Press, Cambridge, UK and New York, NY, USA, pp. 1817–1927, doi:10.1017/9781009325844.015. [Figure 13.22]

Bello, S.F., Lawal, R.A., Adeola, A.C., Nie, Q. (2023). The study of selection signature and its applications on identification of candidate genes using whole genome sequencing data in chicken—a review. *Poult. Sci. J.*, **102**(6):e102657. doi:10.1016/j.psj.2023.102657

Ben-Jemaa, S., Mastrangelo, S., Lee, S-H., Lee, J.H., & Boussaha, M. (2020). Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. *Sci. Rep.*, **10:**e19466. doi:10.1038/s41598-020-76576-3

Berg, J.J., & Coop, G. (2014). A Population Genetic Signal of Polygenic Adaptation. *PLoS Genet.*, **10**(8):e1004412. doi:10.1371/journal.pgen.1004412

Bernatchez, L., Ferchaud, A.L., Berger, C.S., Venney, C.J., & Xuereb, A. (2023). Genomics for Monitoring and Understanding Species Responses to Global Climate Change. *Nat. Rev. Genet.*, **25**(3):165-183. doi:10.1038/s41576-023-00657-y

BirdLife International (2024) Species factsheet: Parus major. Downloaded from http://datazone.birdlife.org/species/factsheet/great-tit-parus-major

Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M., & González, J. (2021). Temperature, Rainfall and Wind Variables Underlie Environmental Adaptation in Natural Populations of Drosophila Melanogaster. *Mol. Ecol.*, **30**(4):938–954. doi:10.1111/mec.15783

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S., & SanCristobal, M. (2010). Detecting Selection in Population Trees: The Lewontin and Krakauer Test Extended. *Genetics*, **186**(1):241–262. doi:10.1534/genetics.110.117275

Boonyaritticha ikij, R., Verbrugghe, E., Dekeukeleire, D., De Beelde, R., Rouffaer, L.O., & Haesendonck, R., Strubbe, D., Mattheus, W., Bertrand, S., Pasmans, F., Bonte, D., Verheyen,

K., Lens, L., & Martel, A.(2017). Salmonella Typhimurium DT193 and DT99 are present in great and blue tits in Flanders, Belgium. *PLoS ONE*, **12**(11):e0187640. doi:10.1371/journal.pone.0187640

Bosse, M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G., McMahon, K., Poissant, J., Verhagen, I., Groenen, M.A.M., van Oers, K., Sheldon, B.C., Visser, M.E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, **358**:365-368. doi:10.1126/science.aal3298

Boyes, D.H., Evans, D.M., Fox, R., Parsons, M.S., & Pocock, M.J.O. (2021). Is light pollution driving moth population declines? A review of causal mechanisms across the life cycle. *Insect Conserv. Divers.*, **14**:167-187. doi:10.1111/icad.12447

Brandt, D.Y.C., César, J., Goudet, J., Meyer, D. (2018). The Effect of Balancing Selection on Population Differentiation: A Study with HLA Genes. *G3: Genes Genomes Genet.*, **8**(8):2805–2815. doi:10.1534/g3.118.200367

Burny, C., Nolte, V., Nouhaud, P., Dolezal, M., Schlötterer, C. (2020). Secondary Evolve and Resequencing: An Experimental Confirmation of Putative Selection Targets without Phenotyping. *Genome Biol. Evol.*, **12**(3):151–159. doi:10.1093/gbe/evaa036

Caizergues, A.E., Le Luyer, J., Grégoire, A., Szulkin, M., Senar, J.-C., Charmantier, A., & Perrier, C. (2022). Epigenetics and the city: Non-parallel DNA methylation modifications across pairs of urban-forest Great tit populations. *Evol. Appl.*, **15**:149–165. doi:10.1111/eva.13334

Campagna, L., & Toews, D.P.L. (2022). The genomics of adaptation in birds. *Curr. Biol.*, **32**(20):R1173-R1186. doi:10.1016/j.cub.2022.07.076

Capblancq, T., & Forester, B.R. (2021). Redundancy Analysis: A Swiss Army Knife for Landscape Genomics. *Methods Ecol. Evol.*, **12**(12):2298–2309. doi:10.1111/2041-210X.13722

Chamberlain, D.E., Cannon, A.R., Toms, M., Leech, D., Hatchwell, B., & Gaston, K. (2009). Avian productivity in urban landscapes: a review and meta-analysis. *Ibis*, **151**:1-18. doi:10.1111/j.1474-919X.2008.00899.x

Charlesworth D. (2006). Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.*, **2(**4):e64. doi:10.1371/journal.pgen.0020064

Charmantier, A., & Gienapp, P. (2014). Climate change and timing of avian breeding and migration: evolutionary versus plastic changes. *Evol Appl.*, **7**(1):15-28. doi: 10.1111/eva.12126

Charmantier, A., Demeyrier, V., Lambrechts, M., Perret, S., & Grégoire, A. (2017). Urbanization Is Associated with Divergence in Pace-of-Life in Great Tits. *Front. Ecol. Evol.*, **5**:e53. doi:10.3389/fevo.2017.00053

Chen, H., Patterson, N., & Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.*, **20**(3):393-402. doi:10.1101/gr.100545.109

Clark, R.D., Catalano, K.A., Fitz, K.S., Garcia, E., Jaynes, K.E., Reid, B.N., Sawkins, A., Snead, A.A., Whalen, J.C. & Pinsky, M.L. (2023). The Practice and Promise of Temporal Genomics for Measuring Evolutionary Responses to Global Change. *Mol. Ecol. Resour.*, **00**:1–17. doi:10.1111/1755-0998.13789

Clements, J.F., Rasmussen, P.C., Schulenberg, T.S., Iliff, M.J., Fredericks, T.A., Gerbracht, J.A., Lepage, D., Spencer, A., Billerman, S.M., Sullivan, B.L., & Wood, C.L. (2023). The eBird/Clements checklist of Birds of the World: v2023. Downloaded from https://www.birds.cornell.edu/clementschecklist/download/

Clobert, J., Perrins, C.M., McCleery, R.H., & Gosler, A.G. (1988). Survival Rate in the Great Tit Parus Major in Relation to Sex, Age, and Immigration Status. *J. Anim. Ecol.*, **57**(1): 287–306. doi:10.2307/4779.

Clutton-Brock, T., & Sheldon, B.C. (2010). Individuals and populations: the role of long-term, individual-based studies of animals in ecology and evolutionary biology. *Trends Ecol. Evol*., **25**(10):562–573. doi:10.1016/j.tree.2010.08.002

Cole, E.F., Regan, C.E., & Sheldon, B.C. (2021). Spatial variation in avian phenological response to climate change linked to tree health. *Nat. Clim. Chang.*, **11**:872–878. doi:10.1038/s41558-021-01140-4

Corcoran, P., Gossmann, T.I., Barton, H.J., Great Tit HapMap Consortium, Slate, J., & Zeng, K. (2017) Determinants of the efficacy of natural selection on coding and noncoding variability in two passerine species. *Genome Biol. Evol.*, **9**(11):2987-3007. doi:10.1093/gbe/evx213

Corsini, M., Schöll, E.M., Di Lecce, I., Chatelain, M., Dubiec, A., & Szulkin, M. (2020). Growing in the city: Urban evolutionary ecology of avian growth rates. *Evol. Appl.*, **14**(1):69-84. doi:10.1111/eva.13081

da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., van Oers, K., Dibbits, B., Slate, J., Crooijmans, R.P.M.A., Visser, M.E., & Groenen, M.A.M. (2019). The Genomic Complexity of a Large Inversion in Great Tits. *Genome Biol. Evol.*, **11**(7):1870–1881. doi:10.1093/gbe/evz106

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, R.M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2):giab008. doi:10.1093/gigascience/giab008

Dehasque, M., Ávila-Arcos, M.C., Díez-del-Molino, D., Fumagalli, M., Guschanski, K., Lorenzen, E.D., Malaspinas, A-S., Marques-Bonet, T., Martin, M.D., Murray, G.G.R., Papadopulos, A.S.T., Overgaard Therkildsen, N., Wegmann, D., Dalén, L., Foote, A.D. (2020). Inference of natural selection from ancient DNA. *Evol. Lett.*, **4**(2):94–108. doi:10.1002/evl3.165

de Jong, M., Caro, S.P., Gienapp, P., Spoelstra, K., Visser, M.E. (2017). Early Birds by Light at Night: Effects of Light Color and Intensity on Daily Activity Patterns in Blue Tits. *J. Biol. Rhythms*, **32**(4):323-333. doi:10.1177/0748730417719168

de Villemereuil, P., & Gaggiotti, O. E. (2015). A New Fst-Based Method to Uncover Local Adaptation Using Environmental Variables. *Methods Ecol. Evol.*, **6**(11):1248–1258. doi:10.1111/2041-210X.12418

de Villemereuil, P., Gaggiotti, O., Mouterde, M., & Till-Bottraud, I. (2016). Common garden experiments in the genomic era: new perspectives and opportunities. *Heredity*, **116**:249–254. doi:10.1038/hdy.2015.93

EBBA2 2022: European Breeding Bird Atlas 2 website. European Bird Census Council. Accessed from: http://ebba2.info (December 2023).
EBBA2 data from: Keller, V., Herrando, S., Voříšek, P., Franch, M., Kipson, M., Milanesi, P., Martí, D., Anton, M., Klvaňová, A., Kalyakin, M.V., Bauer, H-G. & Foppen, R.P.B. (2020). European Breeding Bird Atlas 2: Distribution, Abundance and Change. European Bird Census Council & Lynx Edicions, Barcelona.

Edea, Z., & Kim, K.S. (2014). A whole genomic scan to detect selection signatures between Berkshire and Korean native pig breeds. *J. Anim. Sci. Technol.*, **56**:1–7. doi:10.1186/2055-0391-56-23

Eguchi, K. (1980), The feeding ecology of the nestling great tit, Parus major minor, in the temperate ever-green broadleaved forest II. With reference to breeding ecology. *Popul. Ecol.*, **22**:284-300. doi:10.1007/BF02530852

Ehret, G.B. (2010). Genome-wide association studies: contribution of genomics to understanding blood pressure and essential hypertension. *Curr. Hypertens. Rep.*, **12**(1):17-25. doi:10.1007/s11906-009-0086-6

Elguero, E., Délicat-Loembet, L.M., Rougeron, V., Arnathau, C., Roche, B., Becquart, P., Gonzalez, J. P., Nkoghe, D., Sica, L., Leroy, E.M., Durand, P., Ayala, F.J., Ollomo, B., Renaud, F., & Prugnolle, F. (2015). Malaria Continues to Select for Sickle Cell Trait in Central Africa. *Proc. Natl. Acad. Sci. U S A*, **112**(22):7051–7054. doi:10.1073/pnas.1505665112

Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Ann. Rev. Ecol. Evol. Syst.*, **44**:239-259. doi:10.1146/annurev-ecolsys-110411-160327

Eydivandi, S., Roudbar, M.A., Ardestani, S.S., Momen, M., Sahana, G. (2021). A selection signatures study among Middle Eastern and European sheep breeds. *J. Anim. Breed Genet.*, **138**(5):574–588. doi:10.1111/jbg.12536

Fan, S., Hansen, M.E.B., Lo, Y., & Tishkoff, S. A. (2016). Going Global by Adapting Local: A Review of Recent Human Adaptation. *Science*, **354**(6308):54–59. doi:10.1126/science.aaf5098

Farine, D.R., & Lang, S.D.J. (2013). The early bird gets the worm: foraging strategies of wild songbirds lead to the early discovery of food sources. *Biol. Lett.*, **9**(6):20130578. doi:10.1098/rsbl.2013.0578

Fay, J.C., Wyckoff, G.J., & Wu, C.I. (2002). Testing the neutral theory of molecular evolution with genomic data from Drosophila. *Nature*, **415**(6875):1024–1026. doi:10.1038/4151024a

Feng, S., Stiller, J., Deng, Y., Armstrong, J., Fang, Q., Reeve, A.H., Xie, D., Chen, G., Guo, C., Faircloth, B.C., Petersen, B., Wang, Z., Zhou, Q., Diekhans, M., Chen, W., Andreu-Sánchez, S., Margaryan, A., Howard, J.T., Parent, C., ... & Zhang G. (2020). Dense sampling of bird diversity increases power of comparative genomics. *Nature*, **587**(7833):252-257. doi:10.1038/s41586-020-2873-9

Firth J.A., Sheldon B.C., & Farine D.R. (2016). Pathways of information transmission among wild songbirds follow experimentally imposed changes in social foraging structure. *Biol. Lett.*, **12**:e20160144. doi:10.1098/rsbl.2016.0144

Fisher, R.A. (1919). XV. —The Correlation between Relatives on the Supposition of Mendelian Inheritance. *Earth Environ. Sci. Trans. R. Soc. Edinb.*, **52**:399–433. doi:10.1017/S0080456800012163

Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A.C., Coquelle, C., Delgado, J-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., ...Gautier, M. (2019). A Genomic Map of Climate Adaptation in Mediterranean Cattle Breeds. *Mol. Ecol.*, **28**(5):1009–1029. doi:10.1111/mec.15004

Forester, B.R., Lasky, J.R., Wagner, H.H., Urban, D.L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Mol. Ecol.*, **27**:2215–2233. doi:10.1111/mec.14584

Franks, S.J., & Hoffmann, A.A. (2012). Genetics of climate change adaptation. *Annu. Rev. Genet.*, **46**:185–208. doi:10.1146/annurev-genet-110711-155511

Frichot, E., Schoville, S.D., Bouchard, G., & François, O. (2013). Testing for Associations between Loci and Environmental Gradients Using Latent Factor Mixed Models. *Mol. Biol. Evol.*, **30**(7):1687–1699. doi:10.1093/molbev/mst063

Gautier, M. (2015). Genome-Wide Scan for Adaptive Divergence and Association with Population-Specific Covariates. *Genetics*, **201**:1555–1579. doi:10.1534/genetics.115.181453

Gienapp, P., Laine, V.N., Mateman, A.C., van Oers, K., & Visser, M.E. (2017). Environment-Dependent Genotype-Phenotype Associations in Avian Breeding Time. *Front. Genet.*, **8**:e102. doi:10.3389/fgene.2017.00102

Gosler, A. (1993). The Great Tit. Hamlyn Species Guides, London.

Gosler, A., P. Clement, and D. A. Christie (2020). Great Tit (Parus major), version 1.0. In Birds of the World (S. M. Billerman, B. K. Keeney, P. G. Rodewald, and T. S. Schulenberg, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. doi:10.2173/bow.gretit1.01

Günther, T. & Coop, G. (2013). Robust Identification of Local Adaptation from Allele Frequencies. *Genetics*, **195**(1):205–220. doi:0.1534/genetics.113.152462

Halupka, L., Arlt, D. Tolvanen, J. Millon, A., Bize, P., Adamík, P., Albert, P., Arendt, W.J., Artemyev, A.V., Baglione, V., Bańbura, J., Bańbura, M., Barba, E., Barrett, R.T., Becker, P.H., Belskii, E., Bolton, M., Bowers, E.K., Bried, J., ... & Halupka, K. (2023). The effect of climate change on avian offspring production: A global meta-analysis. *Proc. Natl. Acad. Sci. U S A.*, **120**(19):e2208389120. doi:10.1073/pnas.2208389120

Hewitt, G.M. (1999). Post-glacial re-colonization of European biota. *Biol. J. Linn. Soc.*, **68**(1–2):87-112. doi:10.1006/bijl.1999.0332.

Hinks A.E., Cole E.F., Daniels K.J., Wilkin T.A., Nakagawa S., & Sheldon B.C.. (2015). Scale-dependent phenological synchrony between songbirds and their caterpillar food source. *Am. Nat.*, **186**(1):84–97. doi:10.1086/681572

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A., & Whitlock, M.C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *Am. Nat.*, **188**(4):379-397. doi:10.1086/688018

Hoffmann, A.A., Weeks, A.R., & Sgrò, C.M. (2021). Opportunities and challenges in assessing climate change vulnerability through genomics. *Cell*, **184**(6):1420-1425. doi:10.1016/j.cell.2021.02.006

Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., Nachman, M.W. (2016). Natural history collections as windows on evolutionary processes. *Mol. Ecol.*, **25**(4):864-81. doi:10.1111/mec.13529

Huang, Z., Xu, Z., Bai, H., Huang, Y., Kang, N., Ding, X., Liu, J., Luo, H., Yang, C., Chen, W., Guo, Q., Xue, L., Zhang, X., Xu, L., Chen, M., Fu, H., Chen, Y., Yue, Z., Fukagawa, T., Liu, S., Chang, G., Xu, L. (2023). Evolutionary analysis of a complete chicken genome. *Proc. Natl. Acad. Sci. U S A.*, **120**(8):e2216641120. doi:10.1073/pnas.2216641120

IPCC 2023: Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115. doi:10.59327/IPCC/AR6-9789291691647

Isaksson, C., Sepil, I., Baramidze, V., & Sheldon, B.C. (2013). Explaining variance of avian malaria infection in the wild: the importance of host density, habitat, individual life-history and oxidative stress. *BMC Ecol.*, **13**:e15. doi:10.1186/1472-6785-13-15

Jain, K., & Stephan, W. (2017) Modes of Rapid Polygenic Adaptation, *Mol. Biol. Evol.*, **34**(12):3169–3175. doi:10.1093/molbev/msx240

Jensen, J.D. (2014). On the Unfounded Enthusiasm for Soft Selective Sweeps. *Nat. Comm.*, **5**:e5281. doi:10.1038/ncomms6281

Jensen, E.L., & Leigh, D.M. (2022). Using temporal genomics to understand contemporary climate change responses in wildlife. *Ecol. Evol.*, **12**(9):e9340. doi:10.1002/ece3.9340

Jensen, J.K., Ekroos, J., Watson, H., Salón, P., Olsson, P., & Isaksson, C. (2023). Urban tree composition is associated with breeding success of a passerine bird, but effects vary within and between years, *Oecologia*, **201**:585-597. doi:10.1007/s00442-023-05319-8

Johansson, U.S., Ekman, J., Bowie, R.C.K., Halvarsson, P., Ohlson, J.I., Price, T.D., & Ericson, P.G.P. (2013). A complete multilocus species phylogeny of the tits and chickadees (Aves: Paridae). *Mol. Phylogenetics and Evol.*, **69**(3):852-860. doi:10.1016/j.ympev.2013.06.019

Kaler, A.S., & Purcell, L.C. (2019). Estimation of a significance threshold for genome-wide association studies. *BMC Genomics*, **20**: e618. doi:10.1186/s12864-019-5992-7

Kawecki, T.J., & Ebert, D. (2004). Conceptual issues in local adaptation. *Ecol. Lett.*, **7**:1225-1241. doi:10.1111/j.1461-0248.2004.00684.x

Kawecki, T.J., Lenski, R.E., Ebert, D., Hollis, B., Olivieri, I., & Whitlock, M. C. (2012). Experimental evolution. *Trends Ecol. Evol.*, **27**(10):547–560. doi:10.1016/j.tree.2012.06.001

Kentistou, K.A., Luan, J., Wittemans, L.B.L., Hambly, C., Klaric, L., Kutalik, Z., Speakman, J.R., Wareham, N.J., Kendall, T.J., Langenberg, C., Wilson, J.F., Joshi, P.K., & Morton, N.M. (2023). Large scale phenotype imputation and in vivo functional validation implicate ADAMTS14 as an adiposity gene. *Nat. comm.*, **14**:e307. doi:10.1038/s41467-022-35563-0

Kim, J-M., Santure, A.W., Barton, H.J., Quinn, J.L., Cole, E.F., Great Tit HapMap Consortium, Visser, M.E., Sheldon, B.C., Groenen, M.A.M., van Oers, K., & Slate, J. (2018). A high-density SNP chip for genotyping great tit (Parus major) populations and its application to studying the genetic architecture of exploration behaviour. *Mol. Ecol. Resour.*, **18**:877–891. doi:10.1111/1755-0998.12778

Klassmann A, & Gautier M. (2022). Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. *PLoS One*, **17**(1):e0262024. doi:10.1371/journal.pone.0262024

Korneliussen, T.S., Albrechtsen, A., and Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, **15**:e356. doi:10.1186/s12859-014-0356-4

Krawiec, M., Kuczkowski, M., Kruszewicz, A.G., & Wieliczko, A. (2015). Prevalence and genetic characteristics of Salmonella in free-living birds in Poland. *BMC Vet Res.*, **11**:e15. doi:10.1186/s12917-015-0332-x

Kvist, L., Ruokonen, M., Lumme, J., & Orell, M. (1999). The colonization history and present-day population structure of the European great tit (Parus major major). *Heredity*, **82**:495–502. doi:10.1038/sj.hdy.6885130

Kvist, L., Martens, J., Higuchi, H., Nazarenko, A.A., Valchuk, O.P., & Orell, M. (2003). Evolution and Genetic Structure of the Great Tit (Parus major) Complex. *Proc. Biol. Sci.*, **270**(1523):1447–1454. doi:10.1098/rspb.2002.2321

Kvist, L., & Rytkönen, S. (2006). Characterization of a secondary contact zone of the Great Tit Parus major and the Japanese Tit P. minor (Aves: Passeriformes) in Far Eastern Siberia with DNA markers. *Zootaxa*, **1325**:55–73. doi:https://doi.org/10.11646/zootaxa.1325.1.5

Lachish, S., Knowles, S.C.L., Alves, R., Wood, M.J. and Sheldon, B.C. (2011), Fitness effects of endemic malaria infections in a wild bird population: the importance of ecological structure. *J. Anim. Ecol.*, **80**: 1196-1206. https://doi.org/10.1111/j.1365-2656.2011.01836.x

Lack, D. (1964) A long term study of the great tit (Parus major). *J. Anim. Ecol.*, **33**:159–173. doi:10.2307/2437

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O., Verhoeven, K.J., de Jager, V., Megens, H-J., Warren, W.C., Minx, P., Crooijmans, R.P.M.A., Corcoran, P., The Great Tit HapMap Consortium, Sheldon, B.C., Slate, J., Zeng, K., van Oers, K., Visser, M.E., & Groenen, M.A.M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Comm.*, **7**:e10474. doi:10.1038/ncomms10474

Larsson, L.C., Laikre, L., Palm, S., André, C., Carvalho, G. R., & Ryman, N. (2007). Concordance of Allozyme and Microsatellite Differentiation in a Marine Fish, but Evidence of Selection at a Microsatellite Locus. *Mol. Ecol.*, **16**(6):1135–1147. doi:10.1111/j.1365-294X.2006.03217.x

Lewontin, R.C., & Krakauer, J. (1973). Distribution of Gene Frequency as a Test of the Theory of the Selective Neutrality of Polymorphisms. *Genetics*, **74**(1):175–195. doi:10.1093/genetics/74.1.175

Li, R., Li, C., Chen, H., Li, R., Chong, Q., Xiao, H., & Chen, S. (2020), Genome-wide scan of selection signatures in Dehong humped cattle for heat tolerance and disease resistance. *Anim. Genet.*, **51**:292-299. doi:10.1111/age.12896

Lindner, M., Ramakers, J.J.C., Verhagen, I., Tomotani, B.M., Mateman, A.C., Gienapp, P., & Visser, M.E. (2023). Genotypes selected for early and late avian lay date differ in their phenotype, but not fitness, in the wild. *Sci. Adv.*, **9**,eade6350. doi:10.1126/sciadv.ade6350

Loog, L., Thomas, M.G., Barnett, R., Allen, R., Sykes, N., Paxinos, P.D., Lebrasseur, O., Dobney, K., Peters, J., Manica, A., Larson, G., & Eriksson, A. (2017). Inferring Allele Frequency Trajectories from Ancient DNA Indicates That Selection on a Chicken Gene Coincided with Changes in Medieval Husbandry Practices. *Mol. Biol. Evol.*, **34**(8):1981–1990. doi:10.1093/molbev/msx142

Malhi, Y., Franklin, J., Seddon, N., Solan, M., Turner, M.G., Field, C.B., & Knowlton, N. (2020). Climate Change and Ecosystems: Threats, Opportunities and Solutions. *Philos. Trans. R. Soc. B: Biol.*, **375**(1794). doi:10.1098/rstb.2019.0104

Manolio, T.A., Collins, F.S., Cox, N.J., Goldstein, D.B., Hindorff, L.A., Hunter, D.J., McCarthy, M.I., Ramos, E.M., Cardon, L.R., Chakravarti, A., Cho, J.H., Guttmacher, A.E., Kong, A., Kruglyak, L., Mardis, E., Rotimi, C.N., Slatkin, M., Valle, D., Whittemore, A.S. ... & Visscher, P.M. (2009). Finding the missing heritability of complex diseases. *Nature*, **461**:747–753. doi:10.1038/nature08494

Marková, S., Lanier, H.C., Escalante, M.A., da Cruz, M.O.R., Horníková, M., Konczal, M., Weider, L.J., Searle, J.B., & Kotlík, P. (2023). Local adaptation and future climate vulnerability in a wild rodent. *Nat. Comm.*, **14**:e7840. doi:10.1038/s41467-023-43383-z

Maziarz, M., Broughton, R.K., Beck, K.B., Robinson, R.A., & Sheldon, B.C. (2023). Temporal avoidance as a means of reducing competition between sympatric species. *R. Soc. Open Sci.*, **10**:e230521. doi:10.1098/rsos.230521

McCleery, R.H., Pettifor, R.A., Armbruster, P., Meyer, K., Sheldon, B.C., & Perrins, C.M. (2004). Components of Variance Underlying Fitness in a Natural Population of the Great Tit Parus major. *Am. Nat.*, **164**:3, E62-E72. doi:10.1086/422660

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M.A. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, **20**:1297-1303. doi: 10.1101/gr.107524.110

Mentaschi, L., Duveiller, G., Zulian, G., Corbane, C., Pesaresi, M., Maes, J., Stocchino, A., & Feyen, L. (2022). Global long-term mapping of surface temperature shows intensified intra-city urban heat island extremes. *Global Environ. Change*, **72**:e102441. doi:10.1016/j.gloenvcha.2021.102441.

Mizumo Tomotani, B. (2017). Managing time in a changing world: Timing of avian annual cycle stages under climate change. [Thesis fully internal (DIV), University of Groningen]. University of Groningen.

Montejo-Kovacevich, G., Meier, J.I., Bacquet, C.N., Warren, I.A., Chan, Y.F., Kucka, M., Salazar, C., Rueda-M, N., Montgomery, S.H., McMillan, W.O., Kozak, K.M., Nadeau, N.J., Martin, S.H., & Jiggins, C.D. (2022). Repeated Genetic Adaptation to Altitude in Two Tropical Butterflies. *Nat. Comm.*, **13**(1):1–16. doi:10.1038/s41467-022-32316-x

Morey, R.D., Romeijn, J-W., & Rouder, J.N. (2016). The philosophy of Bayes factors and the quantification of statistical evidence. *J. Math. Psychol.*, **72**:6-18. doi:10.1016/j.jmp.2015.11.001

Nosil, P., Soria-Carrasco, V., Feder, J.L., Flaxman, S.M., & Gompert, Z. (2019). Local and system-wide adaptation is influenced by population connectivity. *Conserv. Genet.*, **20**:45–57. doi:10.1007/s10592-018-1097-0

Novembre J, & Barton N.H. (2018). Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics*, **208**(4):1351-1355. doi: 10.1534/genetics.118.300786
Erratum in: *Genetics*, **209**(2):627

Olazcuaga, L., Loiseau, A., Parrinello, H., Paris, M., Fraimout, A., Guedot, C., Diepenbrock, L.M., Kenis, M., Zhang, J., Chen, X., Borowiec, N., Facon, B., Vogt, H., Price, D.K., Vogel, H., Prud'homme, B., Estoup, A., & Gautier, M. (2020). A Whole-Genome Scan for Association with Invasion Success in the Fruit Fly Drosophila suzukii Using Contrasts of Allele Frequencies

Corrected for Population Structure. *Mol. Biol. Evol.*, **37**(8):2369–2385. doi:10.1093/molbev/msaa098

Oostra, V., Saastamoinen, M., Zwaan, B.J., & Wheat, C. W. (2018). Strong Phenotypic Plasticity Limits Potential for Evolutionary Responses to Climate Change. *Nat. Comm.*, **9**:e1005. doi:10.1038/s41467-018-03384-9

Pardo-Diaz, C., Salazar, C., & Jiggins, C.D. (2015). Towards the identification of the loci of adaptive evolution. *Methods Ecol. Evol.*, **6**(4):445-464. doi:10.1111/2041-210X.12324

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change impacts across natural systems. *Nature*, **421**:37–42 (2003). doi:10.1038/nature01286

Passamonti, M.M., Somenzi, E., Barbato, M., Chillemi, G., Colli, L., Joost, S., Milanesi, M., Negrini, R., Santini, M., Vajana, E., Williams, J.L., Ajmone-Marsan, P. (2021). The Quest for Genes Involved in Adaptation to Climate Change in Ruminant Livestock. *Animals (Basel)*, **11**(10):e2833. doi:10.3390/ani11102833

Patrick, S.C., Chapman, J.R., Dugdale, H.L., Quinn, J.L., & Sheldon, B.C. (2012).Promiscuity, paternity and personality in the great tit. *Proc. R. Soc. B.*, **279**:1724–1730. doi:10.1098/rspb.2011.1820

Perrier, C., & Charmantier, A. (2019). On the importance of time scales when studying adaptive evolution. *Evol. Lett.*, **3**:240-247. doi:10.1002/evl3.86

Perrins, C.M. (1970). The timing of bird's breeding seasons. *Ibis*, **112**:242-255. doi:10.1111/j.1474-919X.1970.tb00096.x

Pickrell, J.K., & Pritchard, J.K. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.*, **8**(11):e1002967. doi: 10.1371/journal.pgen.1002967

Pipoly, I., Preiszner, B., Sándor, K., Sinkovics, C., Seress, G., Vincze, E., Bókony, V., & Liker, A. (2022). Extreme Hot Weather Has Stronger Impacts on Avian Reproduction in Forests Than in Cities. *Frontiers Ecol. Evol.*, **10**:e825410. doi:10.3389/fevo.2022.825410

Price, T.D., Qvarnström, A., & Irwin, D.E. (2003). The role of phenotypic plasticity in driving genetic evolution. *Proc. Biol. Sci.*, **270**(1523):1433-40. doi:10.1098/rspb.2003.2372

Pritchard, J.K., Pickrell, J.K. & Coop, G. (2010). The Genetics of Human Adaptation: Hard Sweeps, Soft Sweeps, and Polygenic Adaptation. *Curr. Biol.*, **20**(4):R208–R215. doi:10.1016/j.cub.2009.11.055

Pritchard, J.K., & Di Rienzo, A. (2010). Adaptation - not by sweeps alone. *Nat. Rev. Genet.*, **11**(10):665–667. doi:10.1038/nrg2880

Privé, F., Luu, K., Vilhjálmsson, B.J., & Blum, M.G.B. (2020). Performing highly efficient genome scans for local adaptation with R package pcadapt version 4. *Mol. Biol. Evol.*, **37**(7):2153–2154. doi:10.1093/molbev/msaa053

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakkerm P.I.W., Daly, M.J., & Sham, P.C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.*, **81**(3):559-575. doi:10.1086/519795

Rachman, M.P., Bamidele, O., Dessie, T., Smith, J., Hanotte, O., & Gheyas, A.A. (2024). Genomic analysis of Nigerian indigenous chickens reveals their genetic diversity and adaptation to heat-stress. *Sci. Rep.*, **14**:e2209. doi:10.1038/s41598-024-52569-4

Raxworthy, C.J., & Smith, B.T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol.*, **36**(11):1049-1060. doi:10.1016/j.tree.2021.07.009

Reid, B.N., Star, B., & Pinsky, M.L. (2023). Detecting parallel polygenic adaptation to novel evolutionary pressure in wild populations: a case study in Atlantic cod (Gadus morhua). *Phil. Trans. R. Soc. B*, **378**(1881):e20220190. doi:10.1098/rstb.2022.0190

Rellstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M., & Holderegger, R. (2015). A Practical Guide to Environmental Association Analysis in Landscape Genomics. *Mol. Ecol.*, **24**(17):4348–4370. doi:10.1111/mec.13322

Rigal, S., Dakos, V., Alonso, H., Auniņš, A., Benkő, Z., Brotons, L., Chodkiewicz, T., Chylarecki, P., de Carli, E., Del Moral, J. C., Domşa, C., Escandell, V., Fontaine, B., Foppen, R., Gregory, R., Harris, S., Herrando, S., Husby, M., Ieronymidou, C., Jiguet, F., … & Devictor, V. (2023). Farmland practices are driving bird population decline across Europe. *Proc. Natl. Acad. Sci. U S A.*, **120**(21):e2216573120. doi:10.1073/pnas.2216573120

Robinson, R.A., Lawson, B., Toms, M.P., Peck, K.M., Kirkwood, J.K., Chantrey, J., Clatworthy, I.R., Evans, A.D., Hughes, L.A., Hutchinson, O.C., John, S.K., Pennycott, T.W., Perkins, M.W., Rowley, P.S., Simpson, V.R., Tyler, K.M., & Cunningham, A.A. (2010) Emerging infectious disease leads to rapid population declines of common British birds. *PLoS ONE*, **55**:e12215. doi:10.1371/journal.pone.0012215

Rocha, J.L., Godinho, R., Brito, J.C., Nielsen, R. (2021). Life in Deserts: The Genetic Basis of Mammalian Desert Adaptation. *Trends Ecol. Evol.*, **36**(7):637-650, doi:10.1016/j.tree.2021.03.007

Rohat, G., Flacke, J., Dosio, A., Pedde, S., Dao, H., van Maarseveen, M. (2019). Influence of changes in socioeconomic and climatic conditions on future heat-related health challenges in Europe. *Global Planet. Change*, **172**:45-59. doi:10.1016j.gloplacha.2018.09.013

Rose, N.H., Bay, R.A., Morikawa, M.K., & Palumbi, S. R. (2018). Polygenic Evolution Drives Species Divergence and Climate Adaptation in Corals. *Evolution*, **72**(1):82–94. doi: 10.1111/evo.13385

Rubenstein, M.A., Weiskopf, S.R., Bertrand, R., Carter, S.L., Comte, L., Eaton, M.J., Johnson, C.G., Lenoir, J., Lynch, A.J., Miller, B.W., Morelli, T.L., Rodriguez, M.A., Terando, A., & Thompson, L.M. (2023). Climate change and the global redistribution of biodiversity: substantial variation in empirical support for expected range shifts. *Environ. Evid.*, **12**:e7. doi:10.1186/s13750-023-00296-0

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S., & The International HapMap Consortium (2007). Genome-Wide Detection and Characterization of Positive Selection in Human Populations. *Nature*, **449**:913–918. doi:10.1038/nature06250

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N.J., Dominoni, D.M., Helm, B., Lundberg, M., Senar, J.C., Sprau, P., Visser, M.E. & Isaksson, C. (2021). Continent-Wide Genomic Signatures of Adaptation to Urbanisation in a Songbird across Europe. *Nat. Comm.*, **12**(1):2983. doi:10.1038/s41467-021-23027-w

Salmón, P., López-Idiáquez, D., Capilla-Lasheras, P., Pérez-Tris, J., Isaksson, C., & Watson, H. (2023). Urbanisation impacts plumage colouration in a songbird across Europe: Evidence from a correlational, experimental and meta-analytical approach. *J. Anim. Ecol.*, **92**:1924–1936. doi:10.1111/1365-2656.13982

Santure, A.W., De Cauwer, I., Robinson, M.R., Poissant, J., Sheldon, B.C., & Slate, J. (2013). Genomic dissection of variation in clutch size and egg mass in a wild great tit (Parus major) population. *Mol. Ecol.*, **22**:3949–3962. doi:10.1111/mec.12376

Santure, A.W., Poissant, J., De Cauwer, I., van Oers, K., Robinson, M.R., Quinn, J.L., Groenen, M.A., Visser, M.E., Sheldon, B.C., & Slate, J. (2015). Replicated analysis of the genetic architecture of quantitative traits in two wild great tit populations. *Mol. Ecol.*, **24**(24):6148-62. doi:10.1111/mec.13452

Santure, A.W. (2016). An ecological model organism flies into the genomics era. *Mol. Ecol. Resour.*, **16**:379-381. doi:10.1111/1755-0998.12491

Saravanan, K.A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., Dutt, T., Mishra,B.P., & Singh, R.K. (2021). Genomic scans for selection signatures revealed candidate

genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, **113**(3):955-963. doi:10.1016/j.ygeno.2021.02.009

Savolainen, O., Lascoux, M., & Merilä, J. (2013). Ecological Genomics of Local Adaptation. *Nat. Rev. Genet.*, **14**(11):807–820. doi:10.1038/nrg3522

Schmitt, T. (2007). Molecular biogeography of Europe: Pleistocene cycles and postglacial trends. *Front. Zool.*, **4**:e11. doi:10.1186/1742-9994-4-11

Schoeman, D.S., Gupta, A. Sen, Harrison, C.S., Everett, J.D., Brito-Morales, I., Hannah, L., Bopp, L., Roehrdanz, P.R., & Richardson, A.J. (2023). Demystifying Global Climate Models for Use in the Life Sciences. *Trends Ecol. Evol.*, **38**(9):843–858. doi:10.1016/j.tree.2023.04.005

Serranito, B., Cavalazzi, M., Vidal, P., Taurisson-Mouret, D., Ciani, E., Bal, M., Rouvellac, E., Servin, B., Moreno-Romieux, C., Tosser-Klopp, G., Hall, S.J.G., Lenstra, J.A., Pompanon, F., Benjelloun, B., & Da Silva, A. (2021). Local adaptations of Mediterranean sheep and goats through an integrative approach. *Sci. Rep.*, **11**:e21363. doi:10.1038/s41598-021-00682-z

Sheldon, B.C., Kruuk, L.E.B., & Alberts, S.C. (2022). The expanding value of long-term studies of individuals in the wild. *Nat. Ecol. Evol.*, **6**(12):1799–1801. doi:10.1038/s41559-022-01940-7

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., Groza, T., Güneş, O., Hall, P., Hayhurst, J., Ibrahim, A., Ji, Y., John, S., Lewis, E., MacArthur, J.A.L., McMahon, A., Osumi-Sutherland, D., Panoutsopoulou, K., Pendlington, Z., ... & Harris, L.W. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Res.*, **51**(D1):D977–D985, doi:10.1093/nar/gkac1010

Song, G., Zhang, R., Machado-Stredel, F., Alström, P., Johansson, U.S., Irestedt, M., Mays Jr., H.L., McKay, B.D., Nishiumi, I., Cheng, Y., Qu, Y., Ericson, P.G.P., Fjeldså, J., Townsend Peterson, A., & Lei, F. (2020). Great journey of Great Tits (Parus major group): Origin, diversification and historical demographics of a broadly distributed bird lineage. *J. Biogeogr.*, **47**:1585–1598. doi:10.1111/jbi.13863

Spurgin, L.G., Bosse, M., Adriaensen, F., Albayrak, T., Barboutis, C., Belda, E., Bushuev, A., Cecere, J.G., Charmantier, A., Cichon, M., Dingemanse, N.J., Doligez, B., Eeva, T., Einar Erikstad, K., Fedorov, V., Griggio, M., Heylen, D., Hille, S., Hinde, C.A., ... & Slate, J. (2024). The great tit HapMap project: a continental-scale analysis of genomic variation in a songbird. *Mol. Ecol. Resour.*, **24**:e13969. https://doi.org/10.1111/1755-0998.13969

Szpiech, Z.A., Novak, T.E., Bailey, N.P., & Stevison, L.S. (2021). Application of a Novel Haplotype-based Scan for Local Adaptation to Study High-altitude Adaptation in Rhesus Macaques. *Evol. Lett.*, **5**:408–421. doi:10.1002/evl3.232

Stumpf, M. (2004). Haplotype diversity and SNP frequency dependence in the description of genetic variation. *Eur. J. Hum. Genet.*, **12**:469–477. doi:10.1038/sj.ejhg.5201179

Sunde, J., Franzén, M., Betzholtz, P-E., Francioli, Y., Pettersson, L.B., Pöyry, J., Ryrholm, N., & Forsman, A. (2023). Century-long butterfly range expansions in northern Europe depend on climate, land use and species traits. *Commun. Biol.*, **6**:e601. doi:10.1038/s42003-023-04967-z

Taberlet, P., Fumagalli, L., Wust-Saucy, A-G., & Cosson, J-F. (1998). Comparative phylogeography and postglacial colonization routes in Europe. *Mol. Ecol.*, **7**:453-464. doi:10.1046/j.1365-294x.1998.00289.x

Tang, K., Thornton, K.R., & Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biol.*, **5**(7):1587–1602. doi:10.1371/journal.pbio.0050171

Thomas, L., Underwood, J.N., Rose, N.H., Fuller, Z.L., Richards, Z.T., Dugal, L., Grimaldi, C.M., Cooke, I.R., Palumbi, S.R., Gilmour, J.P. (2022). Spatially varying selection between habitats drives physiological shifts and local adaptation in a broadcast spawning coral on a remote atoll in Western Australia. *Sci. Adv.*, **8**(17):eabl9185. doi:10.1126/sciadv.abl9185

Tian, S., Zhou, X., Phuntsok, T., Zhao, N., Zhang, D., Ning, C., Li, D., & Zhao, H. (2020). Genomic Analyses Reveal Genetic Adaptations to Tropical Climates in Chickens. *iScience*, **23**(11):e101644. doi:10.1016/j.isci.2020.101644

Tiffin, P. & Ross-Ibarra, J. (2014). Advances and limits of using population genetics to understand local adaptation. *Trends Ecol. Evol.*, **29**(12):673-680. doi:10.1016/j.tree.2014.10.004

Tinbergen, J.M., & Boerlijst, M.C. (1990). Nestling Weight and Survival in Individual Great Tits (Parus Major). *J. Anim. Ecol.*, **59**(3):1113–27. doi:10.2307/5035

Tishkoff, S.A., Reed, F.A., Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., Ibrahim, M., Omar, S.A., Lema, G., Nyambo, T.B., Ghori, J., Bumpstead, S., Pritchard, J.K., Wray, G.A. & Deloukas, P. (2007). Convergent Adaptation of Human Lactase Persistence in Africa and Europe. *Nat. Genet.*, **39**(1):31–40. doi:10.1038/ng1946

Tomczak, A., Mortensen, J.M., Winnenburg, R., Liu, C., Alessi, D.T., Swamy, V., Vallania, F., Lofgren, S., Haynes, W., Shah, N.H., Musen, M.A., & Khatri, P. (2018). Interpretation of biological experiments changes with evolution of the Gene Ontology and its annotations. *Sci. Rep.*, **8**:e5115. doi:10.1038/s41598-018-23395-2

Uffelmann, E., Huang, Q.Q., Munung, N.S., de Vries, J., Okada, Y., Martin, A.R., Martin, H.C., Lappalainen, T., & Posthuma, D. (2021). Genome-Wide Association Studies. *Nat. Rev. Methods Primers*, **1**:e59. doi:10.1038/s43586-021-00056-9

van Balen, J.H. (1073). A Comparative Sudy of the Breeding Ecology of the Great Tit Parus major in Different Habitats. *Ardea*, **55**(1–2):1-93. doi: 10.5253/arde.v61.p1

van Bers, N.E.M., Santure, A.W., Van Oers, K., Cauwer, I.S., Dibbits, B.W., Mateman, C.,Crooijmans, R.P.M.A., Sheldon, B.C., Visser, M.E., Groenen, M.A., & Slate, J. (2012). The design and cross-population application of a genome-wide SNP chip for the great tit Parus major. *Mol. Ecol. Res.*, **12**:753–770. doi:10.1111/j.1755-0998.2012.03141.x

Van der Auwera, G.A., Carneiro, M., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K., Altshuler, D., Gabriel, S., & DePristo, M. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.*, **43**(1110):11.10.1-11.10.33. doi:10.1002/0471250953.bi1110s43

van Oers, K., Santure, A.W., De Cauwer, I., van Bers, N.E.M., Crooijmans, R.P.M.A., Sheldon, B.C., Visser, M.E., Slate, J., & Groenen, M.A.M. (2014). Replicated independent high-density genetic maps of two great tit populations reveal new insights into avian genome structure and evolution. *Heredity*, **112**:307–316. doi:10.1038/hdy.2013.107

van't Hof, A.E., Campagne, P., Rigden, D.J., Yung, C.J., Lingley, J., Quail, M.A., Hall, N., Darby, A.C., & Saccheri, I.J. (2016). The industrial melanism mutation in British peppered moths is a transposable element. *Nature*, **534**(7605):102–105. doi:10.1038/nature17951

Visscher, P.M., Wray, N.R., Zhang, Q., Sklar, P., McCarthy, M.I., Brown, M.A., & Yang, J. (2017). 10 Years of GWAS Discovery: biology, function, and translation. *Am. J. Hum. Genet.*, **101**:5–22. doi:10.1016/j.ajhg.2017.06.005

Visser, M.E., van Noordwijk, A.J., Tinbergen, J.M., & Lessells, C.M. (1998). Warmer springs lead to mistimed reproduction in great tits (Parus major). *Proc. R. Soc. Lond. B.*, **265**:1867–1870. doi:10.1098/rspb.1998.0514

Visser, M.E., & Both, C. (2005). Shifts in Phenology Due to Global Climate Change: The Need for a Yardstick. *Proc. R. Soc. B: Biol. Sci.*, **272**(1581):2561–2569. doi:10.1098/rspb.2005.3356

Visser, M.E., Lindner, M., Gienapp, P., Long, M.C., & Jenouvrier, S. (2021). Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (Parus major*). Proc. R. Soc. B*, **288**:e20211337. doi:10.1098/rspb.2021.1337

Voight, B.F., Kudaravalli, S., Wen, X., & Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.*, **4**(3):446–58. doi:10.1371/journal.pbio.0040072

Wade, D., Ashton-Butt, A., Scott, G., Reid, S.M., Coward, V., Hansen, R.D.E., Banyard, A.C., & Ward, A.I. (2022). High pathogenicity avian influenza: targeted active surveillance of wild birds to enable early detection of emerging disease threats. *Epidemiol Infect.*, **151**:e15. doi: 10.1017/S0950268822001856

Wadgymar, S.M., Lowry, D.B., Gould, B.A., Byron, C.N., Mactavish, R.M., & Anderson, J.T. (2017), Identifying targets and agents of selection: innovative methods to evaluate the processes that contribute to local adaptation. *Methods Ecol. Evol.*, **8**:738-749. doi:10.1111/2041-210X.12777

Wadgymar, S.M., DeMarche, M.L., Josephs, E.B., Sheth, S.N., & Anderson, J.T. (2022). Local Adaptation: Causal Agents of Selection and Adaptive Trait Divergence. *Annu. Rev. Ecol. Evol. Syst.*, **53**(1):87-111. doi:10.1146/annurev-ecolsys-012722-035231

Waldvogel, A-M., Schreiber, D., Pfenninger, M., & Feldmeyer, B. (2020). Climate Change Genomics Calls for Standardized Data Reporting. *Frontiers Ecol. Evol.*, **8**: e242. doi:10.3389/fevo.2020.00242

Walther, G.-R., Post, E., Convey, P., Menzel, A., Parmesank, C., Beebee, T.J.C., Fromentin, J.-M., Hoegh-Guldberg, O., & Bairlein, F. (2002). Ecological Response to Recent Climate Change. *Nature*, **416**:389–395. doi:10.1038/416389a

Wang, K., Dickson, S.P., Stolle, C.A., Krantz, I.D., Goldstein, D.B., & Hakonarson, H. (2010). Interpretation of association signals and identification of causal variants from genome-wide association studies. *Am. J. Hum. Genet.*, **86**:730–742. doi: 10.1016/j.ajhg.2010.04.003

Warren, W., Clayton, D., Ellegren, H., Arnold, A.P., Hillier, L.W., Künstner, A., Searle, S., White, S., Vilella, A.J., Fairley, S., Heger, A., Kong, L., Ponting, C.P., Jarvis, E.D., Mello, C.V., Minx, P., Lovell, P., Velho, T.A.F., Ferris, M., ... & Wilson, R.K.(2010). The genome of a songbird. *Nature*, **464**:757–762. doi:10.1038/nature08819

Wedler, M., Pinto, J.G., & Hochman, A. (2023). More frequent, persistent, and deadly heat waves in the 21st century over the Eastern Mediterranean. *Sci. Total Environ.*, **870**:e161883. doi:10.1016/j.scitotenv.2023.161883

White, N.J., Beckerman, A.P., Snook, R.R., Brockhurst, M.A., Butlin, R.K., & Eyres, I. (2022). Experimental evolution of local adaptation under unidimensional and multidimensional selection. *Curr. Biol.*, **32**(6):1310-1318.e4. doi:10.1016/j.cub.2022.01.048

Wright, S. (1949). The genetical structure of populations. *Ann. Eugen.*, **15**:323-354. doi:10.1111/j.1469-1809.1949.tb02451.x

Xu, S., Li, S., Yang, Y., Tan, J., Lou, H., Jin, W., Yang, L., Pan, X., Wang, J., Shen, Y., Wu, B., Wang, H., Jin, L. (2011). A Genome-Wide Search for Signals of High-Altitude Adaptation in Tibetans. *Mol. Biol. Evol.*, **28**(2):1003–1011. doi:10.1093/molbev/msq277

Yeaman, S. (2022). Evolution of polygenic traits under global vs local adaptation. *Genetics*, **220**(1):iyab134. doi:10.1093/genetics/iyab134

Zhang, G., Li, C., Li, Q., Li, B., Larkin, D.M., Lee, C., Storz, J.F., Antunes, A., Greenwold, M.J., Meredith, R.W., Ödeen, A., Cui, J., Zhou, Q., Xu, L., Pan, H., Wang, Z., Jin, L., Zhang, P., Hu, H., ... & Froman, D.P. (2014). Comparative genomics reveals insights into avian genome evolution and adaptation. *Science*, **346**:1311-1320. doi:10.1126/science.1251385

Zhang, Y., Xue, X., Liu, Y., Abied, A., Ding, Y., Zhao, S., Wang, W., Ma, L., Guo, J., Guan, W., Pu, Y., Mwacharo, J.M., Han, J., Ma, Y., & Zhao, Q. (2021). Genome-wide comparative analyses reveal selection signatures underlying adaptation and production in Tibetan and Poll Dorset sheep. *Sci. Rep.*, **11**:e2466. doi:10.1038/s41598-021-81932-y

2. The genomics of adaptation to climate in European great tit (Parus major) populations.

Stonehouse, J.C., Spurgin, L.G., Laine, V.N., Bosse, M., The Great Tit HapMap Consortium, Groenen, M.A.M., van Oers, K., Sheldon, B.C., Visser, M.E., Slate, J. (2024). The genomics of adaptation to climate in European great tit (*Parus major*) populations. *Evolution Letters*, **8**(1):18-28. doi: https://doi.org/10.1093/evlett/qrad043. Made available under a Creative Commons Attribution Licence (CC BY 4.0).

Article Type: Letter

The genomics of adaptation to climate in European great tit (*Parus major*) populations

Joanne C Stonehouse[1], Lewis G Spurgin[2], Veronika N Laine[3,6], Mirte Bosse[4,7], The Great Tit HapMap Consortium, Martien AM Groenen[4], Kees van Oers[3], Ben C Sheldon[5], Marcel E Visser[3], Jon Slate[1]

1. School of Biosciences, University of Sheffield, Western Bank, Sheffield, S10 2TN, UK
2. School of Biological Sciences, University of East Anglia, Norwich Research Park, United Kingdom
3. Department of Animal Ecology, Netherlands Institute of Ecology (NIOO-KNAW), Wageningen, the Netherlands
4. Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands
5. Edward Grey Institute, Department of Biology, University of Oxford, United Kingdom
6. Finnish Museum of Natural History, University of Helsinki, Helsinki, Finland
7. Amsterdam Institute for Life and Environment (A-LIFE), Section Ecology and Evolution, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

Keywords: climate adaptation, great tit HapMap project, signatures of selection, Genome-environment analysis (GEA), GO term enrichment, climate change.

2.1. Abstract

The recognition that climate change is occurring at an unprecedented rate means that there is increased urgency in understanding how organisms can adapt to a changing environment. Wild great tit (*Parus major*) populations represent an attractive ecological model system to understand the genomics of climate adaptation. They are widely distributed across Eurasia and they have been documented to respond to climate change. We performed a Bayesian genome-environment analysis, by combining local climate data with SNP genotype data from 20 European populations (broadly spanning the species' continental range). We found 36 genes putatively linked to adaptation to climate. Following an enrichment analysis of biological process Gene Ontology (GO) terms, we identified over-represented terms and pathways among the candidate genes. Because many different genes and GO terms are associated with climate variables, it seems likely that climate adaptation is polygenic and genetically complex. Our findings also suggest that geographical climate adaptation has been occurring since great tits left their Southern European refugia at the end of the last ice age. Finally, we show that substantial climate-associated genetic variation remains, which will be essential for adaptation to future changes.

2.2. Impact Summary

Whether and how organisms adapt to variation in climate remain challenging questions. However, advances in ecological genomics tools and statistical genetics methods have made it feasible to identify genomic regions that have possibly responded to selection. Here we study pan-European populations of the ecological model species, the great tit (*Parus major*), to try to identify genes and genetic pathways that have possibly enabled different populations to adapt to past and present climate conditions. By: (i) genotyping 20 different great tit populations at 479,590 SNPs, and looking for associations with climate data collected from each location between the 1970s and 2000s, and (ii) Using Gene Ontology (GO) enrichment analysis to infer the biological roles of climate-associated genes, we aimed to better understand the genetics of climate adaptation. Up to 36 different genes were found to be significantly associated with adaptation to climate, and several different biological processes were over-represented among these genes. Our findings suggest that adaptation to multiple climate variables has occurred, with many genes enabling different populations to adapt to their current and past environmental climatic conditions. It is likely that adaptive genetic variation persists in contemporary great tit populations.

2.3. Introduction

Terrestrial ecosystems face an unprecedented rate of climate change in the coming decades. A 1.5°C global surface temperature warming (above pre-industrial levels), caused by human activities and emissions of greenhouse gases and aerosol pathways, is likely by 2040 (IPCC 2022). Likely impacts on climate include an increase in the number of regions with raised hot and cold temperature extremes per season, and changes to the intensity, frequency and duration of extreme weather events, such as heatwaves, drought and flooding (Eyring et al. 2016). Further consequences of climate change that threaten populations include forest fires, extreme weather events, the spread of invasive species, pests and disease, and the decreased functioning of ecosystem services. Hence, an urgent issue is understanding how organisms respond to the changing environment (IPCC 2022; Root et al. 2003). How a species' physiology and behaviour tolerates changes to the local environment will determine its fitness and survival, and therefore the need to respond to climate change (Cheviron and Brumfield 2012; Khaliq et al. 2014). Sessile plants, low-dispersal insects and some unique or threatened vertebrates e.g. small island populations, are especially vulnerable to local extinction if their response is insufficient (Parmesan and Yohe 2003; Khaliq et al. 2014), although a failure to adapt to climate change is expected to be consequential in any population.

Where populations are able to respond to climate change, they can do so in three possible ways: (1) Populations may shift range to more suitable conditions. For example, poleward distribution shifts in response to anthropogenic climate warming are observed in many plants, birds and butterflies (Parmesan et al. 1999; Hickling et al. 2006; Chen et al. 2011; Devictor et al. 2012; Sunday et al. 2012). (2) Through plasticity of key phenotypic traits, such as behaviour and the timing (phenology) of reproduction and migration (Parmesan 2006). (3) By evolving to the new conditions i.e. adaptation via selection favouring particular alleles which have some form of fitness advantage. However, the evolution of climate adaptation due to specific genomic regions/genes might not be fast enough to match climate change, especially if multiple climatic variables are changing simultaneously and/or generation

times are long. The evolutionary potential to adapt has consequences for population persistence (Khaliq et al. 2014; Parmesan 2006). Furthermore, options (2) and (3) are not necessarily mutually exclusive, as both processes can occur, and plasticity itself may be heritable (i.e. an evolvable trait). Understanding the role of plastic vs microevolutionary changes in response to climate change requires further investigation in many taxa, including birds (Charmantier and Gienapp 2014), the focus of this study. Here we explore associations between genetic variation putatively under selection and climatic variables, in order to better understand microevolutionary responses to climate.

There are two distinct approaches for using genomic data to identify adaptive genetic variation and signatures of natural selection. First, molecular quantitative genetics approaches using phenotypic information. For example, linkage mapping or genomewide association studies (GWAS) can be used to identify quantitative trait loci (QTL) linked to phenotypes involved in adaptation (Santure and Garant 2018). Where phenotypic data are lacking, a second approach, population genomics, uses genomic information (but no phenotypic measurement) to identify signatures of adaptive genetic variation and then attempts to relate genome regions under selection to evolutionary processes and environmental variation (Rellstab et al. 2015). Population genomics attempts to find regions of the genome that are more highly differentiated between populations than is expected from neutral processes – i.e. to find ‘outlier loci’, that are inferred to be involved in local adaptation (Luikart et al. 2003; Butlin 2010). Outlier locus approaches have been used extensively to reveal signatures of local adaptation e.g. to different host plants in walking-stick insects (Nosil et al. 2012), to different shoreline environments in marine snails (Butlin 2010), and to different predator regimes in threespine sticklebacks (Mazzarella et al. 2016). In birds, outlier loci associated with adaptation to living at high altitude in Tibetan Ground tits (Qu et al. 2013), to different forest habitats in blue tits (Perrier et al. 2020) and to urban landscapes in great tits (Salmón et al. 2021) have been found. An extension of the outlier locus approach is to use genome-environment association (GEA) methods to identify loci whose divergence is correlated with an environmental covariate of interest. GEA analyses are more challenging to implement than outlier loci analyses because they require: i)

multiple populations, ii) a measure of the environmental variable of interest and iii) confidence that the environmental variable is the agent of selection. For comparisons of genome-scan methods see Gautier (2015), Forester et al. (2018) and Booker et al. (2024).

Terrestrial species with populations that occupy a wide diversity of habitats/climatic conditions, but where gene flow persists, provide ideal systems to investigate the genomic responses to past (e.g. since the last ice age 10-15 thousand years ago) and more recent (e.g. post-industrialization) climate changes (Rellstab et al. 2015). Ongoing gene flow is useful because in the resulting relatively low differentiation across the genome, make regions responsible for adaptation more likely to stand out as outliers (Luikart et al. 2003). The aim of this paper is to look for evidence across the genome of climate adaptation in a well-studied, model wild vertebrate species, the great tit *Parus major*. Previous genome-wide 'outlier locus' analysis in great tits have not considered whether allele frequencies covary with environmental variables (Bosse et al. 2017; Spurgin et al. 2024), and so the agents of selection could not be identified.

The great tit is a well suited species to understanding the genomics of climate adaptation because: (i) They occupy a wide range of environments and climates across Eurasia, with range edges extending into Eastern Russia, Fenno-Scandinavia and North Africa (Gosler et al. 2016). (ii) As cavity breeders they readily use nest boxes, which has allowed long-term monitoring of wild populations e.g. in Wytham Woods, UK (Lack 1964) and Hoge Veluwe, NL (Van Balen 1973). (iii) They are generally non-migratory, although some migration is observed in northern populations (Kvist et al. 1999), so populations should be well adapted to their local environment. Dispersal distances are usually short, but long-distance dispersal of some individuals facilitates gene-flow among subpopulations, meaning there is only minor population differentiation, with Fst typically around 0.01 (Laine et al. 2016; Lemoine et al. 2016; Spurgin et al. 2024; van Bers et al. 2012). The current distribution has likely spread after the last ice age from a single restricted refugium in South-Eastern Europe (Spurgin et al. 2024).

Great tits are sensitive to climate change, with warmer spring temperatures causing earlier egg laying across Europe (Visser et al. 1998; Visser et al. 2021; Charmantier et al. 2008; Cole et al. 2021). The timing of breeding is especially important in the tri-trophic foodweb (oak bud burst-winter moth caterpillar-great tit), to ensure peak food resources to raise chicks (Perrins 1970; Visser et al. 1998; Hinks et al. 2015; Burgess et al. 2018). An optimal phenology is critical for population persistence (Vedder et al. 2013 but see Reed et al. 2013). Great tits, and passerines more generally, may have sufficient phenotypic plasticity in egg lay date to track environmental changes, enabling them to cope with recent rapid rates of spring warming in the UK (Phillimore et al. 2016), although this plastic response is complex (Ramakers et al. 2018). If limits to phenotypic plasticity prevent an ongoing response to global climate change across parts of the distribution of *Parus major*, then evolutionary changes will be necessary to enable future population resilience and survival.

Rich genomic resources are available for genetic research of great tits. It was only the second passerine to have a high-quality reference genome sequenced and annotated (Laine et al. 2016); a genetic linkage map has been made (van Oers et al. 2014); DNA samples are available from multiple populations; a high density ~500K SNP chip has been produced (Kim et al. 2018); and the SNP chip has been used to genotype individuals from approximately 40 wild populations (Kim et al. 2018; Spurgin et al. 2024; Salmón et al. 2021). Samples for the cross-population analyses were contributed by Great Tit Hap Map Consortium, a group of researchers conducting field studies of great tits across much of the species range. The consortium was established with the aim of characterising inter-population genetic variation and hopefully to identify genes associated with local adaptation. Great tit populations have low Fst across the genome, which means any outlier loci should stand out against a low background of differentiation. Additionally, regions of the genome under natural selection have already been identified using outlier locus approaches (Bosse et al. 2017). Responses to selection appear to be greatest in populations at the range edges, and are perhaps associated with morphology, plumage colour and stress response (Spurgin et al. 2024).

Here we address the question: what regions of the genome have enabled a wild bird species to adapt to different climate conditions? We first describe a genome-environment analysis (GEA) to identify genes that are associated with climate adaptation and show signatures of selection. We note however that an association with climate is not definitive evidence of adaptation to climate. Next, we use a Gene Ontology (GO) term enrichment analysis to identify over-represented biological functions among the genes with climate-associated SNPs.

2.4. Materials and Methods

The Great Tit HapMap project field data and genotyping

Members of the Great Tit HapMap Project consortium (Spurgin et al. 2024) provided blood samples from wild European great tit populations that are the focus of ongoing ecological studies. A total of 838 great tits from 29 locations in Europe were sampled across 22 countries, encompassing much of the species range. In most cases the DNA was extracted using ammonium acetate/high salt extraction protocols (Kim et al. 2018), although a small number of samples were sent as extracted DNA. A custom developed Affymetrix® great tit 650K SNP chip (Kim et al. 2018) was used to genotype the samples at Edinburgh Genomics, UK.

Data Filtering and QC

Following filtering approaches described elsewhere (Bosse et al. 2017; Spurgin et al. 2024), the data set contained a total of 647 great tits from 29 populations, typed at 483,888 single nucleotide polymorphisms (SNPs). Filtering was applied in PLINK v1.90 (Chang et al. 2015) with parameter thresholds: --maf (minor allele frequency) 0.01, --geno (proportion of individuals typed at a SNP) 0.8, --not-chr 35 (remove markers from 'chromosome' 35), --rel-cutoff (relatedness cutoff) 0.25. This removed SNPs that were almost lacking in variation, that had high rates of missing data, or that were assigned to an unknown chromosome (chr 35 is a collection of scaffolds on unknown chromosomes in our Plink files). Only males were included for the Z chromosome, as females are hemizygous. In addition, populations with less than 15 individuals were excluded. The HapMap dataset is known to include some samples of close relatives, such as trios of parents and chick. Having close relatives causes much higher amounts of population structure, which can give misleading results in population genetic analyses. Pairwise-relatedness was estimated in PLINK and the dataset filtered, to exclude one member of a pair when the relatedness was greater than 0.25. This step ensured only one individual per trio was used. After passing quality control, 479,590 SNPs and 535 samples were retained from 20 populations across Europe (**Figure 2.1**; **Table S2.1**).

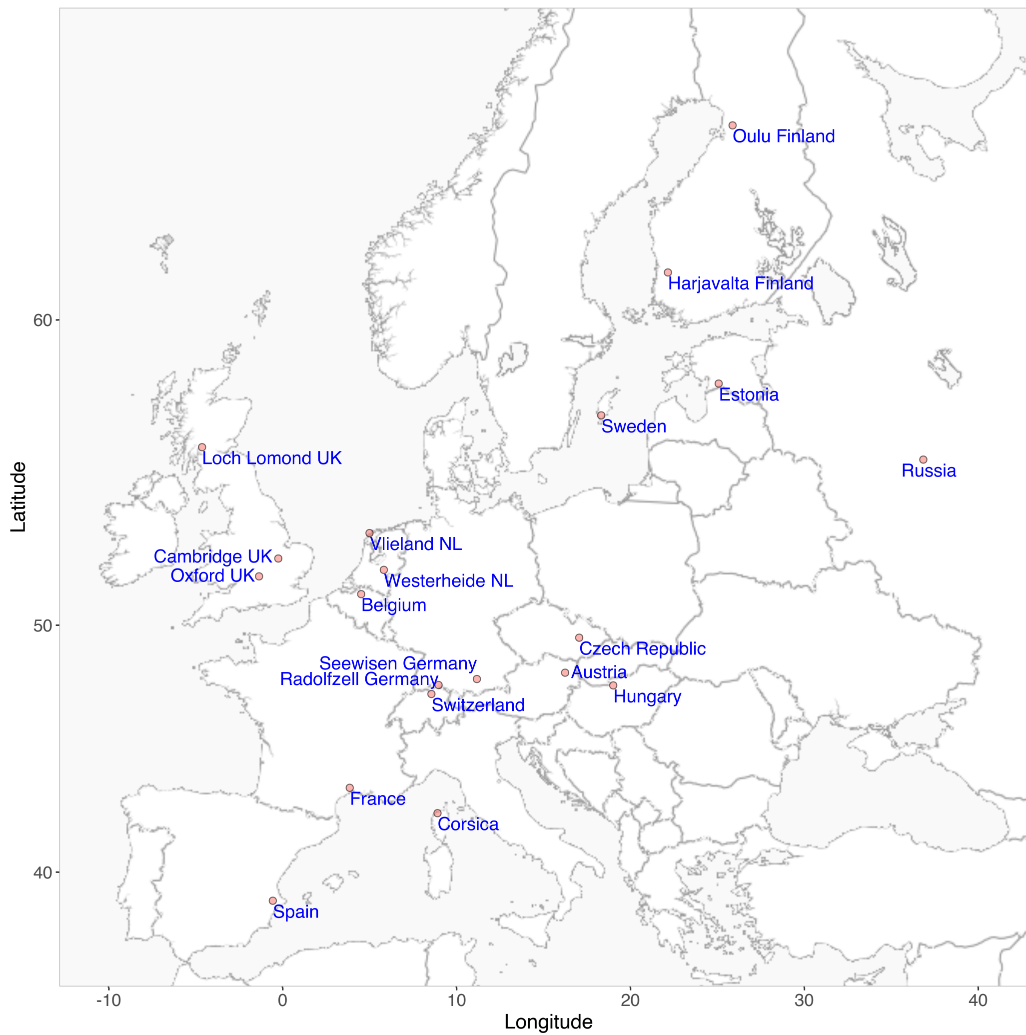


Figure 2.1: The 20 great tit population locations. Further details of each population are in Table S1. Map tiles by Stamen Design (https://stamen.com), under CC BY 3.0. Contains data by OpenStreetMap (http://openstreetmap.org/copyright), under ODbL.

Climate data collection and Principal Component Analysis

The WorldClim v2 data set provides good spatial resolution climate data across Europe (interpolated from monthly-averaged weather station records 1970-2000) (Fick and Hijmans 2017). Twenty two variables (19 bioclimate variables, solar radiation, wind speed and water vapor pressure) were extracted for the 20 population locations (Table S2). The complexity of the dataset was reduced by a principal component analysis (PCA), using the FactoMineR package v2.4 (Lê et al. 2008), run in R version 3.6.1. Downstream analyses used the first four principal components for each location, scaled to mean=0, s.d.=1. Temperature and precipitation have been demonstrated to be associated with constraints on great tit growth rate (Eeva et al. 2020; Rodrignez and Barba 2016). More generally, a review of climatic effects on nestling growth and development in birds (Sauve et al. 2021) has highlighted solar radiation, wind speed and water vapour pressure as additional variables that might affect juvenile traits. Clearly, early life development is sensitive to climate effects (via, for example, changes in food availability and parental care), with potentially long-lasting effects on adult phenotypes and fitness. Thus, we decided to use all 22 available climatic variables, rather than *a priori* choose some and reject others.

Genome-environment analysis: Detecting loci potentially involved in adaptation to climate

To identify climate-associated SNPs, we used BayPass 2.1 (Gautier 2015), a software package that has been used to identify climate-associated loci in similar studies of *Arabidopsis thaliana* (Frachon et al. 2018), barley (Contreras-Moreira et al. 2019) and Mediterranean cattle breeds (Flori et al. 2019). We elected to use BayPass over Redundancy Analysis (RDA) as it controls for population structure, can handle missing genotype data, and is regarded as quite conservative in declaring a SNP as an outlier (Forester et al. 2018). We used BayPass in a two-step process, using default parameters unless stated otherwise. First, we ran the core model to calculate XtX, a differentiation statistic for each SNP. The core model also estimates the covariance in allele frequencies (in the form of a matrix Ω) that arises as a result of population structure; thus demographic effects that can otherwise cause false positives are accounted for. For further details of XtX see Coop et al. (2010), Günther and Coop (2013). The significance of the observed XtX statistics were obtained by generating a

null XtX distribution from pseudo-observed data (POD) simulated with 400K SNPs. Variants identified as highly differentiated ‘outlier’ SNPs at a 1% significance threshold were regarded as being loci putatively under selection. Note that at this stage, no associations with climatic variables have been tested for.

Second, for all 479,520 SNPs, the default AUX model with scaled covariance matrix and climate covariates was run independently three times (with different starting seeds) for the four PCs (analysed jointly in each run). Associations between each SNP and each climate PC were judged from Bayes Factors (expressed in deciban units, dB) under an MCMC sampling method (BFmc); evidence for a SNP being associated with climate adaptation was judged by Jeffreys’ Rule (Jeffreys 1961): $10 < BFmc \leq 15$ “strong evidence”; $15 < BFmc \leq 20$; “very strong evidence”; $BFmc > 20$ “decisive evidence”. The evidence was evaluated by comparing the observed associations with a null distribution generated from POD data simulated using the Ω generated in the core model i.e. to realistically simulate the true demographic history. A variant was evaluated on the strength of evidence threshold across all three runs, i.e. a variant would only be considered as providing ‘decisive’ evidence for association with a climate variable if it scored >20 for all three runs. The best potential candidate loci for climate adaptation were classified as SNPs that were identified as being both associated with climate (from the AUX analysis) and as an ‘outlier’ variant under selection (from the Core analysis).

Annotation of candidate loci for climate adaptation

All SNPs were annotated using the great tit reference genome (GCF_001522545.3_Parus_major1.1_genome.gff) converted to bed file with bedops (Neph et al. 2012), with the closest gene identified using bedtools closest sub-command (Quinlan 2014). This process annotated the gene the SNP was located in, or the nearest gene within 2.5 Kbp in either direction from the SNP. SNPs were regarded as unannotated if no gene was found within 2.5 kbp. An annotation window of 5-Kb was chosen because in great tits linkage disequilibrium (LD) falls to almost baseline levels within 10kbp or less (Spurgin et al. 2024).

Of the 479,590 variants considered, 278,861 variants were annotated and assigned to 14,529 unique genes (i.e. most genes have more than one annotated SNP).

Gene Ontology (GO) term enrichment analysis

Testing for GO term enrichment among potential climate adaptation loci.

A Gene Ontology (GO) term (Ashburner et al. 2000; The Gene Ontology Consortium 2021) analysis on gene scores was performed (separately for each of PC1-PC4), to look for GO terms enriched among the genes with climate-associated SNPs against the background annotated gene universe for all of the SNPs on the chip. For the 14,529 genes containing SNPs, or with SNPs within 2.5kbp, 8,867 great tit GO terms were retrieved from Ensembl (Ensembl Genes 109, *Parus major_1.1*), using the biomaRt package v3.8 (Durinck et al. 2005, 2009). To avoid over-rewarding genes with many SNPs we took an approach similar to one described elsewhere (Dickson et al. 2020). For each gene the number of SNPs in the gene or within 2.5kbp of it was counted. Next, the maximum BayPass BF statistic of any SNP in that gene was determined. Then, all genes with that number of SNPs were ranked by max(BF), and a 'P value' was determined by dividing the rank by the total number of genes with that number of SNPs. For example, if there were 100 genes with 8 SNPs, then the gene with the greatest max(BF) was given a value of 1/100 = 0.01, the gene with the second greatest max(BF) was given a value of 2/100 = 0.02, etc. Relatively few genes had a very large number of SNPs (427 genes had 100 or more SNPs), so binning was performed for genes with 100-149 SNPs, 150-199 SNPs, 200-299 SNPs, 300-499 SNPs and >= 500 SNPs. In total 14,529 genes were included in the analysis. Note that previous approaches calculated the mean test statistic across all SNPs within a gene (Dickson 2020), rather than used the maximum value observed at any SNP. Here, the maximum is appropriate, because linkage disequilibrium declines rapidly in great tits, meaning even true outlier loci are not expected to have a signal that extends across the gene.

Significance testing for GO term enrichment was performed with the topGO v2.36.0 (Alexa and Rahnenfuhrer 2020) Bioconductor (Release 3.9) package in R. The data container used as input for topGO held the following information: (i) the gene universe, functionally

annotated with GO terms, (ii) the gene score 'P value' for climate association, (iii) the gene-to-GO term(s) mapping, and (iv) the GO hierarchy to account for structure among GO terms (provided by the GO.db package v3.8.2 (Carlson 2019). The ontology parameter was set to investigate Biological Process (BP), Molecular function (MF) or Cellular Component (CC) individually. The node size pruned GO terms with less than 10 annotated genes. The default weight01 algorithm (based on the elim and weight methods Alexa et al. 2006) and Kolmogorov-Smirnov (KS) test statistic were implemented. For each enrichment test, the p-values computed by the default 'weight01' method were unadjusted for multiple testing. Multiple testing correcting procedures may be invalid because of the non-independence of the GO term hierarchy (the p-value of a GO term is conditioned on the neighboring terms) and with the annotation of multiple GO terms to one gene (Skarman et al. 2009).

2.5. Results

Climate Data PCA

The climate data set of 22 variables was reduced in dimensionality to four principal components (**Fig S2.1**) that captured 92.4% of the variance (38.7%, 33.2%, 14.5% and 6.0% respectively). Hence, the reported results will focus primarily on PC1 and PC2, with PC3 and PC4 results available in the supplementary material. The key contributing variables to the four climate variables were identified as follows: temperature across the year (PC1), precipitation in the driest and warmest periods (PC2), precipitation in the wettest months (PC3), and wind speed, precipitation seasonality and daily temperature range (PC4). Further details are in **Figs. S2.2-S2.4**.

Whole-genome scan: Detecting loci potentially involved in adaptation to climate

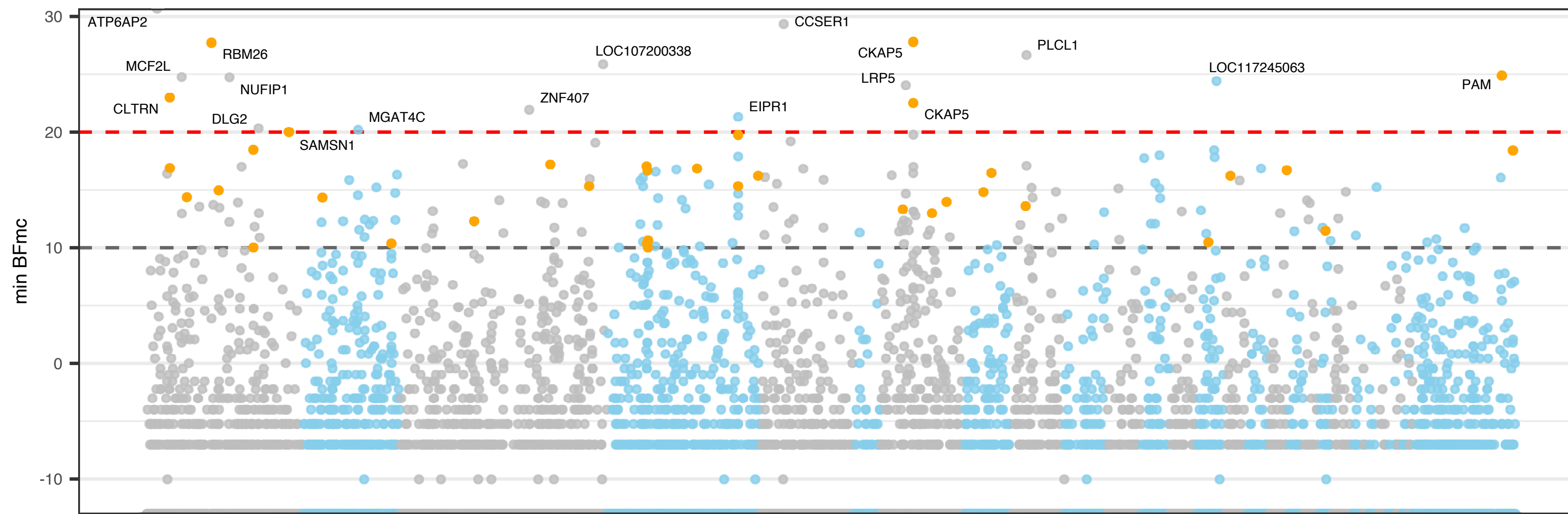
A summary of the covariance matrix, Ω, of SNP allele frequencies is plotted in the supplementary material (**Fig S2.5**). In general, population genetic structure is similar to that described in Spurgin et al. (2024), with the Corsica population standing out as the most differentiated from all other populations.

Identification of outlier loci under selection

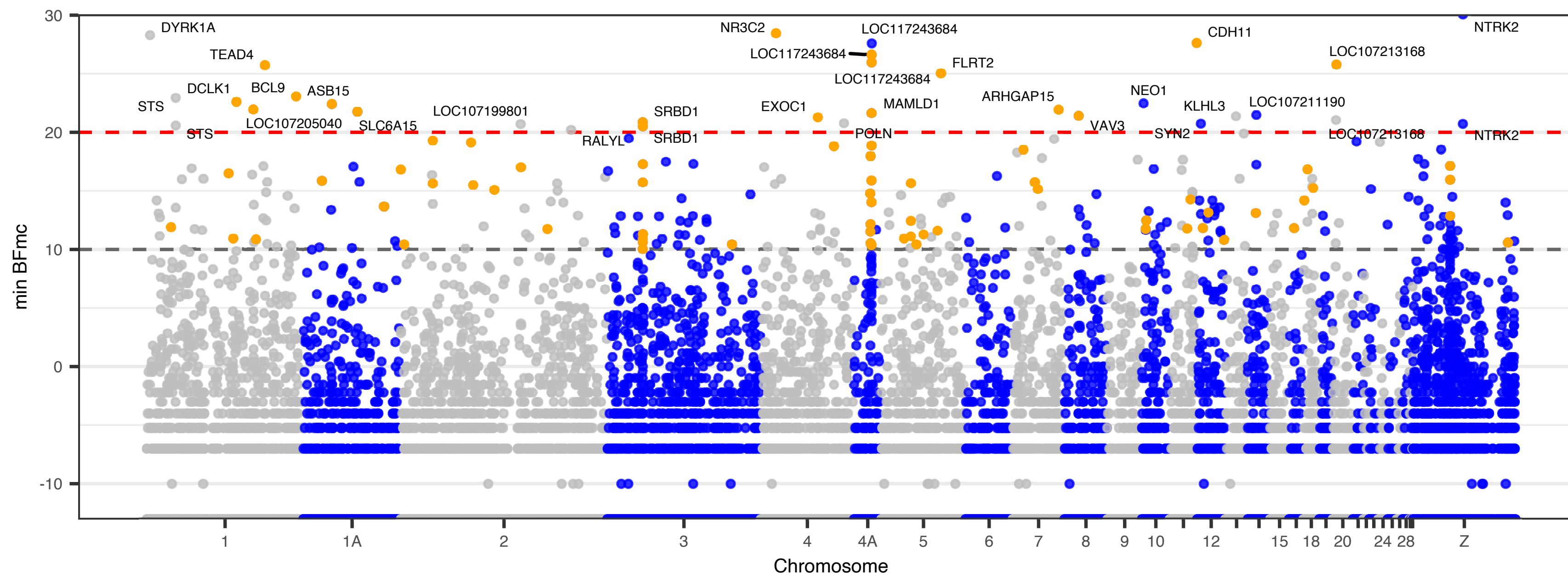
The population differentiation analysis using XtX statistics, simulated XtX values of 400,000 neutral SNPs, to infer a 1% $XtXPOD_{400,000}$ significance threshold of 25.88. Across the three runs of the core model, which gave similar results, a total of 4823 variants were identified as XtX outliers (**Fig S2.6**). These SNPs were situated in or near 1535 unique genes, which between them had 2730 annotations (GO terms). The number of genes and annotations is not expected to be the same because some genes have multiple annotations and some annotations are associated with multiple genes.

Genome-environment analysis (GEA): Identification of genes putatively involved with climate adaptation

At the decisive evidence level (BF ≥ 20 dB) 37 and 67 variants were associated with PC1 and PC2 respectively (**Table S2.3**). A total of 18 (PC1) and 32 (PC2) SNPs were found in annotated genes, of which 6 and 18 SNPs – more than expected by chance (binomial test, $P < 2.2 \times 10^{-16}$ for PC1 and PC2 respectively) - were also outliers in the core model (**Fig 2.2**). Details of each identified and annotated SNP are reported for all four climate PCs in the supplementary material (**Table S2.4**, **Figs S2.7-S2.8**). Allele frequencies in each population at the strongest candidate SNPs potentially involved in climate adaptation are included in **Fig. S2.9** (PC1) and **Fig. S2.10** (PC2).



min BFmc
30
20
10
0
-10
ATP6AP2
MCF2L
CLTRN
RBM26
NUFIP1
DLG2
SAMSN1
MGAT4C
ZNF407
LOC107200338
EIPR1
CCSER1
CKAP5
LRP5
CKAP5
PLCL1
LOC117245063
PAM



min BFmc
30
20
10
0
-10
DYRK1A
TEAD4
DCLK1
BCL9
ASB15
STS
STS
LOC107205040
SLC6A15
LOC107199801
RALYL
SRBD1
SRBD1
NR3C2
EXOC1
LOC117243684
LOC117243684
LOC117243684
MAMLD1
POLN
FLRT2
ARHGAP15
VAV3
NEO1
SYN2
CDH11
KLHL3
LOC107211190
LOC107213168
LOC107213168
NTRK2
NTRK2
1
1A
2
3
4
4A
5
6
7
8
9
10
12
15
18
20
24
28
Z
Chromosome

Figure 2.2: Manhattan plot of annotated genes associated with climate adaptation for a) PC1 and b) PC2 at the decisive evidence threshold (BFmc>20). For all annotated variants, the minimum BFmc score across the three runs is shown. The strong (BFmc>10), very strong (BFmc>15) and decisive evidence thresholds are indicated by dashed lines. SNPs highlighted orange (at threshold BFmc>10) were also found to be under selection in the outlier locus analyses that did not consider climatic data.

Gene Ontology (GO) term enrichment analysis

Biological Process GO term enrichment - PC1

A total of 4,428 GO terms found in 7,674 annotated genes were associated with biological processes. 41 Biological Process GO terms were significantly ($P < 0.01$) enriched within genes associated with PC1. The most enriched GO Term (GO:0061178) is involved in insulin secretion in cellular response to glucose (Table S2.5, **Appendix 2 data file GOtoGene_PC1.txt**). There was one candidate gene, CLTRN, that is potentially associated with climate adaptation (see **Table S2.7**) with a SNP that is significant for both the BF and XtX BayPass statistics and has enriched GO terms (regulation of biological quality GO: 0065008 and regulation of transporter activity GO: 0032409).

Biological Process GO term enrichment - PC2

For PC2, 29 Biological Process GO terms were significantly ($P < 0.01$) enriched. The two most enriched GO Terms were involved in plasma lipid clearance (GO:0034381) and eating behaviour (GO:0042755); see **Table S2.6**, **Appendix 2 data file GOtoGene_PC2.txt**). The enriched GO term ATP biosynthetic process (GO:0006754) was notable for being an annotation of genes with the SNPs that were significantly associated with both PC1 and PC2. Details For PCs 3 and 4, 28 and 23 significantly ($P < 0.01$) enriched Biological Process GO terms were identified (**Table S2.9-S2.10**, **Appendix 2 data files GOtoGene_PC3.txt and GOtoGene_PC4.txt**). Biological Process GO terms over-represented across any two or more of the PCs are provided in **Table S2.8**.

Molecular Function GO term enrichment test

For PC1 to PC4 respectively, 12, 8, 14 and 3 out of 740 Molecular Function GO terms were significantly ($P < 0.01$) enriched (**Tables S2.11-S2.14**).

Cellular Component GO term enrichment test

For PC1 to PC4 respectively, 11, 11, 6 and 6 out of 505 Cellular Component GO terms were identified as significantly ($P < 0.01$) enriched (**Tables S2.15-S2.18**).

2.6. Discussion

We report numerous genes and pathways in wild populations of great tits that are associated with a wide range of climate variables. The genes are spread across the genome, and there is segregating genetic variation within populations in these regions. There are no individual regions of the genome that appear to have undergone dramatic changes, such that populations experiencing the most extreme climatic conditions are fixed for putative positively selected alleles or haplotypes. Taken together, these observations suggest that local adaptation to past and present climate conditions is genetically complex, and has occurred, at least partially, through microevolutionary change. Given the polygenic nature of climate adaptation, individual loci have probably had small phenotypic effects. The data indicate that there remains substantial genetic variation associated with climate adaptation within populations. However, despite the weak genome-wide genetic structure across great tit populations, putative climate-adaptation regions do show some between-population differentiation. Given the considerable number of genes and pathways involved, climate adaptation loci are likely to be a large mutational target (Bay et al. 2018). That, and the presence of within- and between-population genetic variation in loci that possibly cause adaptation means there is the potential for great tits to adapt to future climate changes.

This is the first study to specifically look for genes associated with climate adaptation in this species. However, in the same wild great tit populations Spurgin et al. (2024) reported selection acting on genes thought to be associated with variation in morphology, stress response and colouration, with adaptation possibly strongest at the range edges of the species' distribution. The gene *CALM2* (biological function: calcium ion binding) was identified as a locus under selection in Spurgin et al. (2024) and in this study (here, it was associated with PC3 and PC4). Furthermore, *CALM2* has been reported as a candidate gene

for desert adaptation in sheep (Yang et al. 2016); in sheep it is involved in renal epithelial cell processes of water regulation and natriuresis (sodium excretion to stabilize blood pressure). Spurgin et al. (2024) found *CALM2* to be under selection in the English and Spanish populations. Here it was identified as the strongest candidate gene (i.e. greatest BF) for adaptation to variables influencing PC3 (associated with rain) and PC4 (associated with wind speed, rain seasonality and daily temperature range) (Table S4). Thus, this locus might be under selection in response to climatic conditions. However, none of the other outlier variants in Spurgin et al. (2024) were co-identified in this study, suggesting that signatures of selection at those loci result from adaptation to other selective pressures.

Similarly, a recent study identified loci under divergent selection between Wytham Woods (UK) and two Dutch great tit populations (Bosse et al. 2017). The highly significant outlier regions contained 28 annotated candidate genes involved in skeletal development, morphogenesis, and palate development. These loci were suggested to be involved in the adaptive evolution of longer-bills in UK populations, most notably at the *COL4A5* gene. In our study, two of the best candidate genes potentially associated with climate adaptation found, were also among significant outliers in the Bosse et al. (2017) study; two SNPs were identified in/near *SRBD1* [biological function: mRNA binding] associated with PC2, while eight SNPs were identified in/near *TRPS1* [biological function: cranio skeletal development] associated with PC3. Climate may affect what food sources are available, and that in turn may affect selection on beak and/or skull shape for foraging performance. For example, in Darwin's finches a severe El Niño event led to a scarcity of large seeds, which favoured selection on small beak sizes and led to an evolutionary change (Grant and Grant 1993). Thus, climate adaptation may be the explanation for signatures of selection at these two loci, but more generally, there is not much overlap between the potential climate adaptation loci reported here, and previous attempts to find genes under selection in these populations. Of course, that may be because this study explores populations experiencing a much wider range of climatic variation. Incorporating other possible agents of selection

into GEA models could enhance our understanding of what is shaping genomic variation at other highly differentiated loci.

A key question is whether the same suite of genes are involved in adaptation to climate in different species. We directly compared the set of genes found to be 'decisively' associated with climate in great tit, with twelve positively selected heat stress adaptation genes identified from a similar analysis in chicken populations experiencing desert and monsoon island climates (Tian et al. 2020). In our study nine of the twelve genes had a SNP within 2.5kbp, but none were associated with climate adaptation. However two of the genes (*ADCY1* and *VPS13C*) were identified in the pathways of enriched biological process GO terms (for PC1-4 and PC4 respectively) associated with climate adaptation in great tits (Table S19). Adaptation to heat stress has previously been demonstrated in genes linked to the arachidonic acid metabolism pathway in chickens (Tian et al. 2020), with *CYP2J21* and *CYP2J23* upregulated in pectoral muscles of heat stress treated broiler chickens, with oxidative stress recorded in the pectoral muscle (Liu et al. 2022). Two of those arachidonic acid metabolism genes (*DRD3* and *TNFRSF11A*) are annotated with one of the GO terms (GO:0065008 'regulation of biological quality') that were associated with PC1 in this study (Table S19). Caizergues et al. (2022) identified significant change in methylation linked to the *DRD3* gene in urban populations of great tit compared to forest environments. Extreme heating events in urban areas may lead to heat stress of nestlings and adults (Sumasgutner et al. 2023). Heat stress adaptation has previously been linked with convergent evolution of four genes (*PLA2G12b*, *GRP17, TNFRSF11A*, and *OC90*) between mammals and chicken (Tian et al. 2020). Of these genes, only *OC90* was co-identified in this GEA study, where it was associated with climate (PC4). There were no SNPs within 2.5kbp of *GRP17.* In our analyses, PC2 is the principal component most influenced by climate variables that are associated with extreme heat (Fig S2). There is no overlap between genes associated with PC2 and the heat-stress associated genes in the studies described above. There is little overlap between the 'decisive' candidate climate adaptation genes found here to those found in similar studies of more distantly related (i.e. non-avian) taxa such as *Drosophila* (Bogaerts-Márquez

et al. 2021) and cattle (Flori et al. 2019). An exception is the gene *BTBD7*, which is involved in protein ubiquitination, and was identified as a climate adaptation locus in *Drosophila* (Bogaerts-Márquez et al. 2021). Here, *BTBD7* is associated with PC3 (Table S20). Avian cells respond to temperature induced heat stress, via stimulated protein ubiquitination (Pritchett et al. 2023). Similarly, there is some evidence of overlap of the climate-associated enriched GO terms identified here with those in cattle (Flori et al 2019); these GO terms are associated with cellular signaling and nervous system function.

For both PC1 and PC2 the most strongly over-represented GO terms were not terms where we could find evidence of climate-adaptation in other species. For PC1 the top ranked GO term was 'regulation of insulin secretion'. This term may potentially be linked to climate-related effects on diet and hormones. For example, cold temperatures are thought to prime the glucose stress response in tree swallows (Ryan et al. 2023). For climate PC2, the top ranked GO Term is 'plasma lipoprotein particle clearance' (involved in the release of cholesterol). Climate variation may influence changes in food availability and nestling growth (Keller and Noordwijk 1994). Furthermore, poor nutrition in great tit embryos and early life stage may constrain development and adult health and fitness (Toledo et al. 2016). Thus, it is possible that variation in genes involved in pathways that affect plasma lipoproteins have provided an adaptive response to harmful climatic effects on nutrition.

Overall, the patterns suggest that climate adaptation gene discovery studies have identified different candidate genes that might be involved in adaptation, but that these candidates have some shared gene pathways for biological processes. There may be so little overlap because of the very different investigated taxa (invertebrates, domesticated livestock and birds), and also because of the very different physiological routes to climate adaptation (e.g. morphology, behaviour, physiology, neuro-endocrine, biochemical, metabolism, cellular and molecular responses). Additionally, different genes with related functions may enable adaptation to similar environments. Finally, when conducting GEA studies, it is necessary to be mindful that some 'hits' might be false positives, which are not expected to overlap

between different independent studies. One take-home message from this paper is that identifying individual loci that are definitively associated with climate adaptation is challenging. The combination of multiple climate-related variables driving selection, a polygenic genetic architecture of the (unknown) traits under selection, a lack of good candidates from other systems and an absence of selected loci going to fixation in specific populations all contribute to making gene discovery difficult.

Best practice for the confirmation of candidate genes from genome-environment analysis should ideally include an experimental verification, for example by functional genomics or gene expression studies (Rellstab et al. 2015). While beyond the scope of this study, such an approach is feasible in wild great tit populations, for example by rearing and selective breeding of wild-caught individuals in aviaries where environmental conditions can be controlled. For example, Gienapp et al. (2019) have selected for differences in lay date in aviary birds, and then looked at fitness of early and late layers when they were moved as eggs to nest boxes in the forest.

Methods to reliably forecast how populations may adapt to climate change (changes in allele frequencies) and how climate adaptation genes may be maintained across populations (population differentiation), will require testing of how observable (and possibly repeatable) the genetic effects described here are. Ongoing monitoring of allele frequencies in the study populations, in tandem with an increase in the number of individuals genotyped per population would improve power, as would the inclusion of data from additional populations. For those populations where climate change has been documented and samples are available over many years, it may be possible to look for associations between genotypes at climate-related SNPs and variation in fitness traits. In particular, it could be asked whether any selection at these SNPs varies temporally, in a way predicted by within population changes in climatic conditions (PC1-4). The identity and location of sites under selection could be improved with whole genome sequencing (which is now becoming much closer in cost to SNP chip genotyping). To validate the candidate

genes that have possibly enabled climate adaptation, and to date the timing of adaptive changes, it should be possible to investigate the genomes of museum samples, and formally test whether signatures of selection can be identified in the same genes over the last few hundred years of climate change (Clark et al. 2023).

Finally, it is possible that identifying adaptive genomic variation to climate variation may aid our understanding of the climatic drivers of population declines and genomic vulnerabilities. In summary, we found evidence that great tits have adapted over historical times to different climates through numerous evolutionary changes, and that the process is complicated, caused by selection at many different genes, probably driven by many different climatic variables. Most of the genes reported here have not previously been reported to be associated with climate adaptation. Given that many of the SNPs tagging regions responsible for adaptation are still segregating within populations, it seems likely that there remains substantial genetic variation within and between populations that can help further adaptation to future climate changes.

2.7. Acknowledgements

J.C.S. was funded by a PhD studentship from the Natural Environment Research Council, as part of the Adapting to the Challenges of a Changing Environment (ACCE) Doctoral Training Partnership. This work was supported by grants from the Natural Environment Research Council (grant NE/J012599/1 to J.S. and B.C.S) and the European Research Council (grant 202487 to J.S. and grant 339092 – E-Response to MEV) L.G.S was supported by fellowships from the Edward Grey Institute for Ornithology and the BBSRC (BB/N011759/1). The authors declare no conflicts of interest.

2.8. Author Contributions

JS conceived the study from data collected by the Great Tit HapMap Consortium project. JCS performed the computation and analysed the data. JS, LGS, MB and VNL performed quality control and downstream filtering of the SNP data. JS, KvO, MAMG and MEV supervised genomic data collection. BCS, JS, KvO, B and MEV organised sample collection, with members of the GTHC providing samples and ecological information relevant to their study populations. JS and JCS drafted the initial version of the manuscript and all authors contributed to later versions of the manuscript.

2.9. Data Accessibility

Data and scripts are available at:
https://doi.org/10.5061/dryad.vt4b8gtxd (Stonehouse et al. 2023).

2.10. References

Alexa, A., Rahnenführer, J., & Lengauer, T. (2006). Improved scoring of functional groups from gene expression data by decorrelating go graph structure. *Bioinformatics*, **22**(13), 1600–1607. https://doi.org/10.1093/bioinformatics/btl140

Alexa, A., & Rahnenführer, J. (2020). topGO: Enrichment analysis for Gene Ontology. R package version 2.42.0.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, **25**(1), 25–29. https://doi.org/10.1038/75556

Bay, R. A., Harrigan, R. J., Underwood, V. L., Gibbs, H. L., Smith, T. B., & Ruegg, K. (2018). Genomic signals of selection predict climate-driven population declines in a migratory bird. *Science*, **359**(6371), 83–86. https://doi.org/10.1126/science.aan4380

Bogaerts-Márquez, M., Guirao-Rico, S., Gautier, M., & González, J. (2021). Temperature, rainfall and wind variables underlie environmental adaptation in natural populations of Drosophila melanogaster. *Molecular Ecology*, **30**(4), 938-954. https://doi.org/10.1111/mec.15783

Booker, T. R., Yeaman, S., Whiting, J. R., & Whitlock, M. C. (2024). The WZA: A window-based method for characterizing genotype–environment associations. *Molecular Ecology Resources*, **24**(2): e13768. https://doi.org/10.1111/1755-0998.13768

Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., Gosler, A. G., McMahon, K., Poissant, J., Verhagen, I., Groenen, M. A. M., van Oers, K., Sheldon, B. C., Visser, M. E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, **358**(6361), 365–368. https://doi.org/10.1126/science.aal3298

Burgess, M. D., Smith, K. W., Evans, K. L., Leech, D., Pearce-Higgins, J. W., Branston, C. J., Briggs, K., Clark, J. R., Du Feu, C. R., Lewthwaite, K., Nager, R. G., Sheldon, B. C., Smith, J. A., Whytock, R. C., Willis, S. G., & Phillimore, A. B. (2018). Tritrophic phenological match-mismatch in space and time. *Nature Ecology and Evolution*, **2**, 970–975. https://doi.org/10.1038/s41559-018-0543-1

Butlin, R. K. (2010). Population genomics and speciation. *Genetica*, **138**(4), 409–418. https://doi.org/10.1007/s10709-008-9321-3

Caizergues, A. E., Le Luyer, J., Grégoire, A., Szulkin, M., Senar, J. -C., Charmantier, A., & Perrier, C. (2022). Epigenetics and the city: Non-parallel DNA methylation modifications across pairs of urban-forest Great tit populations. *Evolutionary Applications*, **15**(1), 149–165. https://doi.org/10.1111/eva.13334

Carlson, M. (2019). GO.db: A set of annotation maps describing the entire Gene Ontology. R package version 3.8.2.

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**(7), 1–16. https://doi.org/10.1186/s13742-015-0047-8

Charmantier, A., McCleery, R. H., Cole, L. R., Perrins, C., Kruuk, L. E. B., & Sheldon, B. C. (2008). Adaptive phenotypic plasticity in response to climate change in a wild bird population. *Science*, **320**(5877), 800–803. https://doi.org/10.1126/science.1157174

Charmantier, A., & Gienapp, P. (2014). Climate change and timing of avian breeding and migration: Evolutionary versus plastic changes. *Evolutionary Applications*, **7**(1), 15–28. https://doi.org/10.1111/eva.12126

Chen, I. -C., Hill, J. K., Ohlemüller, R., Roy, D. B., & Thomas, C. D. (2011). Rapid range shifts of species associated with high levels of climate warming. *Science*, **333**(6045), 1024–1026. https://doi.org/10.1126/science.1206432

Cheviron, Z. A., & Brumfield, R. T. (2012). Genomic insights into adaptation to high-altitude environments. *Heredity*, **108**(4), 354–361. https://doi.org/10.1038/hdy.2011.85

Clark, R. D., Catalano, K. A., Fitz, K. S., Garcia, E., Jaynes, K. E., Reid, B. N., Sawkins, A., Snead, A. A., Whalen, J. C., & Pinsky, M. L. (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Molecular Ecology Resources*, **00**, 1–17. https://doi.org/10.1111/1755-0998.13789

Cole, E. F., Regan, C. E., & Sheldon, B. C. (2021). Spatial variation in avian phenological response to climate change linked to tree health. *Nature Climate Change*, **11**(10), 872–878. https://doi.org/10.1038/s41558-021-01140-4

Contreras-Moreira, B., Serrano-Notivoli, R., Mohammed, N. E., Cantalapiedra, C. P., Beguería, S., Casas, A. M., & Igartua, E. (2019). Genetic association with high-resolution climate data reveals selection footprints in the genomes of barley landraces across the Iberian Peninsula. *Molecular Ecology*, **28**(8), 1994–2012. https://doi.org/10.1111/mec.15009

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. (2010). Using environmental correlations to identify loci underlying local adaptation. *Genetics*, **185**(4), 1411–1423. https://doi.org/10.1534/genetics.110.114819

Devictor, V., van Swaay, C., Brereton, T., Brotons, L., Chamberlain, D., Heliölö, J., Herrando, S., Julliard, R., Kuussaari, M., Lindström, A., Reif, J., Roy, D. B., Schweiger, O., Settele, J., Stefanescu, C., van Strien, A., van Turnhout, C., Vermouzek, Z., Wallis de Vries, M. F., ... Jiguet, F. (2012). Differences in the climatic debts of birds and butterflies at a continental scale. *Nature Climate Change*, **2**, 121–124. https://doi.org/10.1038/NCLIMATE1347

Dickson, L. B., Merkling, S. H., Gautier, M., Ghozlane, A., Jiolle, D., Paupy, C., Ayala, D., Moltini-Conclois, I., Fontaine, A., & Lambrechts, L. (2020). Exome-wide association study reveals largely distinct gene sets underlying specific resistance to dengue virus types 1 and 3 in Aedes aegypti. *PLoS Genetics*, **16**(5), e1008794. https://doi.org/10.1371/journal.pgen.1008794

Durinck, S., Moreau, Y., Kasprzyk, A., Davis, S., de Moor, B., Brazma, A., & Huber, W. (2005). BioMart and bioconductor: A powerful link between biological databases and microarray data analysis. *Bioinformatics*, **21**(16), 3439–3440. https://doi.org/10.1093/bioinformatics/bti525

Durinck, S., Spellman, P. T., Birney, E., & Huber, W. (2009). Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nature Protocols*, **4**(8), 1184–1191. https://doi.org/10.1038/nprot.2009.97

Eeva, T., Espín, S., Sánchez-Virosta, P., & Rainio, M. (2020). Weather effects on breeding parameters of two insectivorous passerines in a polluted area. *The Science of the Total Environment*, **729**, 138913. https://doi.org/10.1016/j.scitotenv.2020.138913

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, **9**(5), 1937–1958. https://doi.org/10.5194/gmd-9-1937-2016

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, **37**(12), 4302–4315. https://doi.org/10.1002/joc.5086

Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., Coeur D'Acier, A., Coquelle, C., Delgado, J. -V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., …Gautier, M. (2019). A genomic map of climate adaptation in Mediterranean cattle breeds. *Molecular Ecology*, **28**(5), 1009–1029. https://doi.org/10.1111/mec.15004

Forester, B. R., Lasky, J. R., Wagner, H. H., & Urban, D. L. (2018). Comparing methods for detecting multilocus adaptation with multivariate genotype–environment associations. *Molecular Ecology*, **27**(9), 2215–2233. https://doi.org/10.1111/mec.14584

Frachon, L., Bartoli, C., Carrère, S., Bouchez, O., Chaubet, A., Gautier, M., Roby, D., & Roux, F. (2018). A genomic map of climate adaptation in rabidopsis thaliana at a micro-geographic scale. *Frontiers in Plant Science*, **9**, 967. https://doi.org/10.3389/fpls.2018.00967

Gautier, M. (2015). Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, **201**(4), 1555–1579. https://doi.org/10.1534/genetics.115.181453

Gienapp, P., Calus, M. P. L., Laine, V. N., & Visser, M. E. (2019). Genomic selection on breeding time in a wild bird population. *Evolution Letters*, **3**(2), 142–151. https://doi.org/10.1002/evl3.103

Gosler, A., Clement, P., & Christie, D.A. (2020). Great tit (Parus major), version 1.0. In Birds of the World S.M. Billerman, B. K. Keeney, P. G. Rodewald, T. S.Schulenberg (Eds.), Birds of the World. Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.gretit1.01.

Grant, B. R., & Grant, P. R. (1993). Evolution of Darwin's finches caused by a rare climatic event. *Proceedings of the Royal Society B: Biological Sciences*, **251**(1331), 111–117. https://doi.org/10.1098/rspb.1993.0016

Günther, T., & Coop, G. (2013). Robust identification of local adaptation from allele frequencies. *Genetics*, **195**(1), 205–220. https://doi.org/10.1534/genetics.113.152462

Hickling, R., Roy, D. B., Hill, J. K., Fox, R., & Thomas, C. D. (2006). The distributions of a wide range of taxonomic groups are expanding polewards. *Global Change Biology*, **12**(3), 450–455. https://doi.org/10.1111/j.1365-2486.2006.01116.x

Hinks, A. E., Cole, E. F., Daniels, K. J., Wilkin, T. A., Nakagawa, S., & Sheldon, B. C. (2015). Scale-dependent phenological synchrony between songbirds and their caterpillar food source. *American Naturalist*, **186**(1), 84–97. https://doi.org/10.1086/681572

IPCC (2022). Summary for policymakers [H. -O. Pörtner, D. C. Roberts, E. S. Poloczanska, K. Mintenbeck, M. Tignor, A. Alegría et al. (Eds.)]. In Climate change 2022: Impacts, adaptation and vulnerability (pp. 3–33). Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on climate change [H. -O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegría et al. (Eds.)]. Cambridge University Press, pp. 1–34. https://doi.org/10.1017/9781009325844.001

Jeffreys, H. 1961. Theory of probability (3rd ed.). Clarendon Press.

Keller, L. F., & Noordwijk, A. J. V. (1994). Effects of local environmental conditions on nestling growth. *Ardea*, **82**(2), 349–362. https://doi.org/10.5167/uzh-53659

Khaliq, I., Hof, C., Prinzinger, R., Böhning-Gaese, K., & Pfenninger, M. (2014). Global variation in thermal tolerances and vulnerability of endotherms to climate change. *Proceedings of the Royal Society B: Biological Sciences*, **281**(1789), 20141097. https://doi.org/10.1098/rspb.2014.1097

Kim, J. -M., Santure, A. W., Barton, H. J., Quinn, J. L., Cole, E. F., Great Tit HapMap Consortium, Visser, M. E., Sheldon, B C., Groenen, M. A. M., van Oers, K., & Slate, J. (2018). A high-density SNP chip for genotyping great tit (Parus major) populations and its application to studying the genetic architecture of exploration behaviour. *Molecular Ecology Resources*, **18**(4), 877–891. https://doi.org/10.1111/1755-0998.12778

Kvist, L., Ruokonen, M., Lumme, J., & Orell, M. (1999). The colonization history and present-day population structure of the European great tit (Parus major major). *Heredity*, **82**(Pt 5), 495–502. https://doi.org/10.1038/sj.hdy.6885130

Lack, D. (1964). A long-term study of the great tit (Parus major). *Journal of Animal Ecology*, **33**, 159–173. https://doi.org/10.2307/2437

Laine, V. N., Gossmann, T. I., Schachtschneider, K. M., Garroway, C. J., Madsen, O., Verhoeven, K. J. F., de Jager, V., Megens, H. -J., Warren, W. C., Minx, P., Crooijmans, R. P. M. A., Corcoran, P., Great Tit HapMap Consortium, Sheldon, B C., Slate, J., Zeng, K., van Oers, K., Visser, M. E., & Groenen, M. A. M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nature Communications*, **7**, 10474. https://doi.org/10.1038/ncomms10474

Lemoine, M., Lucek, K., Perrier, C., Saladin, V., Adriaensen, F., Barba, E., Belda, E. J., Charmantier, A., Cichoń, M., Eeva, T., Grégoire, A., Hinde, C. A., Johnsen, A., Komdeur, J., Mänd, R., Matthysen, E., Norte, A. C., Pitala, N., Sheldon, B. C., ...Richner, H. (2016). Low but contrasting neutral genetic differentiation shaped by winter temperature in European great tits. *Biological Journal of the Linnean Society*, **118**(3), 668–685. https://doi.org/10.1111/bij.12745

Lê, S., Josse, J., & Husson, F. (2008). FactoMineR: An R package for multivariate analysis. *Journal of Statistical Software*, **25**(1), 1–18. https://doi.org/10.18637/jss.v025.i01

Liu, Z., Liu, Y., Xing, T., Li, J., Zhang, L., Jiang, Y., & Gao, F. (2022). Transcriptome analysis reveals the mechanism of chronic heat stress on meat quality of broilers. *Journal of Animal Science and Biotechnology*, **13**, 110. https://doi.org/10.1186/s40104-022-00759-3

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. https://doi.org/10.1038/nrg1226

Mazzarella, A. B., Boessenkool, S., Østbye, K., Vøllestad, L. A., & Trucchi, E. (2016). Genomic signatures of the plateless phenotype in the threespine stickleback. *Ecology and Evolution*, **6**(10), 3161–3173. https://doi.org/10.1002/ece3.2072

Neph, S., Kuehn, M. S., Reynolds, A. P., Haugen, E., Thurman, R. E., Johnson, A. K., Rynes, E., Maurano, M. T., Vierstra, J., Thomas, S., Sandstrom, R., Humbert, R., & Stamatoyannopoulos, J. A. (2012). BEDOPS: High-performance genomic feature operations. *Bioinformatics*, **28**(14), 1919–1920. https://doi.org/10.1093/bioinformatics/bts277

Nosil, P., Parchman, T. L., Feder, J. L., & Gompert, Z. (2012). Do highly divergent loci reside in genomic regions affecting reproductive isolation? A test using next-generation sequence data in Timema stick insects. *BMC Evolutionary Biology*, **12**, 164. https://doi.org/10.1186/1471-2148-12-164

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology, Evolution, and Systematics*, **37**(1), 637–669. https://doi.org/10.1146/annurev.ecolsys.37.091305.110100

Parmesan, C., Ryrholm, N., Stefanescu, C., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W. J., Thomas, J. A., & Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature*, **399**, 579–583. https://doi.org/10.1038/21181

Parmesan, C., & Yohe, G. (2003). A globally coherent fingerprint of climate change. *Nature*, **421**, 37–42. https://doi.org/10.1038/nature01286

Perrier, C., Rougemont, Q., & Charmantier, A. (2020). Demographic history and genomics of local adaptation in blue tit populations. *Evolutionary Applications*, **13**(6), 1145–1165. https://doi.org/10.1111/eva.13035

Perrins, C. M. (1970). The timing of birds’ breeding seasons. *Ibis*, **112**(2), 242–255. https://doi.org/10.1111/j.1474-919X.1970.tb00096.x

Phillimore, A. B., Leech, D. I., Pearce-Higgins, J. W., & Hadfield, J. D. (2016). Passerines may be sufficiently plastic to track temperature-mediated shifts in optimum lay date. *Global Change Biology*, **22**(10), 3259–3272. https://doi.org/10.1111/gcb.13302

Pritchett, E. M., Van Goor, A., Schneider, B. K., Young, M., Lamont, S. J., & Schmidt, C. J. (2023). Chicken pituitary transcriptomic responses to acute heat stress. *Molecular Biology Reports*, **50**(6), 5233–5246. https://doi.org/10.1007/s11033-023-08464-8

Qu, Y., Zhao, H., Han, N., Zhou, G., Song, G., Gao, B., Tian, S., Zhang, J., Zhang, R., Meng, X., Zhang, Y., Zhang, Y., Zhu, X., Wang, W., Lambert, D., Ericson, P. G. P., Subramanian, S., Yeung, C., Zhu, H., … Lei, F. (2013). Ground tit genome reveals avian adaptation to living at high

altitudes in the Tibetan plateau. *Nature Communications*, **4**, 2071. https://doi.org/10.1038/ncomms3071

Quinlan, A. R. (2014). BEDTools: The Swiss-Army Tool for genome feature analysis. *Current Protocols in Bioinformatics*, **47**, 11.12.1–34. https://doi.org/10.1002/0471250953.bi1112s47

Ramakers, J. J. C., Gienapp, P., & Visser, M. E. (2018). Phenological mismatch drives selection on elevation, but not on slope, of breeding time plasticity in a wild songbird. *Evolution*, **73**(2), 175–187. https://doi.org/10.1111/evo.13660

Reed, T. E., Grøtan, V., Jenouvrier, S., Sæther, B. E., & Visser, M. E. (2013). Population growth in a wild bird is buffered against phenological mismatch. *Science*, **340**(6131), 488–491. https://doi.org/10.1126/science.1232870

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. (2015). A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, **24**(17), 4348–4370. https://doi.org/10.1111/mec.13322

Rodriguez, S., & Barba, E. (2016). Nestling growth is impaired by heat stress: An experimental study in a Mediterranean great tit population. *Zoological Studies*, **55**, e40. https://doi.org/10.6620/ZS.2016.55-40

Root, T. L., Price, J. T., Hall, K. R., Schneider, S. H., Rosenzweig, C., & Pounds, J. A. (2003). Fingerprints of global warming on wild animals and plants. *Nature*, **421**, 57–60. https://doi.org/10.1038/nature01333

Ryan, T. A., Taff, C. C., Zimmer, C., & Vitousek, M. N. (2023). Cold temperatures induce priming of the glucose stress response in tree swallows. *Comparative Biochemistry and Physiology Part A: Molecular and Integrative Physiology*, **280**, 111419. https://doi.org/10.1016/j.cbpa.2023.111419

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N. J., Dominoni, D. M., Helm, B., Lundberg, M., Senar, J. C., Sprau, P., Visser, M. E., & Isaksson, C. (2021). Repeated genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nature Communications*, **12**, 2983. https://doi.org/10.1038/s41467-021-23027-w

Santure, A. W., & Garant, D. (2018). Wild GWAS—Association mapping in natural populations. *Molecular Ecology Resources*, **18**(4), 729–738. https://doi.org/10.1111/1755-0998.12901

Sauve, D., Friesen, V. L., & Charmantier, A. (2021). The effects of weather on avian growth and implications for adaptation to climate change. *Frontiers in Ecology and Evolution*, **9**, 569741. https://doi.org/10.3389/fevo.2021.569741

Skarman, A., Jiang, L., Hornshøj, H., Buitenhuis, B., Hedegaard, J., Conley, L. N., & Sørensen, P. (2009). Gene set analysis methods applied to chicken microarray expression data. *BMC Proceedings*, **3**(Suppl 4), S8. https://doi.org/10.1186/1753-6561-3-S4-S8

Spurgin, L. G., Bosse, M., Adriaensen, F., Albayrak, T., Barboutis, C., Belda, E., Bushuev, A., Cecere, J. G., Charmantier, A., Cichon, M., Dingemanse, N. J., Doligez, B., Eeva, T., Erikstad, K. E., Fedorov, V., Griggio, M., Heylen, D., Hille, S., Hinde, C. A., … Slate, J. (2024). The great tit HapMap project: A continental-scale analysis of genomic variation in a songbird. *Mol. Ecol. Resour.*, **24**:e13969. https://doi.org/10.1111/1755-0998.13969

Stonehouse, J. C., Slate, J., Spurgin, L. G., Laine, V. N., Bosse, M., Groenen, M., van Oers, K., Sheldon, B. C., Visser, M. E. (2023). The genomics of adaptation to climate in European great tit (Parus major) populations [Dataset]. *Dryad*. https://doi.org/10.5061/dryad.vt4b8gtxd

Sumasgutner, P., Cunningham, S. J., Hegemann, A., Amar, A., Watson, H., Nilsson, J. F., Andersson, M. N., & Isaksson, C. (2023). Interactive effects of rising temperatures and urbanisation on birds across different climate zones: A mechanistic perspective. *Global Change Biology*, **29**(9), 2399–2420. https://doi.org/10.1111/gcb.16645

Sunday, J. M., Bates, A. E., & Dulvy, N. K. (2012). Thermal tolerance and the global redistribution of animals. *Nature Climate Change*, **2**, 686–690. https://doi.org/10.1038/nclimate1539

The Gene Ontology Consortium (2021). The Gene Ontology resource: Enriching a gold mine. *Nucleic Acids Research*, **49**(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Tian, S., Zhou, X., Phuntsok, T., Zhao, N., Zhang, D., Ning, C., Li, D., & Zhao, H. (2020). Genomic analyses reveal genetic adaptations to tropical climates in chickens. *iScience*, **23**(11), 101644. https://doi.org/10.1016/j.isci.2020.101644

Toledo, A., Andersson, M. N., Wang, H. L., Salmón, P., Watson, H., Burdge, G. C., & Isaksson, C. (2016). Fatty acid profiles of great tit (Parus major) eggs differ between urban and rural habitats, but not between coniferous and deciduous forests. *The Science of Nature*, **103**(7), 55. https://doi.org/10.1007/s00114-016-1381-0

van Balen, J. H. (1973). A comparative study of the breeding ecology of the great tit Parus major in different habitats. *Ardea*, **55**(1–2), 1–93. https://doi.org/10.5253/arde.v61.p1

van Bers, N. E., Santure, A. W., Van Oers, K., De Cauwer, I., Dibbits, B. W., Mateman, C., Crooijmans, R. P. M. A., Sheldon, B. C., Visser, M. E., Groenen, M. A. M., & Slate, J. (2012). The design and cross-population application of a genome-wide SNP chip for the great tit Parus major. *Molecular Ecology Resources*, **12**(4), 753–770. https://doi.org/10.1111/j.1755-0998.2012.03141.x

van Oers, K., Santure, A. W., De Cauwer, I., van Bers, N. E. M., Crooijmans, R. P. M. A., Sheldon, B. C., Visser, M. E., Slate, J., & Groenen, M. A. M. (2014). Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. *Heredity*, **112**(3), 307–316. https://doi.org/10.1038/hdy.2013.107

Vedder, O., Bouwhuis, S., & Sheldon, B. C. (2013). Quantitative assessment of the importance of phenotypic plasticity in adaptation to climate change in wild bird populations. *PLoS Biology*, **11**(7), e1001605. https://doi.org/10.1371/journal.pbio.1001605

Visser, M. E., van Noordwijk, A. J., Tinbergen, J. M., & Lessells, C. M. (1998). Warmer springs lead to mistimed reproduction in great tits (Parus major). *Proceeding of the Royal Society B: Biological Sciences*, **265**(1408), 1867–1870. http://doi.org/10.1098/rspb.1998.0514

Visser, M. E., Lindner, M., Gienapp, P., Long, M. C., & Jenouvrier, S. (2021). Recent natural variability in global warming weakened phenological mismatch and selection on seasonal timing in great tits (Parus major). *Proceedings of the Royal Society B: Biological Sciences*, **288**(1963), 20211337. https://doi.org/10.1098/rspb.2021.1337

Yang, J., Li, W. R., Lv, F. H., He, S. G., Tian, S. L., Peng, W. F., Sun, Y. W., Zhao, Y. X., Tu, X. L., Zhang, M., Xie, X. L., Wang, Y. T., Li, J. Q., Liu, Y. G., Shen, Z. Q., Wang, F., Liu, G. J., Lu, H. F., Kantanen, J., …Liu, M. J. (2016). Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Molecular Biology and Evolution*, **33**(10), 2576–2592. https://doi.org/10.1093/molbev/msw129

2.11. Supplementary Material for Chapter 2 digital file

2.11.1. Appendix 2 (supplementaryChapter2.pdf)

Including Tables S2.1-S2.20 and Figures S2.1-S2.10.

Appendix 2 data files:

2.11.2. GOtoGene_PC1.txt

2.11.3. GOtoGene _PC2.txt

2.11.4. GOtoGene _PC3.txt

2.11.5. GOtoGene _PC4.txt

3. Haplotype-based tests for selective sweeps within and between populations

3.1. Abstract

Inter-population variation in the way positive natural selection and purifying selection acts on the genome will lead to local adaptation and genetic differentiation. We previously detected cross-population signatures of recent positive selection using a genome-environment analysis of twenty European great tit populations and identified candidate SNPs associated with climate (GEA method details see **Chapter 2**). However, the populations where selection had been acting were unknown. To address this gap, extended haplotype homozygosity (EHH) tests were performed to identify the populations that have experienced selective sweeps associated with climate. Seventeen candidate SNPs were identified as being under selection, and their closest genes potentially play a role in local adaptation to climate. Closer investigation of the candidate SNPs revealed that they have generally been under selection in just one population or in geographically close populations, rather than being under parallel evolution in different populations. It is important to consider that identifying signatures of recent positive selection by only one or a few approaches (in this case, by extended haplotype-based testing and the association with climate-PCs), requires real caution when inferring causation of local adaptation in a polygenic trait, due to the risk of false-positive results.

3.2. Introduction

3.2.1. Spatial Patterns of Selection and Local Adaptation

Spatial variation in the patterns of natural selection, genetic drift and demography can lead to genetic differentiation among populations (Vucetich and Waite 2003). Positive selection drives the process of local adaptation, where favourable alleles increase in frequency more than expected under neutrality and alter the allele frequencies of genomic variants near the beneficial mutation (Oleksyk et al. 2010). The process of local adaptation by positive selection leaves signatures of selective sweeps on the underlying genomic regions, characterised by reduced diversity and extensive linkage disequilibrium (LD) (Kim and Nielsen 2004). However, see Pritchard and di Rienzo (2010) for a comment that selective sweeps are less easy to detect when polygenic adaptation has taken place i.e. the adaptation involves positive selection acting at many genes with relatively small effects on any evolutionary change.

A haplotype is a set of adjacent polymorphisms on a single chromosome that tend to be inherited as a unit (Crawford and Nickerson 2005). Haplotype-based methods can compare haplotype patterns across populations, based on the association between causal mutations and the ancestral haplotypes on which they arose (Oleksyk et al. 2010). These methods are often used for genetic association studies of human disease, comparing the haplotypes of disease patients and healthy controls (Gabriel et al. 2002). The genetic basis of local adaptation can be elucidated by genome scans for selective sweeps through haplotype homozygosity (Oleksyk et al. 2010; Lange and Pool 2016). The power to detect signatures of selective sweeps via different statistical tests may depend on the effective population size, generation time since selection, the strength of the selective sweep, and the amount of recombination (Burke 2012). For a review into ecological genomics of local adaptation see Savolainen et al. (2013). Recent examples of selective sweeps being detected in domesticated livestock and crops include: in indigenous sheep breeds from the Middle East and South Asia (Eydivandi et al. 2021), during adaptation to harsh environments in West African Dwarf sheep (Álvarez et al. 2020), in indicine cattle (Chen et

al. 2023), in response to selection on agronomic traits and crop yield in rapeseed (Hu et al. 2022), and in domesticated macadamia nuts (Lin et al. 2022). Selective sweeps of *Plasmodium falciparum* mutations that caused drug resistance have been observed to cause malaria outbreaks (Wasakul et al. 2023).

In the context of environmental adaptation, genome scans for selective sweeps have typically focused on altitude, temperature or arid adaptation. For example, high-altitude adaptations were identified in humans (Peng et al. 2011) and ruminants (Friedrich and Wiener 2020). In flowering plants adaptation to the highest altitudes may be linked to adaptation to cold stress and photoperiod (Wei et al. 2023). Selective sweeps associated with adaptation to climate include evidence of adaptation in sheep to hot arid environments (Kim et al. 2016), and in coral to the warmest reefs (Smith et al. 2022). For a review of how to characterise and understand adaptive genetic variation that enables species to respond to climate change see Bernatchez et al. (2023).

Candidate adaptive loci identified from tests of selective sweeps can be investigated for enrichment of gene functions (via GO Term testing) e.g. as used in understanding genomic adaptations to domestication in the macadamia nut (Lin et al. 2022). In contrast to other approaches to detect signatures of selection acting across the genome, the identified candidate adaptive loci from selective sweep approaches can be compared directly to candidate adaptive genes identified from genome-environment association (GEA) studies, to investigate any similarity of gene functions. For example, De La Torre et al. (2021) showed recent/ongoing climate adaptation along temperature and moisture gradients of giant sequoia and coast redwood trees, with some regions characterised by strong selective sweeps and other loci showing weaker evidence of sweeps, perhaps because of many genes being involved in polygenic adaptation.

3.2.2. Parallel adaptive evolution in geographically separated populations

Adaptive evolution can be expected to occur in parallel across different geographically separated populations of a species when exposed to similar selection pressures, because of the shared ancestral patterns of genome-wide diversity (i.e. standing genetic variation) that selection is able to act upon (Barrett and Schluter 2008). In each spatially separated setting, adaptive alleles can be positively selected from the standing genetic variation or alternatively may arise from new mutations, without the adaptive alleles being introduced by gene flow (Olson-Manning et al. 2012). A classic example of an adaptive radiation is found in Darwin's finches, with the *ALX1* haplotype contributing to the diversification of beak sizes and shapes for adaptation to diets of different seed sizes, with the same beak changes observed repeatedly in populations with shared diets (Lamichhaney et al. 2015).

Phenotypic evolution occurring in parallel populations has also been demonstrated for genomic adaptation to altitude in two tropical butterflies found on mountain habitats either side of the Andes, but not in five close relatives (Montejo-Kovacevich et al. 2022). As well as parallel selection occurring simultaneously in spatially separated populations, adaptive evolution can also occur in a repeated fashion, whereby adaptive evolution can occur first in one population and then the process can later be repeated in another separate population, with selection acting on the shared ancestral standing genetic variation. An example of repeated adaptive evolution is found in the colonisation and adaptation of marine sticklebacks to thousands of freshwater streams and lakes over their evolutionary history, with a well-defined set of genome-wide loci that are consistently associated with marine-freshwater divergence (Jones et al. 2012).

Shared adaptive pressures acting on the availability of standing genetic variation often results in parallel evolution at the phenotypic level, but not on the genetic level in polygenic traits. A comparison of genome-wide patterns of differentiation between high- and low-elevation taxon of the Qinghai–Tibet Plateau in all 19 East Asian true tit species (*Paridae*), found nonparallel genetic changes. However, even when different genes are involved in adaptation to high or low elevation, there may still be common genetic pathways involved in oxygen transport at high-elevation and thermogenesis (Cheng et al. 2021).

3.2.3. Detecting selection

Methods to detect genetic diversity and population structure include PCA admixture, phylogenetic analyses, and length of runs of homozygosity (ROH). See Ceballus et al. (2018) for a review of using homozygosity burden from ROH detection in the human genome to investigate how recessive variants contribute to the genetic architecture of stature and cognition. Tests for signatures of selection can be performed on sequenced genomic DNA to identify regions of recent strong positive selection (Oleksyk et al. 2010). Population differentiation methods e.g. fixation index Fst (introduced by Wright 1949; Malécot 1948) and FLK (Bonhomme et al. 2010), can detect differential selection patterns acting in different populations. Haplotype-based statistics can use methods to detect selective sweeps within populations or between populations to reveal signatures of local adaptation, often with methods that test patterns in datasets relative to expectations under a null hypothesis of neutral evolution (Abondio et al. 2022). The challenge for further investigation of detected candidate regions/genes under selection is to discern their biological function, and this functional analysis may assist in understanding the processes involved in adaptation (Hoban et al. 2016).

3.2.4. Haplotype-based tests for selective sweeps

EHH methods are focused on site-specific polymorphisms and consider associations between linked SNPs to infer whether sweeps have taken place (Klassmann and Gautier 2022). The EHH statistic first developed by Sabeti et al. (2002) was demonstrated to detect selective sweeps around the *G6PD*-CH8 and *TNFSF5*-CH4 core haplotypes in genes associated with malaria resistance in humans (**Figure 3.1**). EHH is a linkage disequilibrium (LD)-based measure of EHH surrounding a focal site-specific SNP, used to detect ongoing selective sweeps. Homozygosity in “EHH” refers to the probability that two randomly chosen haplotypes around a focal site-specific SNP from a population are identical. The premise of a test for a selective sweep is that long haplotypes are generated during a selective sweep (where a variant rapidly rises to high frequency), because local LD is not disrupted by recombination. Eventually though, recombination will erode the LD signal.

Thus, a sweep carries a long haplotype to fixation in a population, with observed reduced genetic diversity along the haplotype.

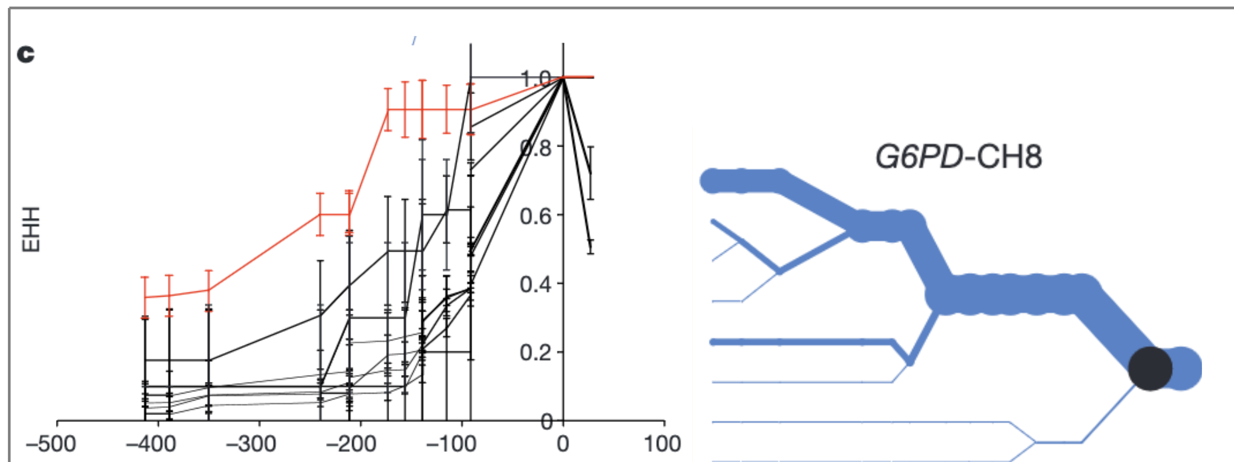


Figure 3.1: A selective sweep was detected by long-range homozygosity around the *G6PD* gene -CH8 core haplotype, associated with malaria resistance in humans. Panel A) The EHH at varying distances from the core region on each core haplotype at *G6PD* demonstrates that G6PD-CH8 (labelled red) has persistent, high EHH values. Panel B) A haplotype bifurcation diagram of G6PD-CH8 shows it has long-range homozygosity that is unusual given its frequency (Figures from Sabeti et al. 2002).

EHH tests detect the signature of relatively recent episodes of selection where the selected allele is fixed or nearly fixed. They have weak power to detect ancient sweeps because recombination and new mutations will erode long haplotypes (Chen et al. 2010). Phased data perform better than unphased data for detecting selective sweeps using EHH tests (Klassmann and Gautier 2022). Genomic data can be phased with packages such as Beagle v5.4 (Browning et al. 2021), ShapeIt2 (O'Connell et al. 2014) and fastPhase (Scheet and Stephens 2006).

3.2.5. Integrated haplotype homozygosity score

The integrated haplotype homozygosity score (iHS) (Voight et al. 2006) is used to detect partial selective sweeps in a single population. The iHS statistic was modified from the

allele specific EHH statistic (Sabeti et al. 2002). The iHS is a measure of the decay of homozygosity with increasing distance from the focal marker (primary locus) and is calculated independently in each direction (upstream/downstream) from the marker. Any two (or more) chromosomes identical in that region constitute a shared haplotype. To detect selection using iHS, both alleles of a site must be present in enough sequences for obtaining a reliable estimate of their respective allele specific EHH. Typically, a minor allele frequency (MAF) of at least 5% is required, which excludes variants near fixation. iHS can be calculated and standardised by the mean across all considered markers with software packages such as Selscan (Szpiech and Hernandez 2014), hapbin (Maclean et al. 2015) or REHH 2.0 (Gautier et al. 2017).

The iHS statistic, based on comparing two alleles, has been independently modified to compare two populations, using two methods that calculate the integrated haplotype homozygosity in subtly different ways: 1) cross-population XP-EHH (see Sabeti et al. 2007 for its derivation), and 2) the ratio between populations (Rsb) (see Tang et al 2007 for its derivation). Candidate regions identified under selection using XP-EHH and Rsb values with phased data are usually highly correlated, and essentially identical with unphased data (Klassman and Gautier 2022), with neither method evaluated to be better at detecting selection than the other (Gautier 2017). In some cases slightly different results are produced; e.g. Ceccobelli et al. (2023) identified 4 and 8 candidate regions respectively with XP-EHH and Rsb when comparing Mediterranean vs Continental Merino sheep. Crucially both XP-EHH and Rsb methods can be used to pinpoint which of the two populations selection has acted in and if parallel selection has occurred. This was not possible in Chapter 2, where GEA methods were used to detect loci possibly involved in climate adaptation.

3.2.6. Complementary Haplotype-based statistics

Eydivandi et al. (2021) applied three complementary haplotype-based test statistics (FLK, XP-EHH, and Rsb) with Fst to detect selective sweeps, using pairwise comparisons of three

categories of sheep breeds (from Iran, Afghanistan, and India). They found stronger correlations (as expected) between Fst and FLK values, and between XP-EHH and Rsb values, across the population comparisons. More convincing evidence of selection was found when two or more haplotype-based test statistics detected sweeps at the same gene, like for the *DOCK*, *ZNF*, *ATP* and *COL* family of genes. These candidate genes are involved in the immune response and animal production traits e.g. milk production, body weight and growth (post-weaning). Nine candidate genes were identified in all four tests, with genes *TNIK*, *DOCK1*, *USH2A*, and *TYW1B* associated with disease resistance and climate adaptation (with some breeds adapted to hot arid conditions).

3.2.7. Signatures of selection in the great tit HapMap dataset

A previous investigation of the great tit HapMap project dataset identified signatures of selection using Fst-based methods (Spurgin et al. 2024). The authors found low overall genomic divergence in European great tit populations, with marginally greater genomic differentiation between Mediterranean Island populations. Unique genomic outlier regions were detected at the range edge populations e.g. Scotland, England, Spain and Finland, with candidate loci putatively involved in morphology, thermal adaptation and colouration. However, this analysis does not reveal which agent is responsible for causing selection. For some of these loci we provided evidence that climate was driving selection, using a genome-environment analysis of twenty European great tit populations (for the GEA method details and results see **Chapter 2**). However, neither investigation could reveal the populations in which selection had been acting, and why specific alleles were favoured in these populations?

In this chapter I attempt to address this gap by identifying the populations in which selection has happened. I performed extended haplotype homozygosity (EHH) tests between pairs of populations, by defining one population as a core population and the other as an outlier population, where core and outlier status are determined from each population's Eigenvector value in the climate variable principal component analyses

described in **Chapter 2**. We expected outlier alleles that are under selection to have extended haplotypes relative to non-selected alleles in that population, or to alleles in the core population(s). To produce a list of initial candidate SNPs, I used the Bayes Factor (BF) associations with climate-PCs (see **Chapter 2**). The other measure of differentiation described in Chapter 2, XtX, did not use the climate data and so is sensitive to all agents of selection, not just climate. Furthermore, by considering different outlier populations, I aimed to detect whether the same genes have been under selection in different (geographically separated) populations, or whether selection has favoured different genes in each population?

An ideal case scenario for this analysis would be to detect if alternative alleles had long haplotypes in both low and high outlier populations (relative to the haplotype length in 'core' populations) for a given climate-related principal component (**Figure 3.2**), implying one is adaptive to one set of conditions and one to another set of conditions (**Figure 3.2c**).

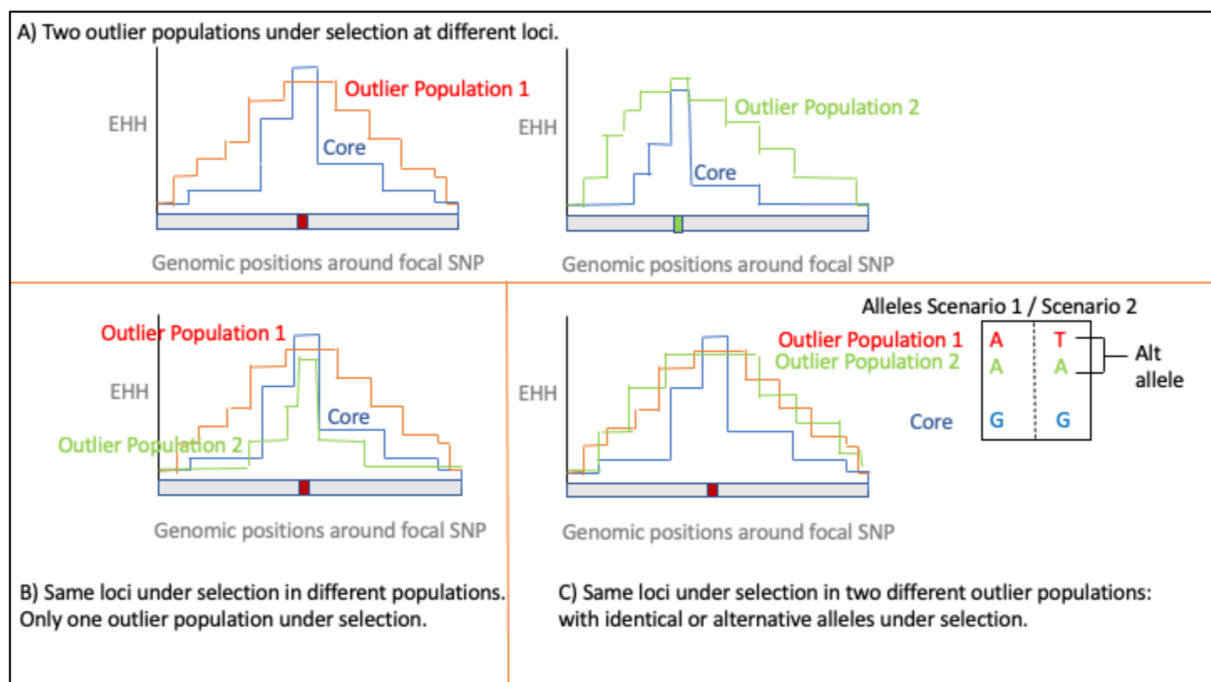


Figure 3.2: The different envisioned scenarios of detecting selection for outlier populations 1 and 2 vs core. A) Different loci are under selection in different outlier populations. B) The

same loci is detected under selection in one outlier population, but not in a different outlier population. C) The same loci under selection in different outlier populations. In the first scenario outlier populations 1 and 2 (which could represent high and low populations from this chapter's data analysis) have identical alleles under selection, while in the second scenario they have alternative alleles under selection.

3.3. Methods

3.3.1. Study system and genomic data

A 500k Affymetrix HD SNP chip (Kim et al. 2018) was developed to genotype blood samples from 29 wild great tit *Parus major* populations, covering much of the species' range. See **Chapter 2** for further details on the assembly and annotation of a reference genome (Laine et al. 2016), the collaboration of the great tit HapMap project and the initial search for signatures of selection (Spurgin et al. 2024). The genotype data set was later phased to construct the most likely haplotypes for each individual, on the following chromosomes: 1-15, 17-24, 26-28, 1A, 4A and Z.

Phasing of SNP genotype data into haplotypes was performed using shapeIT v2.837 (O'Connell et al. 2014). Each chromosome was phased individually, using input files in Plink bed/bim/fam format (Purcell et al. 2007; Chang et al. 2015). Default parameters were chosen, along with the --duoHMM option to use pedigree information to improve phasing accuracy, as many of the samples were in known parent-offspring trios. shapeIT requires a map file which contains the recombination rate between all of the markers. Although the physical map (i.e. genome assembly) position of each SNP was known, far fewer SNPs (~6500) are placed on the great tit linkage map (van Oers et al. 2014). Therefore, to construct the map file we initially estimated the recombination rate (in cM/Mbp) between all SNPs that were on both the linkage map and genome assembly. The remaining SNPs, all of whom had genomic positions, were then placed on the linkage map by linear interpolation. For each interpolated SNP we made the assumption that the recombination rate was constant across the interval between the two nearest flanking SNPs that were on the linkage map. All shapeIT runs were performed on the University of Sheffield High Performance Computing Cluster (ShARC).

In this Chapter I compared haplotypes from outlier vs core populations for climate-principal components (PC) 1-4. See **Chapter 2** for details of the PCA methods that were used to reduce the bioclimate variables of the WorldClim v2 climate data set (Fick and Hijmans

2017) for 20 great tit population locations. Climate-PC1 has strong contributions from temperature variables and climate-PC2 from warm and dry variables. It is important to note that climate-PC1 and PC2 explained ~70% of the variation in the original climate dataset. Climate-PC3 has strong contributions from rain variables and climate-PC4 has strong contributions from wind speed, rain seasonality and daily temp range variables (see **Appendix 2 Fig S2.2** for a visual figure of which climate variables contribute to each PC). The population contribution biplots were inspected (PC1 vs PC2 **Appendix 3 Fig S3.1**; PC3 vs PC4 **Fig S3.2**), and outlier populations were identified along each PC axis at both ends of the axis. Outlier populations are listed in **Table S3.1**. The core populations that remained after identifying outliers were different for each PC. Sometimes outlier populations were not geographically close together - e.g. high value climate-PC1 populations were located in Scotland, France and Corsica. Despite their geographical remoteness and climatic differences, we grouped these populations on the grounds that all three populations occupied similar PC1 space. We expect these outlier populations to be the ones where selection on genes enabling climate adaptation has been happening.

The haplotypes for the samples in high-outlier, low-outlier and core populations were extracted from each phased chromosome file using VCFtools (Danecek et al. 2011) for each PC in turn and indexed with the Parus_major1.1 Genome GFF file (https://www.ncbi.nlm.nih.gov/assembly/GCF_001522545.3) using Bcftools (Danecek et al. 2021) to annotate the SNPs with their nearest genes and genomic features. These three (high-outlier, low-outlier and core) vcf.gz files per chromosome were the input files for the extended haplotype homozygosity (EHH) tests.

3.3.2. Cross-population XP-EHH

Cross-population-EHH, abbreviated to XP-EHH (Sabeti et al. 2007), is a test to detect the selected alleles that have risen to near fixation (i.e. an ongoing selective sweep) in at least one population, but not all populations, by comparing extended homozygosity in population 2 relative to population 1. In XP-EHH methods, the integrated haplotype homozygosity is calculated based on EHHS (iES) values of the average length of shared

haplotypes. The original demonstration of the XP-EHH method identified 26 genes showing evidence of positive selection in the human HapMap dataset (Sabeti et al. 2007).

Common to both XP-EHH and Rsb methods, the ancestral and derived status of alleles at the focal marker are uncharacterised. The XP-EHH and Rsb values can be calculated and standardised to an approximate Gaussian distribution in the REHH package, with a P-value for each SNP assigned. Specifically in the XP-EHH calculation, the standardisation step uses the mean XP-EHH computed over all analysed markers (Klassmann and Gautier 2022; vignette for package REHH 2023). The XP-EHH test is recommended when comparing populations with a low sample size, and grouping low sampled populations with other geographically close or genetically similar populations is suggested to aid detection of cross-population selection signatures (Chen et al. 2016).

3.3.3. Pairwise Rsb

The ratio of EHHS between populations (Rsb) pairwise population statistic (Tang et al. 2007) can be used to detect selection when two populations are used to produce EHHS (site-specific) values. When EHHS values are plotted the allele under selection in one population will have extended 'broad shouldered' haplotype lengths around the focal allele relative to haplotypes not under selection. Rsb is distinguished from XP-EHH methods because the integrated haplotype homozygosity is calculated based on the integrated nEHHS statistic (inES) that represents the average length of shared haplotypes. The unstandardized Rsb (unRsb) value is the log ratio of inES for a marker, and it is then standardised to Rsb, using the median and standard deviation of the unRsb values computed over all analysed markers (and normalised to 1 at the focal position). It has been suggested this standardisation might increase the robustness against different demographic scenarios (Tang et al, 2007; vignette for package REHH 2023).

3.3.4. Extended haplotype homozygosity tests to identify selective sweeps

We compared haplotype lengths between outlier populations (either high-outlier or low-outlier) and core populations, using extended haplotype homozygosity (EHH) statistics, estimated in the R (v4.2) package REHH v3.2.2 (Gautier and Vitalis 2012). Along each chromosome, we used the scan_hh() function, to calculate *iHH*, *inES* and *iES* score matrices for each comparison between an outlier and core population(s). To identify selective sweeps in at least one population, we calculated XP-EHH (Sabeti et al. 2007) with the ies2xpehh() function, which calculates the log-ratio of the *iES* values in populations 1 and 2. We also calculated, Rsb, the ratio of EHHS between populations (Tang et al. 2007) with the ines2rsb() function. This function uses the *inES* statistics from population 1 and 2, to produce Rsb scores. For both the XP-EHH and Rsb statistics, large positive (or negative) results reveal strong extended homozygosity (i.e. a selective sweep) in population one relative to population two (or vice versa). The correlation between the XP-EHH and Rsb statistics was estimated genomewide.

To identify selective sweeps associated with climate, candidate SNPs were identified by two steps. First we looked for evidence of the selective sweeps that have occurred in outlier but not core populations. Here we used the REHH default thresholds of XP-EHH>4 and/or Rsb values >4. A more stringent threshold, e.g. 6, might be suggested when testing a large number of SNPs, to reduce the risk of false positives. However upwardly adjusted thresholds may also risk missing some true positive sweeps. Second, candidate SNPs were further restricted to those that were previously identified as having strong to decisive evidence for an association with climate variation in our previous genome-environment analysis (Chapter 2). To identify candidates, SNPs were filtered by a mean Bayes Factor score >10, calculated across 3 runs of the BayPass Auxilliary model. This resulted in a list of 1815 unique SNPs identified from 286, 505, 327 and 709 SNPs associated with climate-PC1-4, respectively (for method details see **Chapter 2**). Thus, candidates had to have reached statistical significance in both the previous GEA (Chapter 2) and in the haplotype-based tests that identify the population in which selection is acting (this Chapter). Binomial

testing was applied to investigate if positively selected hits occurred more than expected by chance. The null hypothesis is that positively selected SNPs would occur equally often in the outlier and core populations.

For those candidate climate adaptation loci with evidence of selective sweeps in outlier populations from XP-EHH and/or Rsb values, we further investigated whether selection has acted in parallel on the same gene in the different outlier populations. To do this we re-ran the XP-EHH and Rsb tests for distinct outlier populations (each time comparing them to the core populations). From this we could identify whether selection has acted on different loci in each outlier population. If the same sweeps are identified in geographically discrete outlier populations (e.g. in Scotland and in Southern France / Corsica) then we would conclude that parallel evolution has occurred.

A summary of the candidate loci with evidence of selective sweeps associated with climate-PCs 1-4 was compiled from the raw data tables of all the considered population comparisons (**S.Tables S3.2-3.9)**. Data were filtered to retain one best SNP per genomic location. 'Best' SNPs were assigned by counting the number of different test statistics (XP-EHH, Rsb and BF) that provided significant evidence of an association with climate, rather than by prioritising an extreme score in only one test. Each candidate SNP was presented on an extended haplotype homozygosity around a site-specific SNP (EHHS) plot. We also investigated whether there was a bias for significant candidates to be found in regions of low recombination. The local recombination rates (cM/Mbp) for all the candidate SNP were summed, and the same number of hits were randomly assigned to different places of the genome and their local recombination rates were summed. The randomisation was repeated 10,000 times, to test whether sweeps were in more lowly recombining regions than expected by chance. This script was also run on the climate association SNPs (identified in the Chapter 2 GEA) independently for PC1-4, to look for any bias towards the candidate SNPs tending to be in low recombination regions.

3.4. Results

A total of 17 of the previously identified candidate climate adaptation loci showed evidence of selective sweeps in XP-EHH or Rsb tests (**Table 3.1**, **Figures 3.3-3.6**). The XP-EHH and Rsb values generated for each comparison showed a moderate positive correlation, with an average Pearson's product-moment correlation coefficient of 0.64 (**Figures S3.3-4**). For each candidate SNP, a focal EHH plot was drawn showing the nearest gene(s) (**Figures S3.5-21**). The sweeps around the candidate loci were in regions of lower recombination rate than expected by chance ($P = 0.017$), with the 17 candidate SNPs covering 14 distinct regions (**Figure 3.7**). For comparison, the starting set of candidate SNPs excluding those from PC4 (1004 mapped SNPs) were also found in regions of lower recombination than expected ($P = 0.0024$). In contrast, PC4 candidate SNPs tended not to be in regions of lower recombination ($P = 0.81$).

Across PCs 1-3 there were more sweeps detected among candidate climate association loci than expected by chance (**Figures 3.3-3.5**). However, PC4 did not show an excess of selective sweeps over null expectations. We found an excess of these positive significant hits to be in the outlier population (90 SNPs equal to or above the +4 XP-EHH / Rsb thresholds) rather than in the core population (just 20 SNPs equal to or below the -4 thresholds). This excess is statistically significant ($P < 8.2*10^{-12}$ assuming a binomial probability of 0.5). This highly significant *P* value provides support for the GEA results, with evidence that we are reliably detecting sweeps in the outlier populations in the genes that were found in the previous chapter.

Closer investigation of the outlier populations revealed 14 of the candidate regions with selective sweeps were detected in just one population or in geographically close populations, including 11 SNPs undergoing selection only in UK populations. Among the other three SNPs, evidence for a selective sweep was found for SNP AX-100954514 in the Spanish population, SNP AX-100244814 in France and Corsica and SNP AX-100479507 in Finland. Evidence of parallel evolution in geographically remote outlier populations was

only detected for one SNP, AX-100430005, in both Scotland and in the France/Corsica combined populations.

Table 3.1: Seventeen representative candidate climate adaptation loci identified as undergoing sweeps in an outlier population.

SNP	Chr	BP	Core Allele	Outlier Allele	Mean BF	XP-EHH	Rsb	XP-EHH	Rsb	XP-EHH	Rsb	Nearest Gene (kbp)
High PC1 Populations						All outliers		Scotland		France & Corsica		
AX-100958411	1	7,518,315	T	C	**38.3**	3.82	**4.77**	2.04	3.24	3.18	3.41	ATP6AP2 (0.2), MED14 (14.8)
AX-100626511	1A	32,286,754	T	G	17.9	0.36	**4.94**	**4.05**	**5.56**	0.60	1.48	LLPH (12.6), HMGA2 (47.2)
AX-100430005	2	99,275,394	C	T	19.4	**5.61**	**6.71**	**4.26**	2.09	**4.56**	**6.64**	GNAL (11.5)
AX-100212698	3	27,095,579	A	G	16.4	1.92	**4.61**	**7.90**	**6.86**	-1.96	-2.02	LRFN2 (54.6)
AX-100566910	4	45,986,689	T	G	16.8	2.71	**4.95**	**5.78**	**5.52**	0.41	1.54	NSUN7 (0), APBB2 (12.8), RBM47 (47.5)
Low PC1 Populations						All outliers		Finland		Estonia & Russia		
AX-100479507	3	28,881,614	C	T	15.4	**5.58**	**5.59**	**5.01**	**4.87**	2.64	2.81	TTC7A (0), MCFD2 (11.5)
High PC2 Populations						All outliers		Spain		France & Corsica		
AX-100244814	1A	58,759,725	A	G	16.67	2.28	**4.04**	1.50	2.23	2.71	**4.52**	MRPS33 (0), LOC107204196 (14.8), BRAF (18.3)
AX-100954514	Z	26,965,165	T	A	18.41	**4.67**	3.67	**5.45**	2.37	1.52	1.49	JAK2 (0), LOC107215974 (28.9), PLGRKT (38.5)
Low PC2 Populations						All outliers		Scotland		Switzerland & Germany		
AX-100333849	3	25,931,865	G	A	13.69	**4.73**	**10.27**	**6.34**	**6.45**	1.60	3.13	SRBD1 (0)
AX-100053712	3	26,617,245	G	A	16.42	2.12	**4.00**	**6.08**	**4.32**	-0.99	-1.15	CRIPT (16.0), PIGF (23.8), RHOQ (42.8)
AX-100690978	5	21,629,221	T	G	**20.05**	1.06	**4.31**	**5.06**	**4.67**	0.30	-1.35	CD82 (73.3), ALX4 (82.8)
Low PC3 Populations						All outliers		Cambridge UK & Oxford UK		Sweden		
AX-100681217	2	46,317,558	A	C	13.11	**6.00**	**9.04**	**7.54**	**10.36**	0.47	1.39	HERPUD2 (28.8)
AX-100379572	2	12,991,920	T	C	11.25	2.28	**4.79**	3.35	**5.56**	-0.0069	0.58	VPS13B (0)
AX-100271597	2	13,608,283	T	C	**48.95**	**4.27**	**10.15**	**4.27**	**10.15**	-0.96	-0.24	TRPS1 (0)
AX-100386254	3	38,558,688	T	C	**21.1**	**5.29**	**6.17**	**4.76**	**4.97**	1.29	1.90	PDE10A (85.9)
AX-100844877	4A	12,152,707	C	T	13.88	2.97	**5.47**	**4.32**	**6.03**	0.41	1.27	ACSL4 (15.0), COL4A5 (158.4)
Low PC4 Populations						All Outliers		Spain		Switzerland & Belgium		
AX-100523182	15	8,303,211	C	T	**33.05**	2.21	**4.47**	0.96	1.09	-0.62	-0.76	PI4KA (11.6), SERPIND1 (36.9)

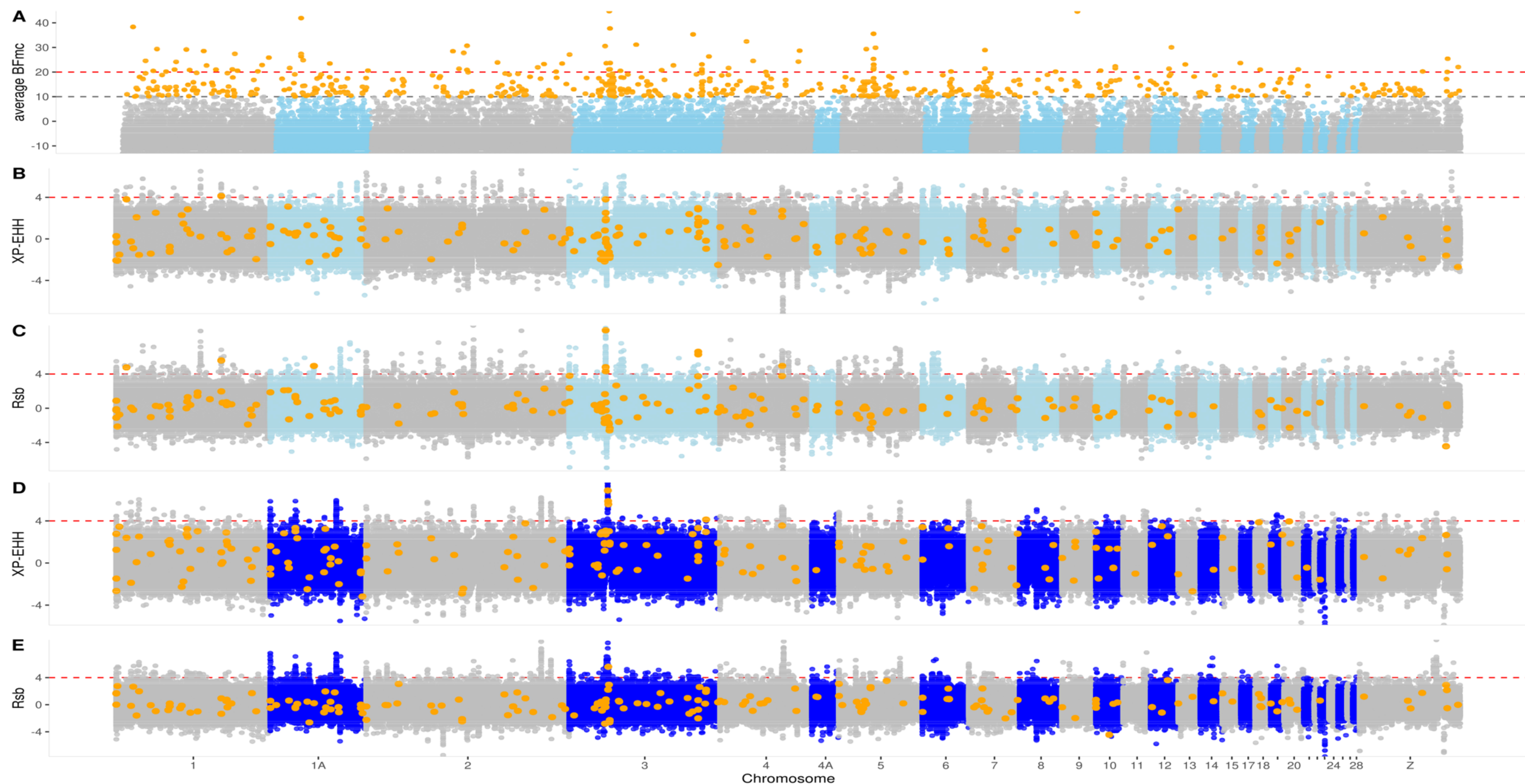


Figure 3.3: EHH statistics for climate-PC1. Panel A highlights 567 climate-PC1 associated SNPs (i.e. BF>10 over 3 runs of BayPass) from a total of 479,111 SNPs. The Manhattan plots display 'high' populations vs core (XP-EHH in panel B and Rsb in panel C) and 'low' vs core (XP-EHH in panel D, and Rsb in panel E). In total, 3, 11, 5 and 3 SNPs are highlighted out of 414, 984, 492, 1120 SNPs above the XP-EHH/Rsb +4 threshold, and with 0, 1, 0, 1 SNPs highlighted out of 125, 216,121 and 246 SNPs below the XP-EHH/Rsb -4 threshold, respectively for panels B-E. For each case using a Binomial test an excess of climate associated SNPs were detected among sweep loci relative to the number expected by chance, except for panel E.

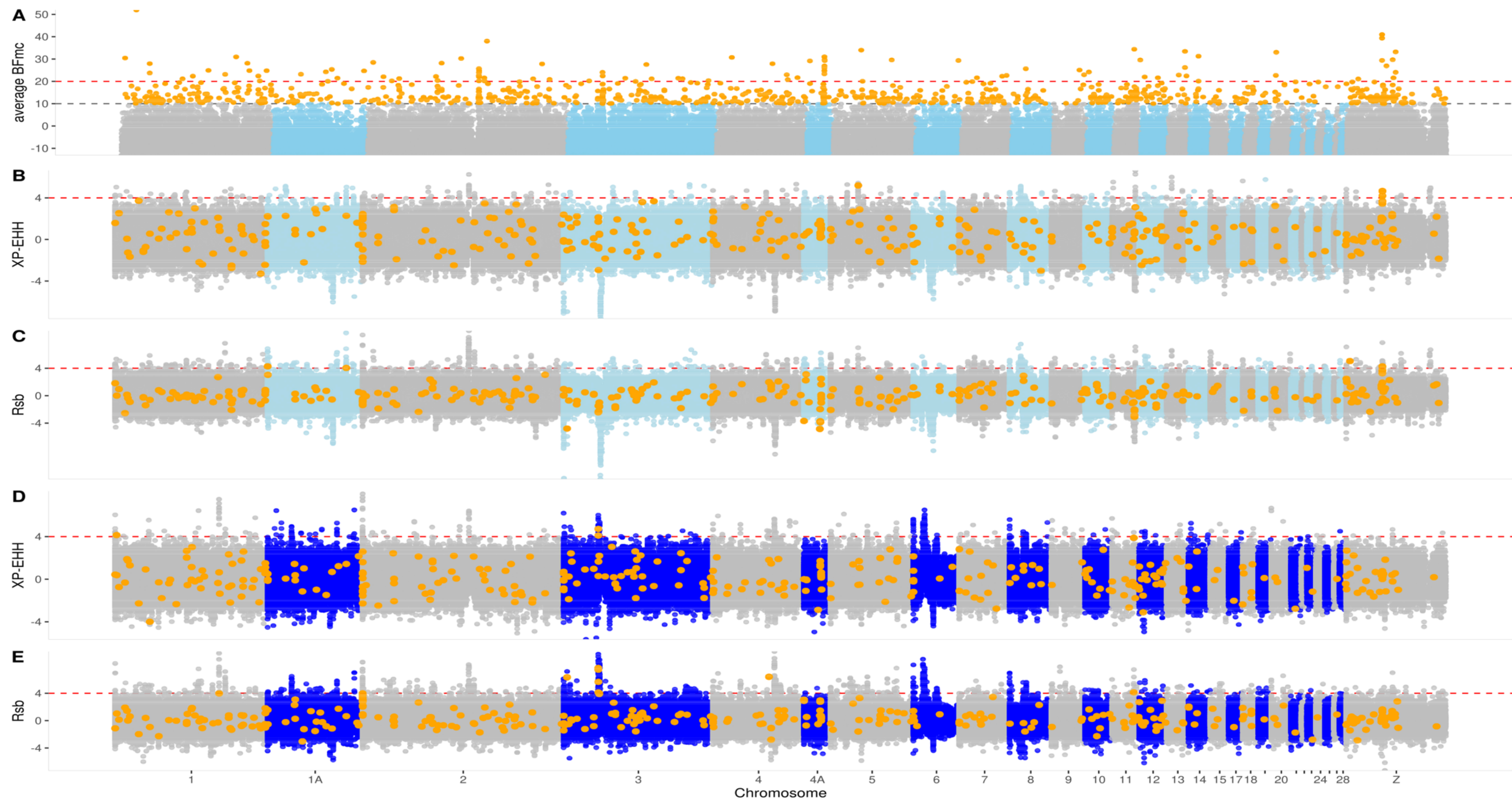


Figure 3.4: EHH statistics for climate-PC2. Plots details as for Figure 3. Panel A highlights 847 climate-PC2 associated SNPs, with 4, 4, 5 and 12 SNPs highlighted out of 308, 755, 485, 1181 SNPs above the XP-EHH/Rsb +4 threshold, and 0, 2, 1, 1 SNPs highlighted out of 329, 646, 93 and 168 SNPs below the XP-EHH/Rsb -4 threshold, respectively for B-E. For each case using a Binomial test an excess of climate associated SNPs were detected among sweep loci relative to the number expected by chance.

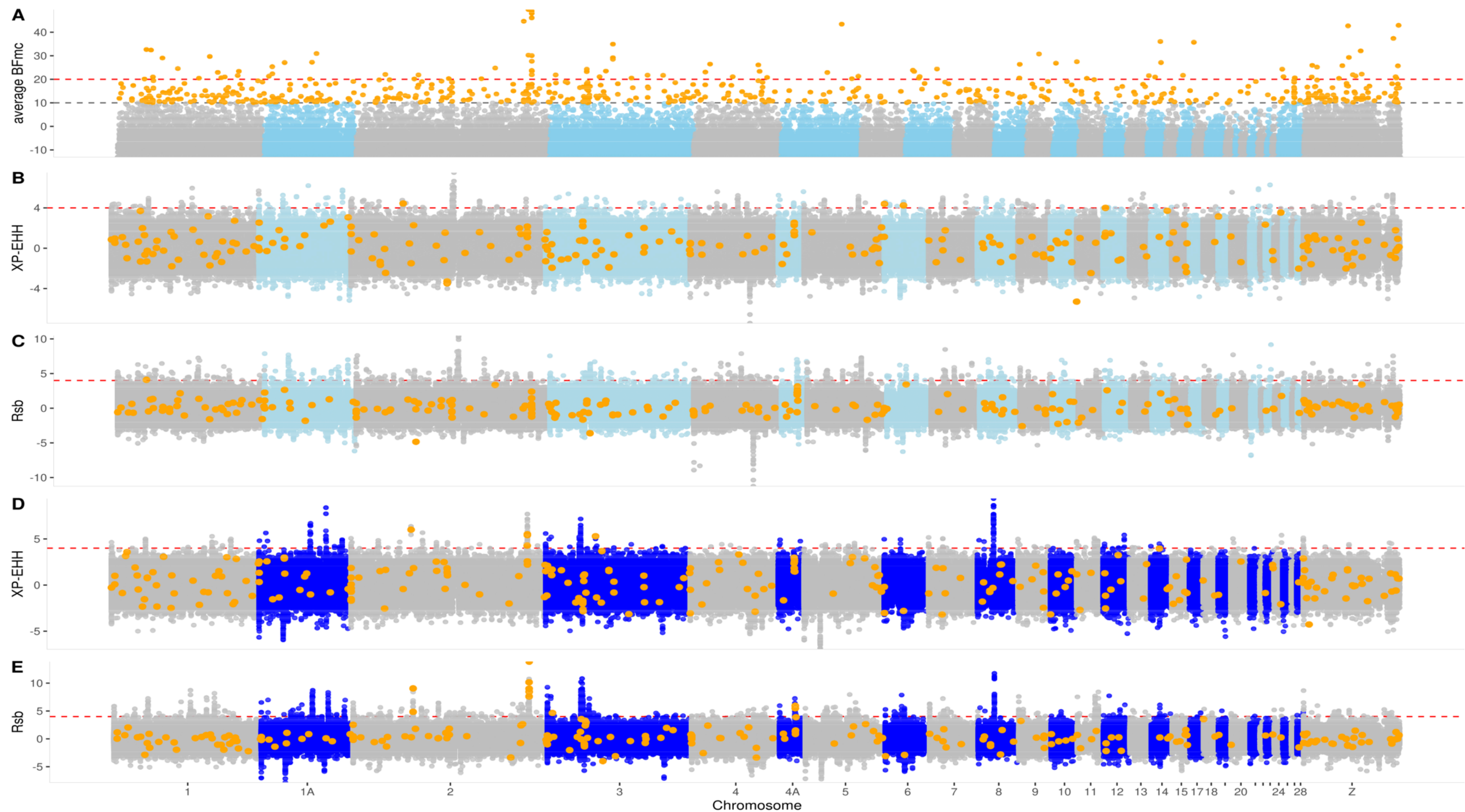


Figure 3.5: EHH statistics for climate-PC3. Plots details as Figure 3. Panel A highlights orange 604 climate-PC3 associated SNPs, with 4, 1, 9 and 18 SNPs highlighted out of 439, 1066, 442, 1213 SNPs above the XP-EHH/Rsb +4 threshold, and 1, 1, 1, 1 highlighted out of 103, 201, 212, 413 below the XP-EHH/Rsb -4 threshold, respectively for B-E. For each case using a Binomial test an excess of climate associated SNPs were detected among sweep loci relative to the number expected by chance, except for panel C.

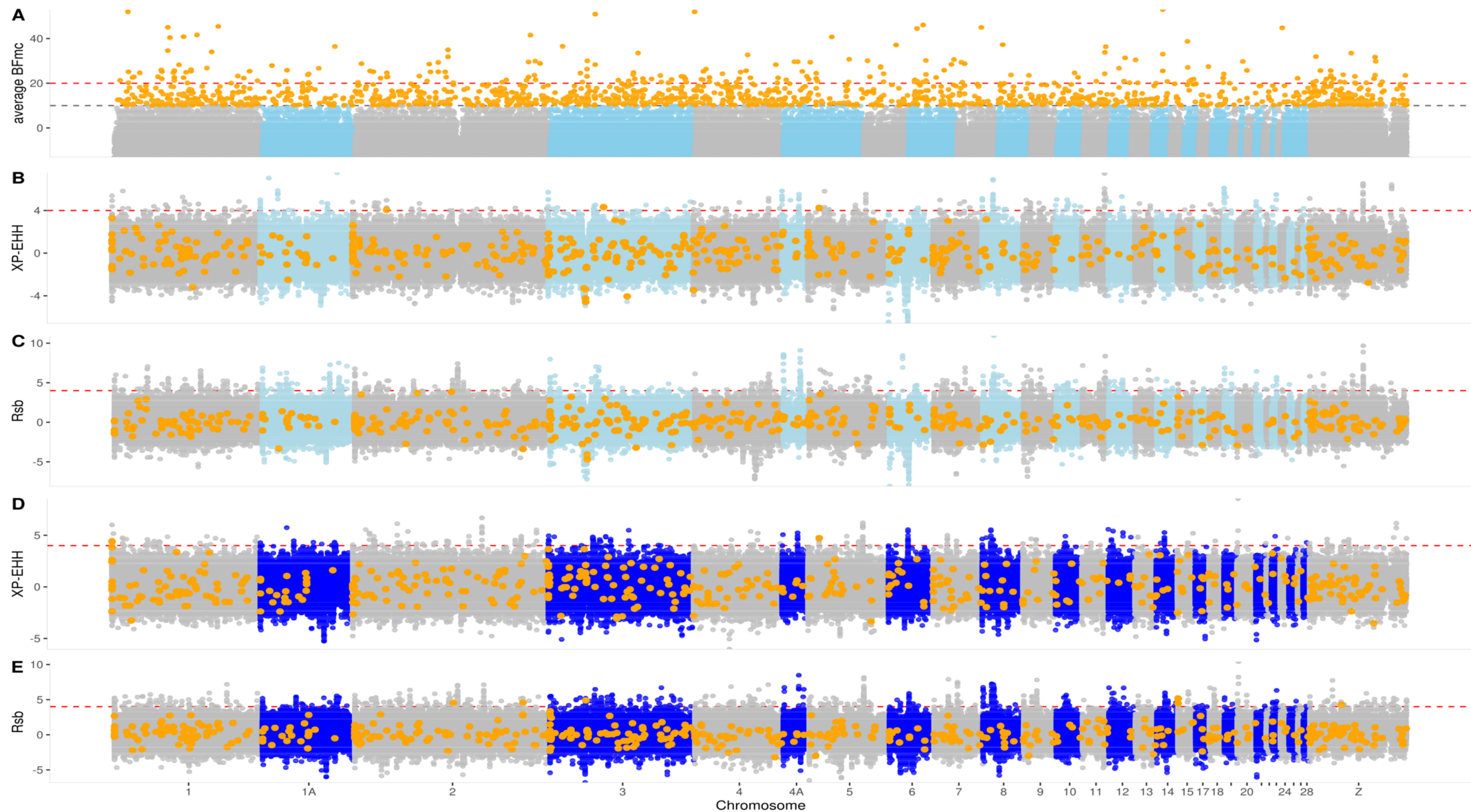


Figure 3.6: EHH statistics for climate-PC4. Plots details as Figure 3. Panel A highlights orange 1343 climate-PC4 associated SNPs, with 3, 0, 2 and 6 SNPs highlighted out of 351, 946, 387, 1017 SNPs above the XP-EHH/Rsb +4 threshold, and 4, 6, 0, 0 SNPs out of 157, 391, 111, 248 SNPs below the XP-EHH/Rsb -4 threshold, respectively for B-E. For each case Binomial testing showed that there was no excess of climate associated SNPs detected among sweep loci relative to the number expected by chance.

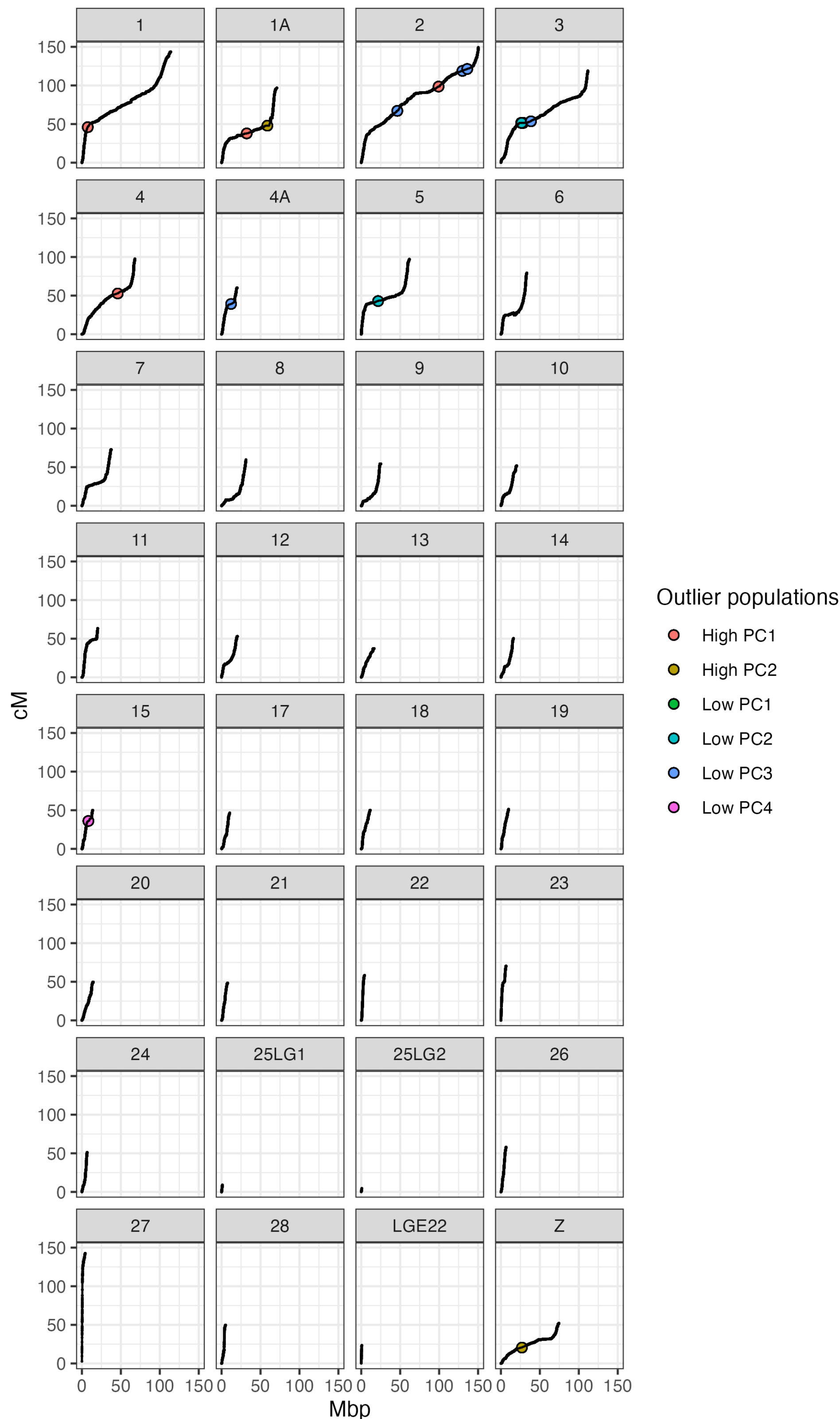


Figure 3.7: The 14 distinct regions of the 17 candidate climate adaptation SNPs shown on the recombination rate map (cM/Mbp) for great tit autosomes using data shared by Prof Jon Slate (personal communication 2022). Flatter slopes indicate regions of low recombination.

3.5. Discussion

Two complementary EHH statistics were used to identify the populations that have experienced selective sweeps associated with climate, informed by the 1815 unique candidate SNPs associated with climate for PC1-4 from the Chapter 2 GEA analysis. Evidence of selective sweeps was found for seventeen of these candidate climate adaptation SNPs. Among the 17 significant candidate SNPs we observed more haplotypes with evidence of a sweep in the outlier populations than in the core populations (confirming what we expected to see). The closest genes to these candidate loci may potentially play a role in local adaptation to climate, and each will be discussed in context for biological process and its relevance to adaptation in the studied great tit populations.

3.5.1. Genes within sweep regions and their effects in other species

Of the 17 candidate loci showing evidence of selective sweeps, 11 were in UK populations (see **Table 3.1** for details). Below I discuss the possible role of each gene on climate adaptation, having first divided them into three broad functional categories: 1) environmental adaptation including hypoxia, 2) body size, skeletal development and bill morphology, and 3) neurodevelopment, disease and sperm motility.

Firstly, candidate genes associated with environmental adaptation include the gene *HERPUD2*, a candidate gene for adaptation in herring to warmer water temperature around the British Isles compared to in the North Atlantic in (Han et al. 2020). The *VPS13B* gene is associated with adaptation to environmental conditions, and has been reported to be under selection in Chinese sheep (Yang et al. 2016) and Mediterranean sheep and goats (Serranito et al. 2021). The *VPS13B* protein has a role in the development of adipocytes during milk formation and shows evidence of a selective sweep in goat breeds (Amiri Ghanatsaman et al. 2023). Two of the candidate genes have a role in hypoxia (low oxygen levels). For example, adaptive alleles of the *CRIPT* gene are associated with oxygen-carrying

in the blood in Tibetan people who live in the highlands (Deng et al. 2019). Natural selection on the *PDE10A* gene has increased the spleen size in nomadic sea-diving human populations, providing a larger reservoir of oxygenated red blood cells, described as an adaptation to hypoxia tolerance (Ilardo et al. 2018).

Second, two of the candidate genes have been associated with body size; *LLPH* a candidate gene for small body size in Caspian horses (Mousavi et al. 2023) and *ACSL4* a candidate gene for body size in Donkey (Zhou et al. 2020). The candidate SNP nearest to *ACSL4* is also 158kbp away from *COL4A5*, a gene identified in Bosse et al. (2017) as being associated with longer bill length and under positive selection in the UK great tit population, possibly as an adaptation to the provision of bird feeders in UK gardens. Two further candidate genes also overlap with the genes that Bosse et al. (2017) suggested were under positive selection and affected great tit craniofacial skeleton and bill morphology. These genes were *TRPS1* and *VPS13B*, with *VPS13B* also identified as playing a role in variation in bill morphology of Darwin's finches (Lawson and Petren 2017; Sendell-Price et al. 2020). The candidate gene *TRPS1* closest to SNP AX-100271597 under selection in the combined Cambridge / Oxford (UK) populations, was one of the best candidate genes in the Chapter 2 GEA (BF > 48, XtX differentiation outlier), and is linked to skeletal dysplasia and hair and craniofacial abnormalities in humans (Cho et al. 2019). A strong selective sweep at *TRPS1* has occurred in all African rainforest hunter-gatherer groups, with selection linked to short stature (Lopez et al. 2019).

The third category of gene functions less obviously includes candidate for climate adaptation. One of the candidate genes, *LRFN2* (also termed *SALM1*), is linked to synaptic activity, with *LRFN2*-deficient mice displaying inhibited social behaviours, enhanced synaptic plasticity and enhanced memory formation (Morimura et al. 2017). Mutations in the *SRBD1* gene increase the risk of the eye condition glaucoma, which can lead to a loss of vision (Kanemaki et al. 2013). Mutations of *CD82*, a tumour suppressor gene, likely influence the immune response and metastasis suppression (Tsai and Weissman 2011). Polymorphisms in the gene *NSUN7* increase the risk of neuroblastoma (Zhou et al. 2020b)

and are associated with sperm motility defects in mice (Khosronezhed et al. 2015). In addition, a further significant sweep was detected for a UK population in combination with other high PC1 populations (Scotland, France and Corsica) near *ATP6AP2*. Despite being identified in Chapter 2 as one of the loci with the strongest association to climate (BF > 30), *ATP6AP2* is a (pro)renin receptor with tissue-specific activation required for vertebrate development (Wendling et al. 2017), but with a gene function less obviously linked to climate adaptation.

3.5.2. Selective sweeps in non-UK outlier populations

Selective sweeps in non-UK outlier populations were reported for four different populations, with candidate genes identified with functions related to adaptation to altitude and thermotolerance:

1) In Finland there was a sweep centered on a SNP close to candidate gene *TTC7A* (associated with climate-PC1). Differential gene expression at the *TTC7A* has been reported in a comparison of great tit embryos that were grown in conditions resembling high and low altitudes (Chen et al. 2022).

2) In the Spanish population there was a sweep near to candidate gene *JAK2* (associated with PC2). Mutations in *JAK2* lead to the production of too many red blood cells (clinically a blood cancer termed polycythemia vera), with cases observed to be more frequent in people living at high-altitude (Moliterno et al. 2023). Blockade of the *JAK2* pathway protects mice against hypoxia-induced hypertension (Zhang et al. 2020).

3) In the combined France/Corsica population there is a sweep close to candidate gene *MRPS33* (associated with climate-PC2). Selection at *MRPS33* has been detected in African cattle, and the gene is suggested to play a role in the evolution of thermo-tolerance (Taye et al. 2017).

4) In the combined low PC4 populations (Spain, Switzerland and Belgium) there was a sweep near *PI4KA*, identified in Chapter 2 as 'decisively' associated with climate (BF > 30); however the gene displays a less obvious adaptive gene function. Human patients with

mutations in *PI4KA* have a severe related metabolic disorder with causes changes in brain and neurodevelopment (Verdura et al. 2021).

In summary, the candidate genes identified in positive sweep regions have known biological processes (as referenced in the above detailed descriptions for each gene) associated with environmental adaptation (*HERPUD2*, *VPS13B*), thermo-tolerance (*MRPS33*), hypoxia (*CRIPT*, *PDE10A*, *TTC7A*, *JAK2*), body size (*LLPH*, *ACSL4*), skeletal development (*TRPS1*), bill morphology (*VPS13B*), synaptic activity and memory (*LRFN2*), vertebrate development (*ATP6AP2*), disease including glaucoma (*SRBD1*), metabolic disorder (*PI4KA*), metastasis suppression (*CD82*), neuroblastomas (*NSUN7*), and sperm motility (*NSUN7*).

3.5.3. Do selective sweeps provide support for parallel evolution?

Closer investigation of the outlier populations revealed candidate genes had generally been selected in just one population or geographically close populations, with only one candidate SNP under parallel evolution in different outlier populations. This SNP, AX-100430005, appears to have undergone parallel selective sweeps in Scotland and in the combined France/Corsica populations. The closest gene to this SNP is *GNAL*. Mutations in *GNAL* cause primary muscle dystonia, affecting movement in human patients (Fuchs et al. 2012). Thus, only one account of parallel evolution is observed, which provides low evidence of parallel evolution occurring for the identified selective sweeps of this Chapter.

3.5.4. Are candidate genes in regions of low recombination?

Some researchers have suggested that false positives are more likely in regions of low recombination (e.g. Burri et al. 2015; Burri 2017) including in our study populations (Perrier and Charmantier 2019; in response to Bosse et al. 2017). For example, regions detected with increased differentiation, reduced genetic diversity and increased LD may look like

selective sweeps, but could have resulted from low recombination and background selection (Burri et al. 2015). To consider the attribution of selective sweeps, we looked at the location of our candidate SNPs with respect to local recombination, and tested these relative to the recombination rate at randomly chosen SNP regions of the genome. We found evidence that our sweep regions were disproportionately likely to be in regions with low local recombination rate. This does not necessarily mean they are false positives, as it could be that the power to detect sweeps is greater in those regions. Signatures of genuine sweeps should last longest in regions with low recombination rates. We also found evidence that the SNPs associated with climate for PC1-PC3 from the GEA in Chapter 2 were biased for SNPs in regions of low recombination, which may reflect why more candidate SNPs were also found in regions of low recombination. In other words, any bias that has occurred could have happened in the GEA analyses, rather than the haplotype-based ones.

3.5.5. Limitation of methods to detect sweeps

In this chapter the general method of choosing outlier population groups was based on shared climate-PC space, which did not always reflect geographical distance or expected population gene flow from migration. A more targeted selection of which populations to compare could help investigate local adaptation in finer detail, for example like the replicated choice of urban vs rural paired great tit populations (Salmón et al. 2021). Comparisons between populations could be aided by sampling more birds per population and more populations distributed across the species range.

It is important to consider that identifying signatures of recent positive selection by only one or a few approaches (in this case, by extended haplotype-based testing and by the BF evidence for associations with climate), requires real caution when inferring causation of local adaptation in a polygenic trait. The false detection of positive results is statistically likely when hundreds of thousands of SNPs are tested. More stringent thresholds can be considered to reduce false-positives, especially in larger SNP data sets with a higher number of false-positives to be expected (Klassmann and Gautier 2022). However, true sites

of selection may then be missed. Other uncertainties in the methods may include the handling of large gaps between markers, and the clustering of significant scores (see vignette for package REHH 2023 for ways to deal with data considerations). Klassmann and Gautier 2022 recommend in general that only a handful of the most extreme outlier regions can be justified to be investigated in further detail, with caution needed if reporting a large number of putative selective sweeps, due to false-detection rates (FDR) introduced by the statistical method.

3.5.6. Combined methods to detect sweeps

This Chapter focused on cross population XP-EHH and Rsb haplotype statistics because they were commonly co-reported in the literature of detecting selective sweeps in vertebrates, and both values could be calculated for ~400K SNPs in a REHH package framework on a HPC cluster. The calculated XP-EHH and Rsb values of this data analysis were moderately correlated, but different enough to justify reporting both values. Notably, overall we found more of our candidate loci (from Chapter 2) subsequently produced evidence of selective sweeps by combined XP-EHH and Rsb evidence (this chapter) than expected by chance.

Other available EHH tests that can be applied to compare populations and genome scan for local adaptation, include: 1) the cross population-composite likelihood ratio test (XP-CLR), developed by Chen et al. (2010) to detect sweep regions of the genome where the change in allele frequency has occurred too quickly relative to expectations under a model of neutral genetic drift. For example, this test was used to detect selective sweeps in domesticated macadamia nuts (Lin et al. 2022). 2) the cross-population number of segregating sites by length (XP-nSL). This test was developed by Szpiech et al. (2021) to compare the length of haplotype homozygosity between alleles from derived and ancestral populations, in terms of mutation patterns. XP-nSL is able to detect ongoing and recently completed hard and soft sweeps, and was originally applied to identify high altitude

adaptation in rhesus macaques (see Garud 2023 for an explanation of rapid adaptation signatures in soft sweeps).

Different haplotype-based tests provide better resolution for detecting selective sweeps depending on whether the focal allele has moderate frequencies (e.g. iHS) or has become (nearly) fixed within one population (e.g. XP-EHH, Rsb, XP-CLR) or to detect ongoing sweeps (e.g. XP-nSL) (Chen et al. 2016). Haplotype-based methods to detect selection can be combined, with the aim of providing better resolved and more robust evidence of the signatures of selection. For example, Lin et al. (2022) identified genomic regions under selection in populations of cultivated cucumber, by combining four haplotype-based approaches (iHS, XP-nSL, XP-EHH and Rsb) with Fisher's combination method (Fisher 1954). An alternative, suggestion on how to improve the resolution of identifying genomic regions under selection, is to combine the haplotype statistics via de-correlated composition of multiple signals (DCMS) of the P-values (Ma et al. 2015). Overall, the different tests for selection are considered complimentary and the combination of these tests an optimal strategy for resolving selection signatures with different features (Chen et al. 2016).

Ultimately the classification of an adaptation gene would require experimental validation or the combination of genomic data with phenotype and life history data from natural populations. Only then is it possible to demonstrate that selected alleles result in increased fitness and survival, which is difficult to directly measure in the wild. For example, natural selection for variants of the gene EPAS1 is observed in the Tibetan human population. These alleles have helped with adaptation to living at high altitude, while individuals not carrying the adaptive variants use oxygen less efficiently and struggle with the effects of altitude sickness (Simonson et al. 2010).

In summary, this chapter detected the populations where selection may have acted. Seventeen candidate climate adaptation SNPs were identified as having undergone selective sweeps, mostly in the outlier populations. The closest genes to these candidate

SNPs may potentially play a role in adaptation to climate, with suggested roles for transporting oxygen and for thermoregulation and body size/skeletal adaptations to deal with local climatic and temperature conditions. Future studies may focus investigation on the 12 selective sweeps associated with climate identified in the UK populations. Whole genome sequencing of museum great tit specimens from UK populations could be performed to test if significant changes in allele frequencies have occurred in these candidate climate adaptation genes over a time-series of samples (**Chapter 4**).

Thinking more generally, climate did not explain all of the detected positive sweeps, so understanding adaptation to other environmental factors and other agents of selection could be investigated further. Further understanding of the genomics of climate adaptation in natural populations of great tit could be gained from comparing replicated matched pairs of populations, contrasting climatic extremes especially in more peripheral populations e.g. those exposed to high levels of heat stress or winter cold tolerance vs a control population from mainland Europe. Perhaps haplotype-based methods will be developed to reduce the implication of sweeps being falsely detected in low recombination regions, to aid the detection of candidate adaptive loci.

3.6. References

Abondio, P., Cilli, E., Luiselli, D. (2022). Inferring Signatures of Positive Selection in Whole-Genome Sequencing Data: An Overview of Haplotype-Based Methods. *Genes*, **13**(5):926. doi: 10.3390/genes13050926

Álvarez, I., Fernández, I., Traoré, A., Pérez-Pardal, L., Menéndez-Arias, N.A. & Goyache, F. (2020). Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments. *Sci. Rep.*, **10**:2824. doi:10.1038/s41598-020-59839-x

Amiri Ghanatsaman, Z., Ayatolahi Mehrgardi, A., Asadollahpour Nanaei, H. & Esmailizadeh, A. (2023). Comparative genomic analysis uncovers candidate genes related with milk production and adaptive traits in goat breeds. *Sci. Rep.*, **13**:8722. doi:10.1038/s41598-023-35973-0

Barrett, R.D.H. & Schluter, D. (2008). Adaptation from standing genetic variation. *Trends Ecol. Evol.*, **23**(1):38-44 doi:10.1016/j.tree.2007.09.008

Bernatchez, L., Ferchaud, A.L., Berger, C.S., Venney, C. J., & Xuereb, A. (2023). Genomics for monitoring and understanding species responses to global climate change. *Nat. Rev. Genet.*, **25**:165-183. doi:10.1038/s41576-023-00657-y

Bonhomme, M., Chevalet, C., Servin, B., Boitard, S., Abdallah, J., Blott, S. & Sancristobal, M. (2010). Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*, **186**(1):241-62. doi: 10.1534/genetics.104.117275

Bosse, M., Spurgin, L. G., Laine, V. N., Cole, E. F., Firth, J. A., Gienapp, P., Gosler, A. G., McMahon, K., Poissant, J., Verhagen, I., Groenen, M. A. M., van Oers, K., Sheldon, B. C., Visser, M. E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, **358**(6361), 365–368. doi:10.1126/science.aal3298

Browning, B.L., Tian, X., Zhou, Y. & Browning, S.R. (2021) Fast two-stage phasing of large-scale sequence data. *Am. J. Hum. Genet.*, **108**(10):1880-1890. doi:10.1016/j.ajhg.2021.08.005

Burke, M.K. (2012). How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc. R. Soc. B.*, **279**:5029–5038. doi:10.1098/rspb.2012.0799

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L., Suh, A., Dutoit, L., Bureš, S., Garamszegi, L.Z., Hogner, S., Moreno, J., Qvarnström, A., Ružić, M., Sæther, S-A., Sætre, G-P., Török, J., & Ellegren, H. (2015). Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of Ficedula flycatchers. *Genome Res.*, **25**:1656–1665. doi:10.1101/gr.196485.115

Burri, R. (2017). Interpreting differentiation landscapes in the light of long-term linked selection. *Evol. Lett.*, **1**:118–131. doi:10.1002/evl3.14

Ceballos, F.C., Joshi, P.K., Clark, D.W., Ramsay, M. & Wilson, J.F. (2018). Runs of homozygosity: windows into population history and trait architecture. *Nat. Rev. Genet.*, **19**: 220–234. doi:10.1038/nrg.2017.109

Ceccobelli, S., Landi, V., Senczuk, G., Mastrangelo, S., Sardina, M.T., Ben-Jemaa, S., Persichilli, C., Karsli, T., Bâlteanu, V-A., Raschia, M.A., Poli, M.A., Ciappesoni, G., Muchadeyi, F.C., Dzomba, E.F., Kunene, N.W., Lühken, G., Deniskova, T.E., Dotsev, A.V., Zinovieva, N.A., ... & Pilla, F. (2023). A comprehensive analysis of the genetic diversity and environmental adaptability in worldwide Merino and Merino-derived sheep breeds. *Genet. Sel. Evol.*, **55**:e24. doi:10.1186/s12711-023-00797-z

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S.M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, **4**(7), 1–16. doi :10.1186/s13742-015-0047-8

Chen, H., Patterson, N. & Reich, D. (2010). Population differentiation as a test for selective sweeps. *Genome Res.*, **20**(3):393–402. doi:10.1101/gr.100545.109

Chen, M., Pan, D., Ren, H., Fu, J., Li, J., Su, G., Wang, A., Jiang, L., Zhang, Q. & Liu, J-F. (2016). Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genet. Sel. Evol.*, **48**:76. doi:10.1186/s12711-016-0254-5

Chen, N., Xia, X., Hanif, Q., Zhang, F., Dang, R., Huang, B., Lyu, Y., Luo,X., Zhang, H., Yan, H., Wang, S., Wang, F., Chen, J., Guan, X., Liu, Y., Li, S., Jin, L., Wang, P., Sun, L. ... Lei, C. (2023). Global genetic diversity, introgression, and evolutionary adaptation of indicine cattle revealed by whole genome sequencing. *Nat. Comm.*, **14**:7803. doi:10.1038/s41467-023-43626-z

Chen, X., Ji, Y., Cheng, Y., Hao, Y., Lei, X., Song, G., Qu, Y. & Lei, F. (2022). Comparison between short-term stress and long-term adaptive responses reveal common paths to molecular adaptation. *iScience*, **25**(3):103899. doi: 10.1016/j.isci.2022.103899

Cheng, Y., Miller, M.J., Zhang, D., Xiong, Y., Hao, Y., Jia, C., Cai, T., Li, S-H., Johansson, U.S., Liu, Y., Chang, Y., Song, G.,Qu, Y., & Lei, F. (2021). Parallel genomic responses to historical climate change and high elevation in East Asian songbirds. *Proc. Natl. Acad. Sci U S A.*, **118**(50):e2023918118. doi: 10.1073/pnas.2023918118

Cho, K. Y., Kelley, B. P., Monier, D., Lee, B., Szabo-Rogers, H., & Napierala, D. (2019). Trps1 Regulates Development of Craniofacial Skeleton and Is Required for the Initiation of Palatal Shelves Fusion. *Front. Physiol.*, **10**, 513. doi:10.3389/fphys.2019.00513

Crawford, D.C., & Nickerson, D.A. (2005). Definition and clinical importance of haplotypes. *Ann. Rev. Med.*, **56**:303–320. doi:10.1146/annurev.med.56.082103.104540

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., & 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics*, **27**(15):2156-8. doi: 10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2), giab008. doi:10.1093/gigascience/giab008

DeGiorgio, M., Huber, C.D., Hubisz, M.J., Hellmann, I. & Nielsen, R. (2016). SweepFinder2: increased sensitivity, robustness and flexibility. *Bioinformatics*, **32**(12):1895–1897. doi:10.1093/bioinformatics/btw051

De La Torre, A.R., Sekhwal, M.K., & Neale, D.B. (2021). Selective Sweeps and Polygenic Adaptation Drive Local Adaptation along Moisture and Temperature Gradients in Natural Populations of Coast Redwood and Giant Sequoia. *Genes*, **12**(11):1826. doi:10.3390/genes12111826

Deng, L., Zhang, C., Yuan, K., Gao, Y., Pan, Y., Ge, X., He, Y., Yuan, Y., Lu, Y., Zhang, X., Chen, H., Lou, H., Wang, X., Lu, D., Liu, J., Tian, L., Feng, O., Khan, A., Yang, Y., Jin, Z-B., Yang, J., Lu, F., Qu, J., Kang, L., Su, B. & Xu, S. (2019). Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *Natl. Sci. Rev.*, **6**(6):1201–1222. doi:10.1093/nsr/nwz108

Ellegren, H. (2013). The Evolutionary Genomics of Birds. *Annu. Rev. Ecol. Evol. Syst.*, **44**(1):239-259. doi:10.1146/annurev-ecolsys-110411-160327

Eydivandi, S., Roudbar, M.A., Karimi, M.O. & Sahana, G. (2021). Genomic scans for selective sweeps through haplotype homozygosity and allelic fixation in 14 indigenous sheep breeds from Middle East and South Asia. *Sci. Rep.*, **11**:2834. doi:10.1038/s41598-021-82625-2

Fariello, M.I., Boitard, S., Naya, H., SanCristobal, M., & Servin, B. (2013). Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*, **193**(3):929–941. doi:10.1534/genetics.112.147231

Fay, J.C. & Wu, C.I. (2000). Hitchhiking under positive Darwinian selection. *Genetics*, **155**(3):1405-13. doi: 10.1093/genetics/155.3.1405

Fick, S.E. & Hijmans, R.J. (2017). WorldClim 2: New 1-km Spatial Resolution Climate Surfaces for Global Land Areas. *Int. J. Climatol.*, **37**:4302-4315. doi:10.1002/joc.5086

Fisher, R.A., (1954). Statistical methods for research workers, 12th ed. Edinburgh: Oliver and Boyd.

Friedrich, J. and Wiener, P. (2020), Selection signatures for high-altitude adaptation in ruminants. *Anim. Genet*., **51**:157-165. doi:10.1111/age.12900

Fuchs, T., Saunders-Pullman, R., Masuho, I., San Luciano, M., Raymond, D., Factor, S., Lang, A.E., Liang, T-W., Trosch, R.M., White, S., Ainehsazan, E., Hervé, D., Sharma, N., Ehrlich, M.E., Martemyanov, K.A., Bressman, S.B. & Ozelius, L.J. (2013). Mutations in GNAL cause primary torsion dystonia. *Nat. Genet*., **45**:88–92. doi:10.1038/ng.2496

Gabriel, S.B., Schaffner, S.F., Nguyen, H., Moore, J.M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S.N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E.S., Daly, M.J., & Altshuler, D. (2002). The Structure of Haplotype Blocks in the Human Genome. *Science*, **296**:2225-2229. doi:10.1126/science.1069424

Gautier, M., Klassmann, A. and Vitalis, R. (2017), rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol. Ecol. Resour*., **17**: 78-90. doi:10.1111/1755-0998.12634

Gautier, M. & Vitalis, R. (2012). rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics*, **28**(8):1176–1177. doi:10.1093/bioinformatics/bts115

Han, F., Jamsandekar, M., Pettersson, M.E., Su, L., Fuentes-Pardo, A.P., Davis, B.W., Bekkevold, D., Berg, F., Casini, M., Dahle, G., Farrell, E.D., Folkvord, A. & Andersson, L. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *eLife*, **9**:e61076. doi:10.7554/eLife.61076

Hansson, B., Ljungqvist, M., Dawson, D., Mueller, J.C., Olano-Marin, J., Ellegren, H. & Nilsson, J- Å. (2010). Avian genome evolution: insights from a linkage map of the blue tit (Cyanistes caeruleus). *Heredity*, **104**:67–78. doi:10.1038/hdy.2009.107

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., Poss, M.L., Reed, L.K., Storfer, A & Whitlock, M.C. (2016). Finding the Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions. *Am. Nat.*, **188**(4):379-397. doi:10.1086/688018

Hu, J., Chen, B., Zhao, J., Zhang, F., Xie, T., Xu, K., Gao, G., Yan, G., Li, H., Li, L., Ji, G., An, H., Li, H., Huang, Q., Zhang, M., Wu, J., Song, W., Zhang, X., Luo, Y. ... Wu, X. (2022). Genomic selection and genetic architecture of agronomic traits during modern rapeseed breeding. *Nat. Genet.*, **54**: 694–704. doi:10.1038/s41588-022-01055-6

Ilardo, M. A., Moltke, I., Korneliussen, T. S., Cheng, J., Stern, A. J., Racimo, F., de Barros Damgaard, P., Sikora, M., Seguin-Orlando, A., Rasmussen, S., van den Munckhof, I. C. L., Ter Horst, R., Joosten, L. A. B., Netea, M. G., Salingkat, S., Nielsen, R., & Willerslev, E. (2018). Physiological and Genetic Adaptations to Diving in Sea Nomads. *Cell*, **173**(3), 569–580.e15. doi:10.1016/j.cell.2018.03.054

Jones, F., Grabherr, M., Frank Chan, Y. Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M.C., Myers, R.M., Miller, C.T., Summers, B.R., Knecht, A.K., ... Kingsley, D.M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**:55–61. doi:10.1038/nature10944

Kanemaki, N., Tchedre, K.T., Imayasu, M., Kawarai, S., Sakaguchi, M., Yoshino, A., et al. (2013) Dogs and Humans Share a Common Susceptibility Gene SRBD1 for Glaucoma Risk. *PLoS ONE*, **8**(9):e74372. doi:10.1371/journal.pone.0074372

Kim, E.-S., Elbeltagy, A.R., Aboul-Naga, A.M., Rischkowsky, B., Sayre,B., Mwacharo, J.M. & Rothschild, M.F. (2016). Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity*, **116**:255–264, doi:10.1038/hdy.2015.94

Kim, J-M., Santure, A.W., Barton, H.J., Quinn, J.L., Cole, E.F., Great Tit HapMap Consortium, Visser, M.E., Sheldon, B.C., Groenen, M.A.M., van Oers, K. & Slate, J. (2018). A high-density SNP chip for genotyping great tit (Parus major) populations and its application to studying the genetic architecture of exploration behaviour. *Mol. Ecol. Res.*, **18**(4):877–891. doi:10.1111/1755-0998.12778

Kim, Y., & Nielsen, R. (2004). Linkage Disequilibrium as a Signature of Selective Sweeps. *Genetics*, **167**(3):1513–1524. doi:10.1534/genetics.103.025387

Klassmann, A. & Gautier, M. (2022). Detecting selection using extended haplotype homozygosity (EHH)-based statistics in unphased or unpolarized data. *PLoS One*, **17**(1):e0262024. doi: 10.1371/journal.pone.0262024

Khosronezhad, N., Hosseinzadeh Colagar, A., Mortazavi, S.M. (2015). The Nsun7 (A11337)-deletion mutation, causes reduction of its protein rate and associated with sperm motility defect in infertile men. J. Assist. Reprod. *Genet.*, **32**(5):807-15. doi:10.1007/s10815-015-0443-0

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O., Verhoeven, K.J.F., de Jager, V., Megens, H-J., Warren, W.C., Minx, P., Crooijmans, R.P.M.A., Corcoran, P., Great Tit HapMap Consortium, Sheldon, B.C., Slate, J., Zeng, K., van Oers, K., Visser, M.E. & Groenen, M.A.M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Comm.*, **7**:10474. doi:10.1038/ncomms10474

Lamichhaney, S., Berglund, J., Almén, M.S., Maqbool, K., Grabherr, M., Martinez-Barrio, A., Promerová, M., Rubin, C-J., Wang, C., Zamani, N., Grant, B.R., Grant, P.R., Webster, M.T., & Andersson, L. (2015). Evolution of Darwin's finches and their beaks revealed by genome sequencing. *Nature*, **518**:371–375. doi:10.1038/nature14181

Lange, J. D., & Pool, J. E. (2016). A haplotype method detects diverse scenarios of local adaptation from genomic sequence variation. *Mol. Ecol.*, **25**(13):3081–3100. doi:10.1111/mec.13671

Lawson, L.P. & Petren, K. (2017). The adaptive genomic landscape of beak morphology in Darwin's finches. *Mol. Ecol.*, **26**(19), 4978–4989. doi:10.1111/mec.14166

Lewontin, R. & Krakauer, J. (1973). Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. *Genetics*, **74**:175–195. doi:10.1093/genetics/74.1.175

Lin, J., Zhang, W., Zhang, X., Ma, X., Zhang, S., Chen, S., Wang, Y., Jia, H., Liao, Z., Lin, J., Zhu, M., Xu, X., Cai, M., Zeng, H., Wan, J., Yang, W., Matsumoto, T., Hardner, C., Nock, C.J. & Ming, R. (2022). Signatures of selection in recently domesticated macadamia. *Nat. Comm.*, **13**:242. doi:10.1038/s41467-021-27937-7

Lin, X., Zhang, N., Song, H., Lin, K. & Pang, E. (2022). Population-specific, recent positive selection signatures in cultivated Cucumis sativus L. (cucumber). *G3 (Bethesda)*, **12**(7):jkac119. doi: 10.1093/g3journal/jkac119

Lopez, M., Choin, J., Sikora, M., Siddle, K., Harmant, C., Costa, H.A., Silvert, M., Mouguiama-Daouda, P., Hombert, J.M., Froment, A., Le Bomin, S., Perry, G.H., Barreiro, L.B., Bustamante, C.D., Verdu, P., Patin, E., & Quintana-Murci, L. (2019). Genomic Evidence for Local Adaptation of Hunter-Gatherers to the African Rainforest. *Curr. Biol.*, **29**(17):2926–2935.e4. doi:10.1016/j.cub.2019.07.013

Ma, Y., Ding, X., Qanbari, S., Weigend, S. & Zhang, Q. (2015). Properties of different selection signature statistics and a new strategy for combining them. *Heredity*, **115**:426–436. doi:10.1038/hdy.2015.42

Maclean, C.A., Chue Hong, N.P., & Prendergast, J. G. (2015). hapbin: An Efficient Program for Performing Haplotype-Based Scans for Positive Selection in Large Genomic Datasets. *Mol. Biol. Evol.*, **32**(11):3027–3029. doi:10.1093/molbev/msv172

Malécot G. 1948. Les mathématiques de l'hérédié. Masson & Cie, Paris, France.

Moliterno, A.R., Kaizer, H. & Reeves, B.N. (2023). JAK2V617F allele burden in polycythemia vera: burden of proof. *Blood*, **141**(16):1934–1942. doi:10.1182/blood.2022017697

Montejo-Kovacevich, G., Meier, J.I., Bacquet, C.N., Warren, I.A., Frank Chan, Y. Kucka, M., Salazar, C., Rueda-M, N., Montgomery, S.H., McMillan, W.O., Kozak, K.M., Nadeau, N.J., Martin, S.H. & Jiggins, C.D. (2022). Repeated genetic adaptation to altitude in two tropical butterflies. *Nat. Comm.*, **13**:4676. doi:10.1038/s41467-022-32316-x

Morimura, N., Yasuda, H., Yamaguchi, K., Katayama, K.-i., Hatayama, M., Tomioka, N.H., Odagawa, M., Kamiya, A., Iwayama, Y., Maekawa, M., Nakamura, K., Matsuzaki, H., Tsujii, M., Yamada, K., Yoshikawa, T. & Aruga, J. (2017). Autism-like behaviours and enhanced memory formation and synaptic plasticity in Lrfn2/SALM1-deficient mice. *Nat. Comm.*, **8**:15800. doi:10.1038/ncomms15800

O'Connell, J., Gurdasani, D., Delaneau, O., Pirastu, N., Ulivi, S., Cocca, M., Traglia, M., Huang, J., Huffman, J.E., Rudan, I., McQuillan, R., Fraser, R.M., Campbell, H., Polasek, O., Asiki, G., Ekoru, K., Hayward, C., Wright, A.F., Vitart, V., … Marchini, J. (2014). A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet.*, **10**(4): e1004234. doi:10.1371/journal.pgen.1004234

Oleksyk, T.K., Smith, M.W. & O'Brien, S.J. (2010). Genome-wide scans for footprints of natural selection. *Philos. Trans. R. Soc. Lond. B Biol. Sci.*, **365**(1537):185-205. doi: 10.1098/rstb.2009.0219

Olson-Manning, C., Wagner, M. & Mitchell-Olds, T. (2012). Adaptive evolution: evaluating empirical support for theoretical predictions. *Nat. Rev. Genet.*, **13**:867–877. doi:10.1038/nrg3322

Pavlidis, P., Živkovic, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: likelihood-based detection of selective sweeps in thousands of genomes. *Mol. Biol. Evol.*, **30**(9):2224–2234. doi:10.1093/molbev/mst112

Peng, Y., Yang, Z., Zhang, H., Cui, C., Qi, X., Luo, X., Tao, X., Wu, T., Ouzhuluobu, Basang, Ciwangsangbu, Danzengduojie, Chen, H., Shi, H. & Su, B. (2011). Genetic Variations in Tibetan Populations and High-Altitude Adaptation at the Himalayas, *Mol. Biol. Evol.*, **8**(2):1075–1081. doi:10.1093/molbev/msq290

Perrier, C. & Charmantier, A. (2019). On the importance of time scales when studying adaptive evolution. *Evol. Lett.*, **3**(3):240–247. doi:10.1002/evl3.86

Pritchard, J. & Di Rienzo, A. Adaptation – not by sweeps alone. (2010). *Nat. Rev. Genet.*, **11**, 665–667. doi:10.1038/nrg2880

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A., Bender, D., Maller, J., Sklar, P., de Bakker, P.I., Daly, M.J. & Sham, P.C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.*, **81**(3): 559-75. doi:10.1086/519795

Ravinet, M., Elgvin, T.O., Trier. C., Alibadian, M., Gavrilov, A. & Sætre, G-P. (2018). Signatures of human-commensalism in the house sparrow genome. *Proc. R. Soc.*, **285**(1884): 20181246. doi:10.1098/rspb.2018.1246

Sabeti, P.C., Reich, D.E., Higgins, J.M., Levine, H.Z., Richter, D.J., Schaffner, S.F., Gabriel, S.B., Platko, J.V., Patterson, N.J., McDonald, G.J., Ackerman, H.C., Campbell, S.J., Altshuler, D., Cooper, R., Kwiatkowski, D., Ward, R., & Lander, E.S. (2002). Detecting recent positive selection in the human genome from haplotype structure. *Nature*, **419**(6909):832–837. doi:10.1038/nature01140

Sabeti, P.C., Varilly, P., Fry, B., Lohmueller, J., Hostetter, E., Cotsapas, C., Xie, X., Byrne, E.H., McCarroll, S.A., Gaudet, R., Schaffner, S.F., Lander, E.S. & The International HapMap Consortium (2007). Genome-wide detection and characterization of positive selection in human populations. *Nature*, **449**:913–918. doi:10.1038/nature06250

Salmón, P., Jacobs, A., Ahrén, D., Biard, C., Dingemanse, N.J., Dominoni, D.M., Helm, B., Lundberg, M., Senar, J.C., Sprau, P., Visser, M.E., & Isaksson, C. (2021). Repeated genomic signatures of adaptation to urbanisation in a songbird across Europe. *Nat. Comm.*, **12**:2983. doi:10.1038/s41467-021-23027-w

Savolainen, O., Lascoux, M. & Merilä, J. (2013). Ecological genomics of local adaptation. *Nat. Rev. Genet.*, **14**:807–820. doi:10.1038/nrg3522

Scheet, P. & Stephens, M. (2006). A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.*, **78**:629–644. doi: 10.1086/502802

Sendell-Price, A.T., Ruegg, K.C. & Clegg, S.M. (2020). Rapid morphological divergence following a human-mediated introduction: the role of drift and directional selection. *Heredity*, **124**:535-549. doi: 10.1038/s41437-020-0298-8

Serranito, B., Cavalazzi, M., Vidal, P., Taurisson-Mouret, D., Ciani, E., Bal, M., Rouvellac, E., Servin, B., Moreno-Romieux, C., Tosser-Klopp, G., Hall, S.J.G., Lenstra, J.A., Pompanon, F., Benjelloun, B. & Da Silva, A. (2021). Local adaptations of Mediterranean sheep and goats through an integrative approach. *Sci. Rep.*, **11**:1–17. doi:10.1038/s41598-021-00682-z

Simonson, T.S., Yang, Y., Huff, C.D., Yun, H., Qin, G., Witherspoon, D.J., Bai, Z., Lorenzo, F.R., Xing, J., Jorde, L.B., Prchal, J.T. & Ge, R. (2010). Genetic Evidence for High-Altitude Adaptation in Tibet. *Science*, **329**(5987):72-75. doi:10.1126/science.1189406

Smith, E.G., Hazzouri, K.M., Young Choi, J., Delaney, P., Al-Kharafi, M., Howells, E.J., Aranda, M. & Burt, J.A. (2022). Signatures of selection underpinning rapid coral adaptation to the world's warmest reefs. *Sci. Adv.*, **8**(2):eabl7287. doi:10.1126/sciadv.abl7287

Spurgin, L.G., Bosse, M., Adriaensen, F., Albayrak, T., Barboutis, C., Belda, E., Bushuev, A., Cecere, J.G., Charmantier, A., Cichon, M., Dingemanse, N.J., Doligez, B., Eeva, T., Erikstad, K.E., Fedorov, V., Griggio, M., Heylen, D., Hille, S., Hinde, C.A., … Slate, J. (2024). The great tit HapMap project: A continental-scale analysis of genomic variation in a songbird. *Mol. Ecol. Resour.*, **24**:e13969. https://doi.org/10.1111/1755-0998.13969

Szpiech, Z.A. & Hernandez, R.D. (2014). selscan: an efficient multi-threaded program to calculate EHH-based scans for positive selection. *Mol. Biol. Evol.*, **31**(10): 2824-2827. doi:10.1093/molbev/msu211

Szpiech, Z.A., Novak, T.E., Bailey, N.P. & Stevison, L.S. (2021). Application of a novel haplotype-based scan for local adaptation to study high-altitude adaptation in rhesus macaques. *Evol. Lett.*, **5**:408-421. doi:10.1002/evl3.232

Tajima F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**(3):585–595. doi:10.1093/genetics/123.3.585

Tang, K., Thornton, K.R. & Stoneking, M. (2007). A New Approach for Using Genome Scans to Detect Recent Positive Selection in the Human Genome. *PLoS Biol.*, **5**(7):e171. doi:10.1371/journal.pbio.0050171

Taye, M., Lee, W., Caetano-Anolles, K., Dessie, T., Hanotte, O., Mwai, O.A., Kemp, S., Cho, S., Oh, S.J., Lee, H.K., & Kim, H. (2017). Whole genome detection of signature of positive selection in African cattle reveals selection for thermotolerance. *Anim. Sci. J.*, **88**(12):1889–1901. doi:10.1111/asj.12851

Tsai, Y. C. & Weissman, A. M. (2011). Dissecting the diverse functions of the metastasis suppressor CD82/KAI1. *FEBS Lett.*, **585**(20):3166–3173. doi:10.1016/j.febslet.2011.08.031

van Oers, K., Santure, A.W., De Cauwer, I., van Bers, N.E., Crooijmans, R.P., Sheldon, B.C., Visser, M.E., Slate, J., & Groenen, M.A. (2014). Replicated high-density genetic maps of two great tit populations reveal fine-scale genomic departures from sex-equal recombination rates. *Heredity*, **112**(3):307–316. doi:10.1038/hdy.2013.107

Verdura, E., Rodríguez-Palmero, A., Vélez-Santamaria, V., Planas-Serra, L., de la Calle, I., Raspall-Chaure, M., Roubertie, A., Benkirane, M., Saettini, F., Pavinato, L., Mandrile, G., O'Leary, M., O'Heir, E., Barredo, E., Chacón, A., Michaud, V., Goizet, C., Ruiz, M., Schlüter, A., ... Pujol, A. (2021). Biallelic PI4KA variants cause a novel neurodevelopmental syndrome with hypomyelinating leukodystrophy. *Brain*, **144**(9):2659–2669. doi:10.1093/brain/awab124

Vignette for package REHH (2023).
https://cran.r-project.org/web/packages/rehh/vignettes/rehh.html#data-considerations

Voight, B.F., Kudaravalli, S., Wen, X. & Pritchard, J.K. (2006). A map of recent positive selection in the human genome. *PLoS Biol.*, **4**(3):e72. doi: 10.1371/journal.pbio.0040072. Erratum in: *PLoS Biol.* 2006 Apr;**4**(4):e154. Erratum in: PLoS Biol. 2007 Jun;5(6):e147

Vucetich, J.A. & Waite., T.A. (2003). Spatial patterns of demography and genetic processes across the species' range: Null hypotheses for landscape conservation genetics. *Conserv. Genet.*, **4**:639–645. doi:10.1023/A:1025671831349

Wasakul, V., Disratthakit, A., Mayxay, M., Chindavongsa, K., Sengsavath, V., Thuy-Nhien, N.,Pearson, R.D., Phalivong, S., Xayvanghang, S., Maude, R.J., Gonçalves, S., Day, N.P., Newton, P.N., Ashley, E.A., Kwiatkowski, D.P., Dondorp, A.M. & Miotto, O. (2023). Malaria outbreak in Laos driven by a selective sweep for Plasmodium falciparum kelch13 R539T

mutants: a genetic epidemiology analysis. *Lancet. Infect. Dis.*, **23**:568–77. doi:10.1016/S1473-3099(22)00697-1

Wei, K., Silva-Arias, G.A. & Tellier, A. (2023). Selective sweeps linked to the colonization of novel habitats and climatic changes in a wild tomato species. *New Phytol.*, **237**:1908–1921. doi: 10.1111/nph.18634

Wendling, O., Champy, MF., Jaubert, S., Pavlovic, G., Dubos, A., Lindner, L., Jacobs, H., Mark, M., Combe, R., Goncalves Da Cruz, I., Luche, H., Mudgett, J.S., Rosahl, T., Sorg, T., Malissen, M., Reilly, P.T. & Hérault, Y. (2017). Atp6ap2 ablation in adult mice impairs viability through multiple organ deficiencies. *Sci. Rep.*, **7**:9618. doi:10.1038/s41598-017-08845-7

Wright, S. (1949). The genetical structure of populations. *Ann. Hum. Genet.*, **15**(4):323–354. doi:10.1111/j.1469-1809.1949.tb02451.x

Yang, J., Li, W-R., Lv, F-H., He, S-G., Tian, S-L., Peng, W-F., Sun, Y-W., Zhao, Y-X., Tu, X-L., Zhang, M., Xie, X-L., Wang, Y-T., Li, J-Q., Liu, Y-G., Shen, Z-Q., Wang, F., Liu, G-J., Lu, H-F., Kantanen, J., Han, J-L., Li, M-H. & Liu, M-J. (2016). Whole genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.*, **33**(10):2576–2592. doi:10.1093/molbev/msw129

Zhang, L., Wang, Y., Wu, G., Rao, L., Wei, Y., Yue, H., Yuan, T., Yang, P., Xiong, F., Zhang, S., Zhou, Q., Chen, Z., Li, J., Mo, B-W., Zhang, H., Xiong, W. & Wang, C-Y. (2020). Blockade of JAK2 protects mice against hypoxia-induced pulmonary arterial hypertension by repressing pulmonary arterial smooth muscle cell proliferation. *Cell Prolif.*, **53**(2):e12742. doi:10.1111/cpr.12742

Zhou, Z., Fan, Y., Wang, G., Lai, Z., Gao, Y., Wu, F., Lei, C., Dang, R. (2020a). Detection of Selection Signatures Underlying Production and Adaptive Traits Based on Whole-Genome Sequencing of Six Donkey Populations. *Animals*, **10**(10):1823. doi:10.3390/ani10101823

Zhou, C., Wang, Y., He, L., Zhu, J., Li, J., Tang, Y., Zhou, H., He, J. & Wu, H. (2020b). Association between NER pathway gene polymorphisms and neuroblastoma risk in an eastern Chinese population. *Mol. Ther. Oncolytics*, **20**:3-11. doi:10.1016/j.omto.2020.12.004

3.7. Supplementary Material Chapter 3 digital files

3.7.1. Appendix 3 (supplementaryChapter3.pdf)

Including Table S3.1 and Figures S3.1-S3.21.

3.7.2. S.Tables S3.2-S3.9

Raw data tables for all XP-EHH and Rsb testing of outlier vs core population comparisons.

4. Temporal trends in allele frequencies of potential climate adaptation genes in the UK great tit museum record

4.1. Abstract

Whole-genome sequencing was performed on DNA extracted from 95 great tit museum specimens, that were originally field collected in the UK since the 1870s. I formally tested whether signatures of selection for climate adaptation were identified in the same genomic locations as identified in Chapters 2 and 3, using this independent temporal dataset of UK great tit specimens. Temporal trends in allele frequencies were calculated by linear regression of allele counts over time for each SNP. The distribution of the regression coefficients was compared between candidate regions and the rest of the genome; the rest of the genome can be considered a null distribution for what might be expected by chance under a scenario of no adaptation to climate. Three SNPs out of 11 potential climate adaptation regions showed temporal trends that were greater than expected by chance ($P < 0.05$). However not all of the changes in allele frequency were in the expected direction. The closest genes to these three SNPs are the best climate adaptation genes, because they were significant in a pan-European Genome-Environment Analysis (GEA), they had long haplotypes consistent with positive selection in the UK population, and they showed an unusually strong temporal trend in the UK over the last century.

4.2. Introduction

Sequencing the genomes of natural history museum collections of plant and animal specimens and comparing the allele frequencies to those in contemporary samples can help researchers investigate evolutionary questions related to temporally-changing climates (Holmes et al. 2016; Clark et al. 2023). Natural history museum collections are a source of 'historical' (collected in the last ~200 years) and sometimes of far older 'ancient' DNA (Raxworthy and Smith 2021).

The genetic information from preserved and stored bird specimens is often extracted from skin (Irestedt et al. 2022) or toe pads (Mundy et al. 1997). DNA degradation, caused by exposure to moisture or storage chemicals, may occur continuously over storage time, and DNA damage may occur in the extraction process (Fong et al 2023). Challenges of sequencing the extracted DNA include having an available reference genome, obtaining adequate sequencing coverage, removing contaminated DNA, correcting for DNA degradation, and minimising the amount and impact of missing data (Raxworthy and Smith 2021).

Despite being challenging to work with in genomics studies, museum collections may be utilised as a time-series genetic resource for a focal species (Holmes et al. 2016). For example, the Mauritius kestrel (*Falco punctatus*) underwent a dramatic population bottleneck, associated with human activities including deforestation and habitat loss, pesticide use and the introduction of non-native species. As a result, there was only a single endangered breeding pair in 1974. An intensive conservation project 1983-1994 released 331 captive-reared kestrels into the wild population, raised from wild caught or captive-bred eggs with controlled provision and reduced predation risk (Jones et al. 1995). The high levels of genetic variation in historical Mauritius kestrels (as determined from 12 microsatellite markers of museum skins), informed researchers how much genetic variation was present in the pre-bottleneck population, and showed them it was not necessary to add genetic diversity from individuals from donor populations (i.e. assisted

gene flow). Following conservation efforts, a growing population of 200 pairs was reported around the Millennium (Groombridge et al. 2000).

The comparison of genetic changes over century-long timescales in museum collections (Lalueza-Fox 2022), allows for investigation of evolutionary adaptation in response to climate changes (Raxworthy and Smith 2021). For example, Bi et al. (2019) demonstrated that adaptive responses (in this case, selection on the gene *ALOX15*) were detectable in an alpine chipmunk species (*Tamias alpinus*), with DNA extracted from museum specimens collected in 1911 that were compared to 2012 samples. However, the occurrence and distribution of a species across time may introduce bias into which individual specimens get identified and collected for museum storage (Fong et al 2023). The availability of comparable museum specimens from different timepoints will subsequently influence the monitoring of genetic variation in a population over time (Holmes et al. 2016; Raxworthy and Smith 2021; Fong et al 2023).

In a recent example of historical DNA being used to investigate genome-wide shifts in genetic structure and climate adaptation, Turbek et al. (2023) first performed a comparative genomic study of 17 museum specimens of willow flycatchers (*Empidonax traillii*) (collected 1888-1909 from San Diego, USA) and 18 contemporary individuals (collected 2010-2011). Gene flow over the last century was demonstrated, occurring from neighbouring populations into the San Diego population. The increased genetic diversity in the contemporary San Diego population is suggested to have enhanced its adaptive potential to future challenges. Second, a total of 200 contemporary willow flycatchers from across the species range (Southwest USA) were sequenced, and regions of the genome with allele frequencies correlated with dew point temperature (associated with air moisture) and precipitation variables were identified as genetic loci likely involved in climate adaptation. Third, the authors found a decrease in genetic diversity at the genes linked to climate adaptation (particularly rain and air moisture variables) between the historical and contemporary San Diego samples. The authors demonstrated that the contemporary San Diego population showed evidence of adaptation (and potential for further adaptation) to wetter and more humid climate conditions (Turbek et al. 2023).

We set out to take this temporal approach to study adaptive evolution in response to changing climate conditions using great tit (*Parus major*) specimens from UK natural history collections dating back to the ~1870s. The great tit is an ecological model species with an assembled and annotated genome available (Laine et al. 2016). For a case study of the long-term monitoring of individuals in the wild at Wytham Woods in Oxford, UK, see Sheldon et al. (2022). UK specimens of great tits are well-suited for investigation of adaptive evolutionary studies because the species is predominantly resident in the UK and should be locally adapted to environmental conditions, with only a few long-distance migrant individuals arriving from differently adapted populations in Europe each year (Kvist et al. 1999). The ability to extract DNA from bird specimens stored in UK museums (Mundy et al. 1997), has allowed us to whole-genome sequence 95 great tit individuals (at ~12x coverage), collected at different time-points over the last century in the UK.

In Chapter 2 I described the methods used for genome-environment analysis, to identify SNPs that are associated with adaptation to climate across 20 European great tit populations. I reported the closest gene for 179, 275 and 204 SNPs associated with climate variable principal components PC1, PC2 and PC3 respectively, at a threshold of Bayes Factor (BF) >10 ('good evidence' of an association). PC1 largely described year-round temperature, PCs 2 and 3 largely described precipitation levels in the warmer months and wettest months respectively. Briefly, the data suggest that there has been adaptation to the wide range of climates across Europe, and that this adaptation involves a large set of genes with different gene ontologies, although some ontologies are over-represented among adaptation loci. Adaptation to climate appears to be complex and polygenic. In Chapter 3 the populations in which selection was happening at these candidate climate adaptation genes were identified (determined by either or both of the haplotype homozygosity statistics XP-EHH and Rsb exceeding a threshold of 4).

The aim of this chapter is to test the prediction that the SNPs of interest for climate adaptation in the UK have undergone greater changes in allele frequency over the last 150 years, compared to the rest of the genome; i.e. that they exceed expectations from genetic

drift. I calculated temporal trends in allele frequencies at all of the sequenced SNPs, and made comparisons between candidate regions and the rest of the genome; the rest of the genome can be considered a null distribution of what might be expected by chance under a scenario of drift with no adaptive evolution.

4.3. Methods

Among the 28 SNPs of interest (from 12 genomic regions) for the UK identified from extended haplotype statistics (see **Appendix 3 S.Tables S3.2, S3.5 and S3.7**), SNPs which were transversions (i.e. G/C or A/T) were removed, because it was impossible to know whether the museum sequencing data is reading the same strand as the SNP chip, and therefore it is not possible to know whether the allele identified as being under selection in the UK is increasing or decreasing in frequency in the museum samples. This list was further reduced, to include only one SNP within each distinct genomic region. The top SNP for each candidate region was chosen as the one that met the most thresholds across the three tests described in the earlier chapters (BF, XP-EHH and Rsb). If more than one SNP was significant for the same number of tests, then then priority was given to SNPs that tended to comfortably exceed thresholds across multiple tests rather than to SNPs with very extreme values at just one test.

The final variant list contained 12 SNPs of interest: with 4 variants associated with climate-PC1 and identified as under selection in Scotland (AX-100430005, AX-100212698, AX-100566910, AX-100626511), 3 variants associated with climate-PC2 and identified as under selection in Scotland PC2 (AX-100333849, AX-100053712, AX-100690978), and 5 variants associated with climate-PC3 and identified as under selection in Cambridge and Oxford (AX-100681217, AX-100379572, AX-100271597, AX-100386254, AX-100844877) (see **Table 3.1**). The focus on particular UK populations and not all three combined is because the Loch Lomond, Scotland population was identified as an outlier population for contributions to climate-PC1 and climate-PC2, whereas the more southerly Cambridge and Oxford populations were both identified as outliers for contributions to climate-PC3 (**see**

Appendix 3 Figs S1 and S2 for visualisation of how the outlier and core populations for each climate PC were identified on the climate PCA population bi-plots).

4.3.1. DNA sample collection

DNA was extracted using the Qiagen Dneasy kit from the toe pads of 119 great tit specimens that were caught 1868-1982. The samples were obtained from five natural history museums around the UK (Liverpool, Newcastle, Oxford, York, and Cardiff). Most samples (around 2/3 of the total) had over 100ng of extracted DNA. Metadata on each individual specimen was recorded, including sex, age, subspecies, year caught and location (**Appendix 4.2 file sampleInfo.txt**). 95 samples passed quality control to be sent for sequencing.

4.3.2. Paired-end sequencing, trimming and alignment

The 95 great tit museum samples and a negative control were paired-end sequenced at the NERC Environmental Omics Facility in the Centre for Genomic Research, University of Liverpool (the sequencing was externally funded by NEOF - project 1403 (**Appendix 4.3 project proposal NEOF_application.docx**)). Libraries were first prepared, and then each batch of 50 samples (genome size ~1 Gb) were sequenced on a lane of the NovaSeq 6000 S4 (2x150 bp) to achieve approximately 12x coverage per sample. Forward (R1) and reverse (R2) paired-end reads were produced, along with unpaired (R0) reads produced as the result of removal of one of the pair due to poor sequence quality or the presence of an adaptor. The raw Fastq files were trimmed in two steps: First, the Illumina adapter sequences were removed using Cutadapt version 1.2.1 (Martin 2011), with the -O 3 filter specified, so the 3' end of any reads which matched the adapter sequence for 3 bp or more were trimmed. Second, the reads were trimmed using Sickle version 1.200 (Joshi and Fass 2011) and filtered for a minimum window quality score of 20. Reads shorter than 15 bp after trimming were removed (and if only one of a read pair passed this filter, it was included in the R0 file).

The summary statistics were generated using fastq-stats from EAUtils (Aronesty 2011, 2013), which reported that 94 of the trimmed samples had an overall read length >50M, with the only exception being sample 88_GT104. The total number of reads for each trimmed sample (except 88_GT104) ranged from 95M to 197M (sample coverage approx. 10-20x), with most of the samples achieving coverage >12x. The trimmed data were made available to download in fastq.gz format, such that they could be read directly for read alignment and filtering. All of the downstream data processing steps were performed on the high-performance computer (HPC) cluster ShARC, at the University of Sheffield.

Trimmed reads for each individual museum sample were aligned to the great tit reference genome (https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_001522545.3/) (Laine et al. 2016) in fasta format, using bwa mem version 0.7.17 (Li & Durbin 2009). Forward paired-end sequences (R1) for lane 1 and lane 2 were combined and mapped, reverse paired-end sequences (R2) for lane 1 and lane 2 were combined and mapped. The unpaired reads (R0) of both lanes were concatenated and mapped. SAMtools (Danecek et al. 2021) view was used to convert these three output files (R0,R1,R2 for each sample) from .sam to .bam format; they were individually sorted, and the three files per sample were merged and indexed, to produce a single binary alignment map (.bam) file for each of the museum samples.

4.3.3. Genotype Calling

Genotype calling was performed on the 95 great tit museum samples using ANGSD software v0.940 (Korneliussen et al. 2014). ANGSD has been recommended for low coverage (12x can be regarded as low coverage) sequencing, such as is typically obtained from museum samples (Korneliussen et al. 2014). See Olkkonen & Löytynoja (2023) for a published protocol using ANGSD on low-coverage sequencing data in an endangered seal with low genetic diversity. Genotype calling was run separately for each chromosome, with the following ANGSD filtering options: output into PLINK tfam/tped files (-doPlink 2); calculate the genotype likelihood using the samtools method (-GL 1); allele frequency threshold

>0.05 (-minMaf 0.05); infer major & minor alleles from the genotype likelihoods (-doMajorMinor 1); *P* value (used for calling SNPs, where the null hypothesis is the site is invariant) less than 1e-6 (-SNP_pval 1e-6); frequency -fixed major, unknown minor (-doMaf 2); remove reads with mapping quality scores below 30 (-minMapQ 30); remove bases with a quality score <19 (-minQ 20); remove if < 20 individuals have genotype information (-minInd 20); write the major and minor alleles, and write the called genotype directly: e.g. AA,AC etc (-doGeno 5); use the allele frequency as prior (-doPost 1); and use a minimum genotype posterior probability cutoff (-postCutoff 0.95).

The ANGSD output data files for all of the individual chromosomes were converted into standard PLINK format files (bim, .bed, .fam) and merged together into a file containing 11,033,725 genome-wide SNP variants, using PLINK version 1.9 (Chang et al. 2015). These approx. 11M SNPs had a genotyping rate of 0.87 for the 95 UK museum great tit samples. Among the called SNPs, 450,794 variants were on the SNP chip used in the previous chapters. This subset of SNPs had a 0.84 genotyping rate among the 95 museum birds. The sample location was classified regionally as having been collected in Channel Islands, Isle of Man, Midlands, North England, Scotland, South England, Wales or unknown location. The samples were further divided into four time periods: 1868-1909, 1910-1929, 1930-1949, and 1950-1990, with a minimum of 18 samples in each time-bin.

The paired-end sequencing report for the 95 UK museum samples of great tit specimens provides the total number of trimmed reads per sample (**Appendix 4.4 file sequencing_report.pdf**). Quality control (QC) of the sample data against year of collection was investigated, with sequencing depth, SNP call rate, and rate of heterozygosity (which might be a proxy for whether there is allelic dropout) used as measures of quality (**Figure 4.1**). Sequencing depth was calculated as the total number of paired-end reads (i.e. R0 + R1 + R2) * 150bp / total genome size (1,020,293,992 bp). In addition, the relationship between heterozygosity and sequence depth was plotted (**Figure 4.2**). Seven great tit museum samples with elevated heterozygosity rates (over 0.5) were identified (Samples: 3, 6, 35, 39, 63, 64, 88) and removed from the data set.

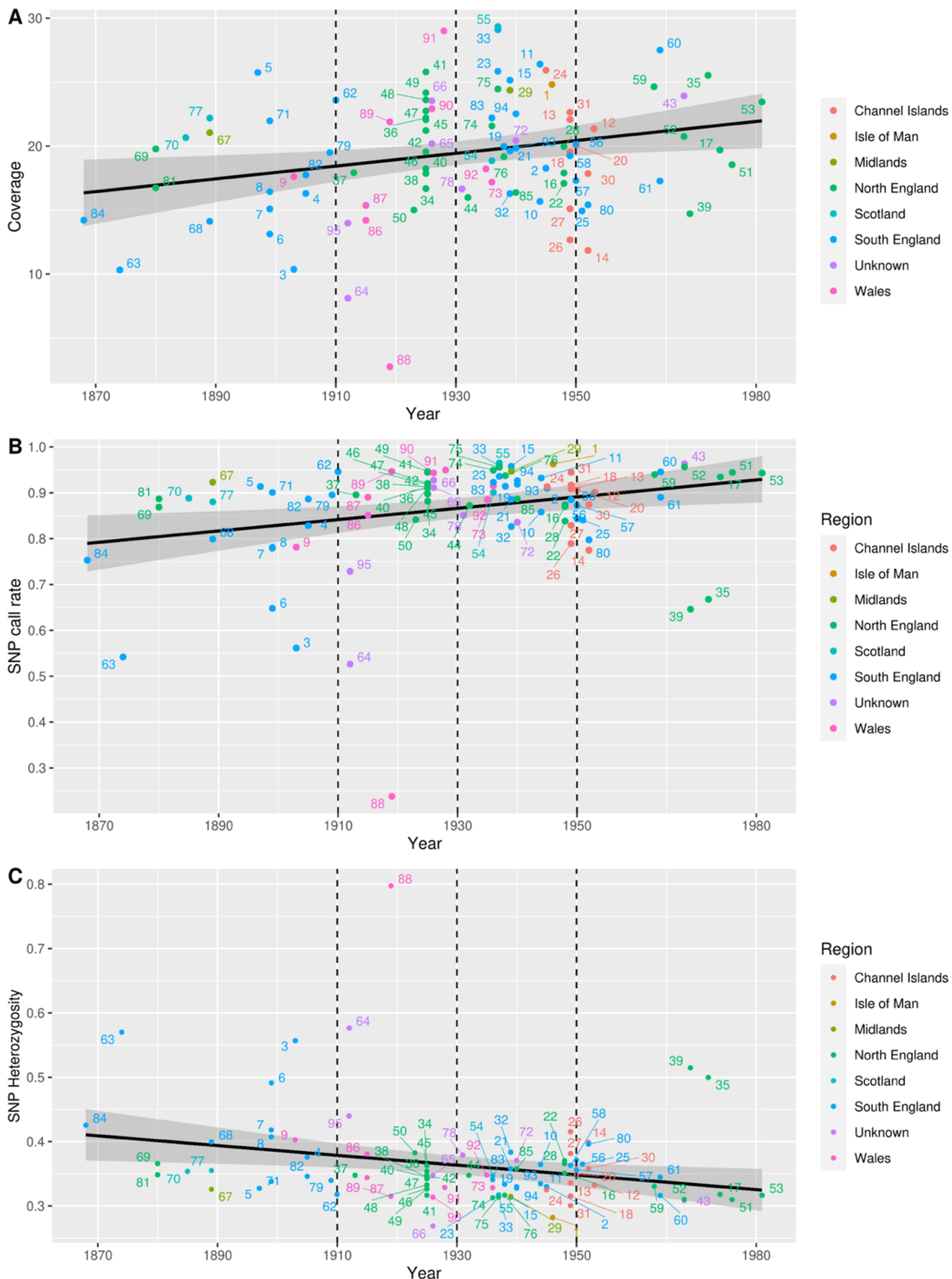


Figure 4.1: Quality control of the 95 great tit museum specimens across the time-series (with the four time-periods indicated by vertical dashed lines), and with the location of field collection in the UK and Isle of Man displayed. Sample ID codes are available in **Appendix 4.2 file sampleInfo.txt**. (A) Sequence Depth ('coverage'). (B) SNP call rate. (C) Observed SNP heterozygosity. High heterozygosity may be a proxy for whether there is insufficient read depth of one allele i.e. allelic dropout.

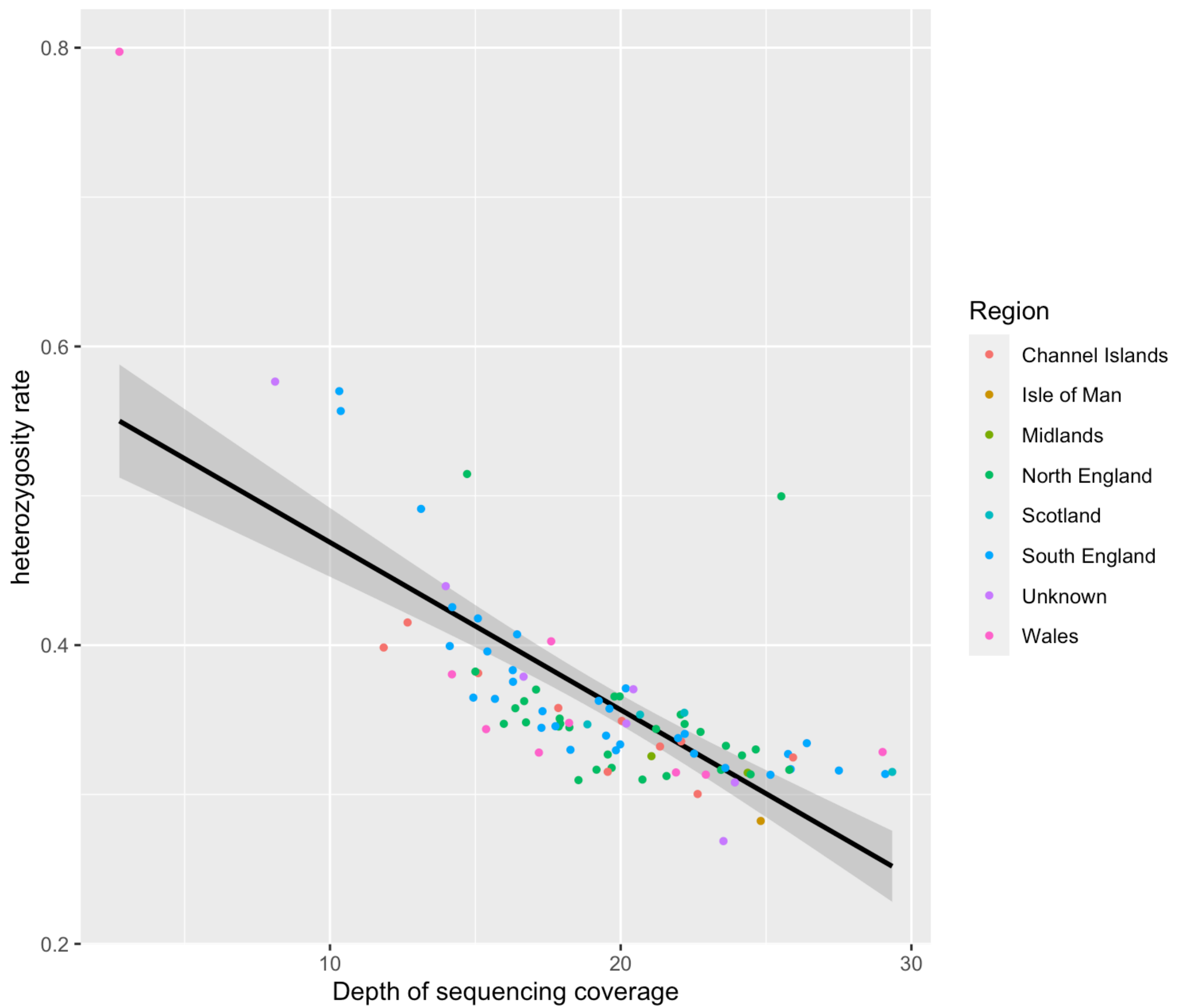


Figure 4.2: Sequence depth against heterozygosity for the 95 great tit museum samples. For quality control, the seven samples shown with heterozygosity around 0.5 or above were removed from the subsequent data analysis.

4.3.4. Investigate Z-linked markers erroneously mapped to an autosome

The sex of each sample was identified to look at heterozygosity separately in males and females and tests of Hardy-Weinberg equilibrium were run on each SNP. This test was performed because an initial principal component analysis of the autosomal data suggested discrete clustering by sex. We suspected that some sex-linked regions may have been erroneously placed on autosomes in the great tit assembly, or perhaps some sex-linked reads are being aligned to autosomes because of high sequence similarity. The great tit genome assembly does not include the W chromosome because the reference bird was a male. If W-chromosome reads are being aligned to a highly-similar autosomal region then any minor sequence differences between the autosomal and W-linked regions will tend to get falsely called as SNPs, with genotypes always being heterozygous in females and homozygous in males. Nearly every SNP where the observed heterozygosity in females is much lower than the expected heterozygosity is on the Z chromosome (NC_031799.1) as expected for a Z-linked SNP. We found ~286,00 SNPs with greater heterozygosity in females and ~127,000 SNPs with greater heterozygosity in males (calculated with between sex observed heterozygosity differences of 0.30 or more).

Further filtering of problematic Z-linked SNPs and less strict SNP calling thresholds were applied in PLINK to retain the majority of target identified SNPs. Z chromosome (NC_031799.1) markers with much higher heterozygosity in females than males (a sign they could not be Z-linked), SNPs with MAF less than 0.05, SNPs that were strongly out of HWE ($P < 1 \times 10^{-6}$), and SNPs with missing genotype call rates greater than 0.3 i.e. 30% were all removed. This filtered data set contained 8,709,994 variants for 88 samples. An additional stringent filtered data set was also produced, with the missing genotype call rates no greater than 6%, while all other parameters remained the same. However, many of the candidate SNPs were not retained for comparison.

PCA was run on the original filtered data set which contained 3,994,526 SNPs and 88 samples, with the time period and collection location visualised on the PCA plots (**Figure**

4.3). We observed the great majority of loci weighting PC2 are on chromosome 7 (NC_031776.1) with no obvious reason why (**Figure 4.3 c-d**). Binned regions over chromosome 7 were investigated in further detail, as there is suggestive evidence of a chromosomal inversion on this chromosome in great tits (da Silva et al. 2019 see SFig.1).

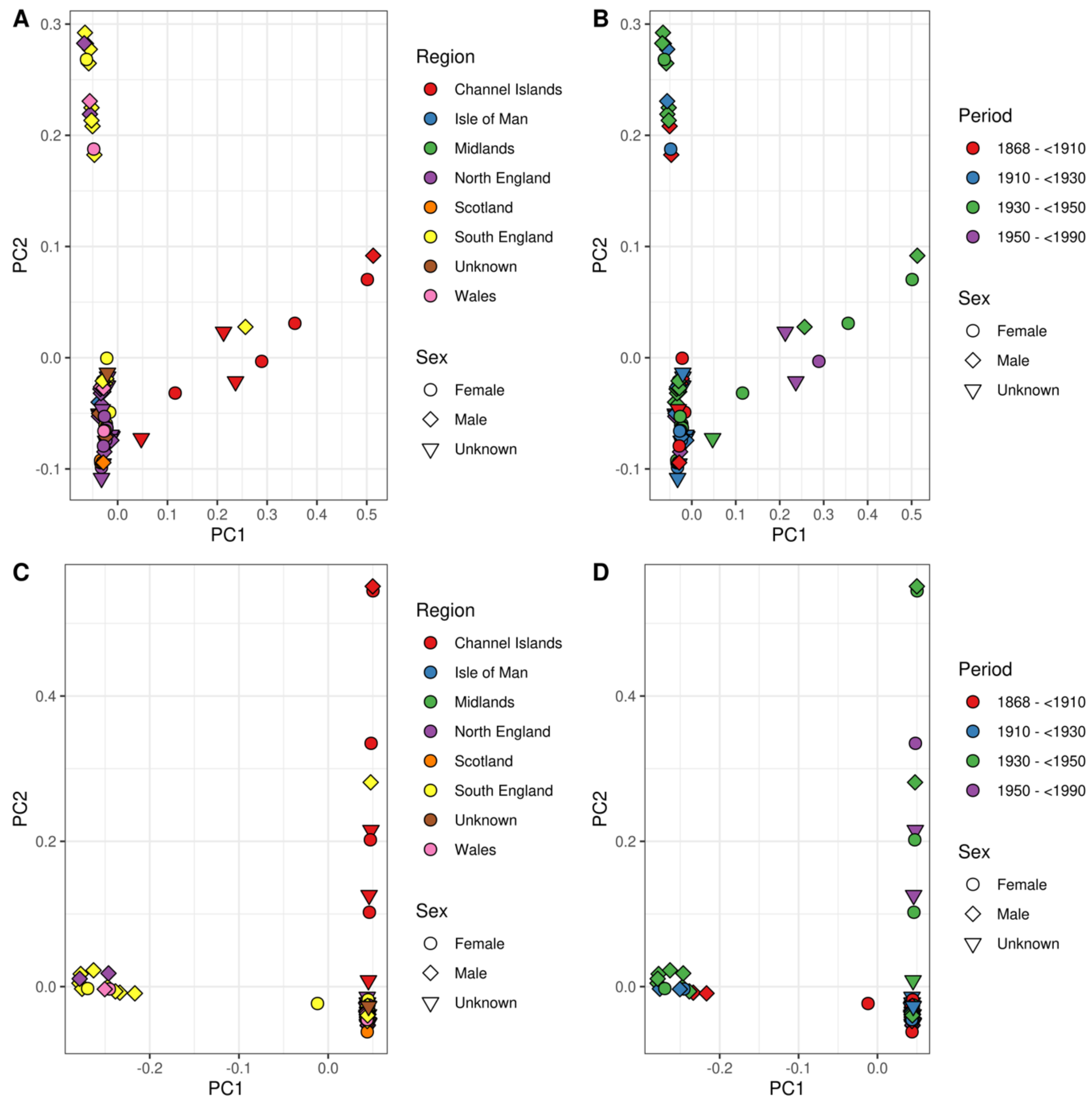


Figure 4.3: PCA of the 88 great tit museum samples was performed after filtering and removal of problematic SNPs with high differences in heterozygosity between the sexes. (A) PCA of all ~4 million SNPs, with sex of the samples and region of collection displayed. (B) PCA of all SNPs, with sex of sample and time period of sample collection displayed. (C) PCA of 175,284 SNPs on Chromosome 7 markers, with sample sex and collection region indicated. (D) PCA of only Chromosome 7 markers, with sample sex and time period of collection indicated.

An allele frequency report (.frq file) and genotype count (.frqx) report were produced for each of the filtered ~8M SNP and additional stringent filtered ~4M SNP data sets with 88 samples. The sample collection time-bins (1868-1909, 1910-1929, 1930-1949 and 1950-1990) contained 15, 22, 35, 16 individuals respectively, and a stratified allele frequency (.frq.strat) report file was produced, whereby the allele frequencies at each SNP in each different time period were recorded.

Temporal trends in allele frequencies were calculated in R (version 4.2.1), by performing a linear regression of allele counts (weighted by the number of scored genotypes per bin) over time for each SNP in the filtered SNP data set and later processed for the stringent filtered data set. The overall distribution of SNP regression coefficients is considered the null distribution of what might be expected by chance under a scenario of no adaptation to climate. To determine the significance of each candidate SNP, relative to the overall set of SNPs, we first converted each regression coefficient to an absolute value, to take into account that ~half of the SNPs will have negative coefficients and ~half will have positive coefficients. Absolute regression coefficients were ranked largest to smallest, and an empirical P-value was generated for each candidate SNP by dividing the ranked position by the total number of SNPs. A Bonferroni correction for multiple tests threshold was applied to the P-value, to take into account that when testing large SNP datasets, it is certain that some SNPs will be significant ($P < 0.05$), just by chance.

If one of the 12 SNPs of interest was filtered out or non-significant, but an equivalent high scoring (e.g. for BF, XP-EHH or Rsb) SNP was identified in exactly the same gene region and was tested to be significant, this SNP would provide a replacement: this occurred in two cases, for SNPs of interest associated with climate-PC1 AX-100879265 and climate-PC2 AX-100459612, replaced with SNPs AX-100212698 and AX-100333849, respectively (see **Table 4.1** for full details of the updated 12 SNPs of interest). Furthermore, a binomial test was run to investigate if more candidate SNPs were significant at $P < 0.05$ than expected by chance.

The significant SNPs were interpreted on the basis that the SNP chip allele on the extended haplotype identified under selection in a UK population (Chapter 3) was the one that was expected to have increased in frequency in the museum samples. However, for those SNPs that are transversions and where the two alleles are complementary (i.e. G/C or A/T SNPs) it is impossible to know whether alleles on the same strand or opposite strands are being compared between the two datasets. For example, at a G/C SNP, if allele G was the SNP chip allele on the haplotype under selection, then G would be expected to be increasing in frequency in the museum samples. However, because we didn't know which strand was typed on the SNP chip, then we cannot be sure which allele in the museum sequencing data (G or C) is equivalent to allele G on the chip. This problem is not encountered with transition SNPs (G/A or C/T) or non-complementary transversion SNPs (A/C and G/T), because the SNP chip and Illumina sequence alleles can always be matched unambiguously. Fortunately, transitions are much more common than transversions.

4.4. Results

For the 95 great tit museum samples, we demonstrated that as the depth of sequence coverage increased, the observed heterozygosity decreased (**Figure 4.2**). Some samples had impossibly high heterozygosity values (e.g. one sample had a heterozygosity of almost 0.8). The observed relationship suggests that the genotype calling algorithm in ANGSD is reluctant to call a site as homozygous if coverage is low, and is therefore biased towards calling heterozygous genotypes. This problem may be removed if genotype probabilities had been called rather than the most likely genotypes.

The plots for regions 10-20Mbp and 20-30Mbp show birds from the Channel Islands as distinct from the rest, which is broadly the trend seen genomewide (**Figure 4.3c**). PCA over binned regions of chromosome 7 shows clustering of the markers weighting PC2 in the first 10mbp, suggestive of an inversion (**Figure 4.4**). In addition, PC1 explains ~22% of the variation in the first region on chromosome 7, but for the subsequent regions PC1 only explains 2-3% of the variation. This is consistent with the SNPs in the first region being in

much higher LD with each other than is the case for the second and third regions, again consistent with an inversion.

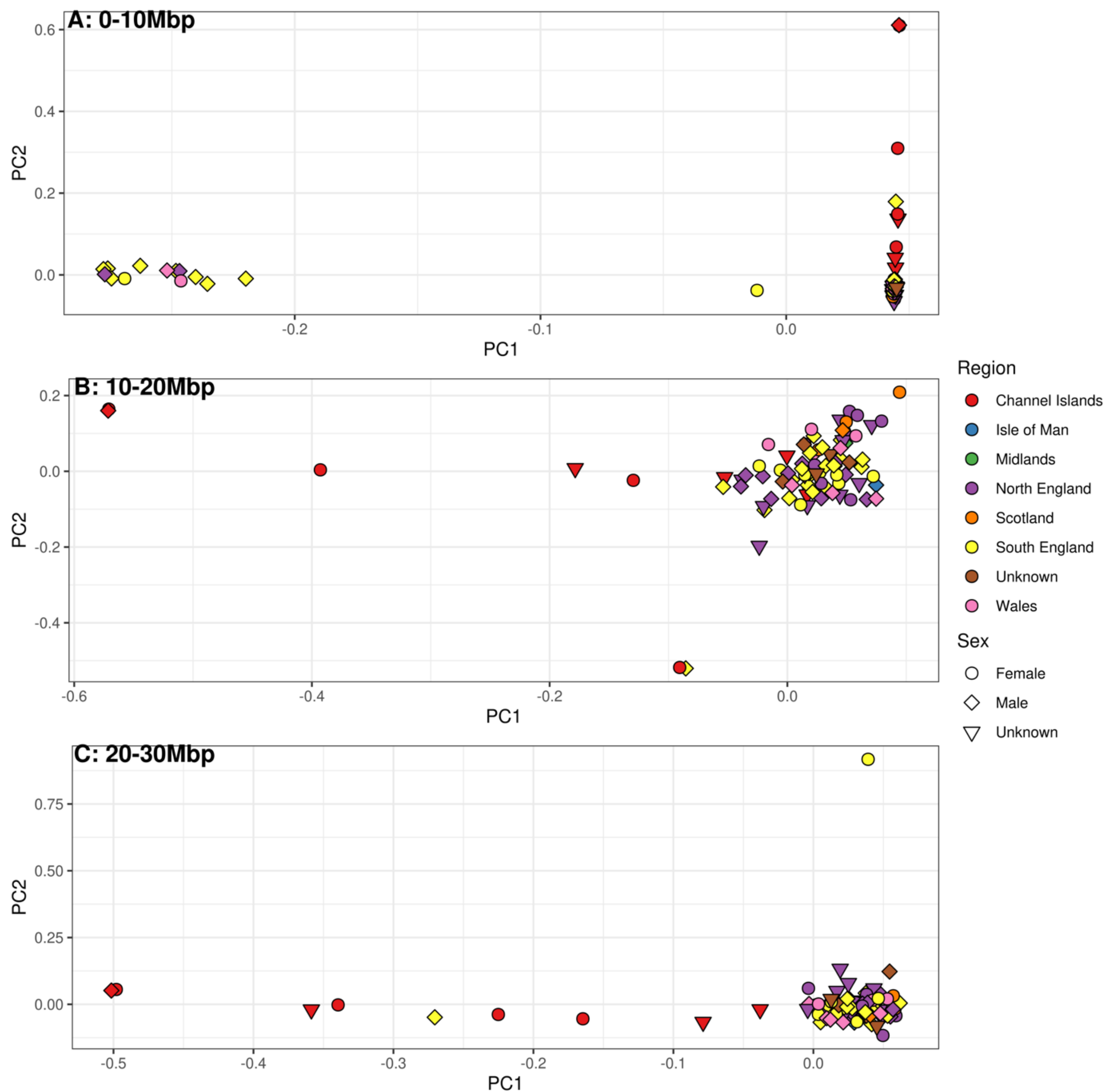


Figure 4.4: PCA of 175,284 SNPs on only chromosome 7, for regions 0-10Mbp, 10-20Mbp and 20-30Mbp. The sex and collection region for the 88 great tit museum samples are indicated.

The distribution of the regression co-efficient for all ~8 M genotyped called filtered SNPs i.e. a plot of the null distribution, is displayed in **Figure 4.5**. Three of the 12 SNPs of interest (each for a unique gene region) produced significant empirical P-values ($P < 0.05$), with two of the SNP's closest genes identified within 28kbp of the SNP (**Table 4.1**). None of the three significant SNPs (AX-100879265, AX-100459612, AX-100681217) were individually significant

at the Bonferroni-corrected significance threshold of P <0.05/12 (P < 0.0042). However, a Binomial test on a total of three significant candidate SNPs out of 12 unique tested regions, found that more SNPs were significant at P < 0.05 than expected by chance (P = 0.0196). The later analysis of the stringent filtered SNP data set only included three out of the 12 SNP regions of interest, so the results will not be discussed in detail. However, two SNPs had significant ranked regression co-efficients in the stringent filtered data set (although they did not show the expected directionality in their allele frequency trends). SNP AX-100681217 was co-identified as showing a significant trend in both the strict and more relaxed datasets, with the allele frequency changes in the expected direction (**Table 4.1**).

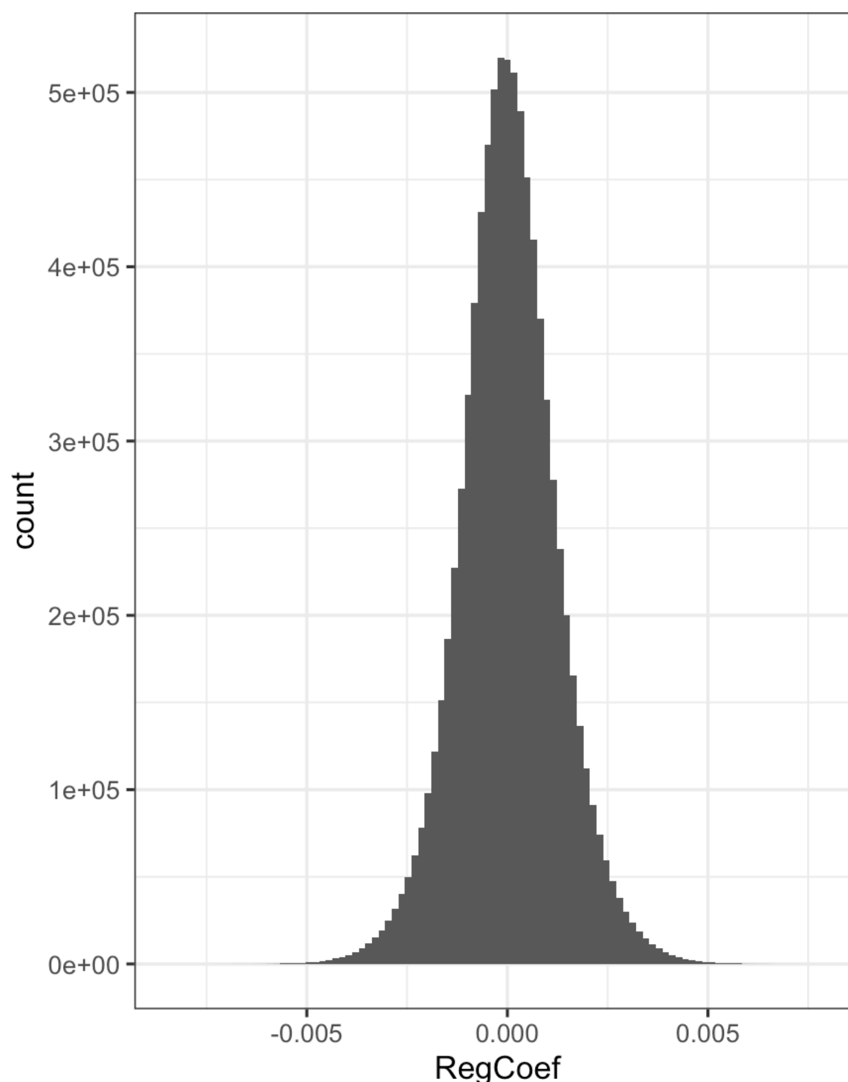


Figure 4.5: Histogram of the 'null' distribution of regression co-efficient values for all ~8 million genotype called and filtered SNPs, calculated across the 88 great tit museum samples.

Table 4.1: The set of 12 SNPs of interest for climate adaptation evolution, previously identified as associated with climate-PCs and as under selection in UK populations. Each SNP is represented by a row in the table, with information on SNP identification (ID), chromosome number and base position. Mean Bayes Factor refers to Chapter 2, with values > 10 indicating 'good' evidence, > 15 'very good' evidence and >20 'decisive' evidence for association with an indicated climate-PC. From Chapter 3, the UK 'outlier' population allele inferred to be under selection from XP-EHH and Rsb statistics, and the core population non-selected allele are indicated. For the temporal trends in allele frequencies among the museum samples, each SNP's calculated regression coefficient and empirical *P* value are reported. For significant SNPs, the allele frequencies in each time bin are reported. The final column indicates the closest gene(s) to each SNP.

SNP ID (SNPchip)	Genome ID (Chr_BP)	Chr	Mean BF†	Outlier popn(s)	XP-EHH‡	Rsb‡	Outlier / Core alleles	Museum allele freqs (Periods 1-4)	Reg. coef. (x 10^{-3})	*P* value / stringent *P* Val	Closest Gene (kbp)
AX-100430005	NC_031769.1_99275394	2	19.4 (PC1)	Scotland	**4.3**	2.1	T / C	T (0.60, 0.55, 0.52, 0.79)	-2.5	0.05/ **0.03**	GNAL (11.5)
AX-100879265	NC_031770.1_26910322	3	16.1 (PC1)	Scotland	**9.1**	**6.3**	C / T	G (0.53, 0.50, 0.29, 0.35)	-3.0	**0.02**	SOCS5 (213.5) LRFN2 (240.7)
AX-100566910	NC_031771.1_45986689	4	16.8 (PC1)	Scotland	**5.8**	**5.5**	G /T		-1.2	0.29	NSUN7 (0) APBB2 (12.8) RBM47 (47.5)
AX-100626511	NC_031773.1_32286754	30	17.9 (PC1)	Scotland	**4.1**	**5.6**	G / T		-1.1	0.35	LLPH (12.6) HMGA2 (47.2)
AX-100459612	NC_031770.1_25920756	3	12.4 (PC2)	Scotland	**5.7**	**5.1**	G / A	C (0.65, 0.62, 0.36, 0.46)	-3.1	**0.02**	SRBD1 (0)
AX-100053712	NC_031770.1_26617245	3	16.4 (PC2)	Scotland	**6.1**	**4.3**	A / G		0.5	0.66	CRIPT (16.0) PIGF (23.8) RHOQ (42.8)
AX-100690978	NC_031774.1_21629221	5	**20.1** (PC2)	Scotland	**5.1**	**4.7**	G / T		2.2	0.08	CD82 (73.4) ALX4 (82.8)
AX-100681217	NC_031769.1_46317558	2	13.1 (PC3)	Cambridge & Oxford	**7.5**	**10.4**	C / A	G (0.33, 0.52, 0.56, 0.59)	-2.6	**0.04/ 0.02**	HERPUD2 (28.0)
AX-100379572	NC_031769.1_129919204	2	11.3 (PC3)	Cambridge & Oxford	3.4	**5.6**	C / T		-0.2	0.83	VPS13B (0)
AX-100271597 *AX-100293521	NC_031769.1_136082831 *NC_031769.1_136162421	2 *2	**49.0** (PC3) ***21.96**	Cambridge & Oxford	**7.5** ***4.4**	**12.1** ***9.5**	C / T *A/G	*T (0.50, 0.43, 0.44, 0.29)	-1.6 *-2.4	0.18 / * ***0.03**	TRPS1 (0) * TRPS1 (0)
AX-100386254	NC_031770.1_38558688	3	**21.1** (PC3)	Cambridge & Oxford	**4.8**	**5.0**	C / T		0.9	0.40	PDE10A (85.9)
AX-100844877	NC_031772.1_12152707	4A	13.9 (PC3)	Cambridge & Oxford	**4.3**	**6.0**	T / C		1.8	0.13	ACSL4 (15.0) COL4A5 (158.4)

†From GEA in Chapter 2, ‡From REHH analysis in Chapter 3, * An equivalent nearby SNP of interest in the gene region with a significant ranked *P* value.

4.5. Discussion

The discussion will focus on the filtered SNP data set, to include all of the twelve candidate SNP regions of interest. We found the first 10Mbp region on chromosome 7 is suggestive of an inversion, consistent with Da Silva et al. 2019 investigation of inversion polymorphisms in great tits. We identified three SNPs with significant temporal trends in allele frequencies for UK great tit museum samples. The finding of this set of three SNPs out of twelve potential climate adaptation regions was more than expected by chance ($P = 0.02$). However, two SNPs had allele frequency changes that were not in the expected direction, and no individual SNP was found to be significant after correcting for multiple testing.

The SNP AX-100681217 (on chromosome 2) had allele frequency changes that were in the expected direction, and this SNP was also co-identified as significant in the stringent filtered dataset. An increase in allele G across the museum time-series, was consistent with signatures of selection in the Cambridge and Oxford populations and was associated with climate-PC3. The closest gene to this SNP, *HERPUD2* (*Homocysteine inducible ER protein with ubiquitin-like domain 2*) has therefore, potentially been involved in climate adaptation. *HERPUD2* is expressed in skeletal muscle cells, with the primary function of skeletal muscles in birds being flight and thermogenesis e.g. shivering (Swanson et al. 2022).

Very strong genetic differentiation has been observed at the *HERPUD2* gene (with signatures of selection detected in a further 11 loci) between super-pools of autumn and spring spawning herring from the warmer Baltic Sea and colder Atlantic Ocean (Han et al. 2020). Ecological adaptation at candidate loci is not driven by differences in salinity. However, it is not discernible whether genetic adaptation is driven by water temperature variation or a temperature-correlated environmental variable (Andersson et al. 2023).

In contrast, significant SNPs AX-100879265 and AX-100459612 associated with climate PC1 and PC2 respectively, did not show the expected selected allele frequency trends in the museum genomes, with an overall decrease in the identified allele under selection in the UK population up to the 1950s, although there was a slight increase from 1950 to 1982.

Notably, these two significant SNPs are located within 1 Mbp of each other on Chromosome 3. Firstly, the SNP AX-100879265, consistent with selection in the Scotland population and associated with climate-PC1, is located between, but not very close to, the two nearest genes *SOCS5* (~213 kbp) and *LRFN2* (~241 kbp). The gene *SOCS5* (*suppressor of cytokine signaling 5*), when deficient in mice, increases susceptibility to infection by the influenza virus and heightens illness severity (Kedzierski et al. 2017). The gene *LRFN2* (*leucine rich repeat and fibronectin type III domain containing 2*), when mutated in mice, shows suppressed synaptic activity in the brain, affecting working memory and startle responses (Li et al. 2018). Neither seems like a strong candidate for climate adaptation.

The second SNP AX-100459612 also did not shown the expected allele frequency trend. This SNP, which is consistent with selection in the Scotland population, and is associated with climate-PC2, is located within the gene *SRBD1* (*S1 RNA Binding Domain 1*). Mutations in *SRBD1*, a susceptibility gene for glaucoma eye disease, may cause vision loss (Writing Committee for the Normal Tension Glaucoma Genetic Study Group of Japan Glaucoma Society 2010), which would result in changes in behaviour and fitness of wild birds. However, there is no obvious reason why positive selection at this gene might occur in response to climate change.

It is important to consider that the closest genes to these three significant SNPs are the best candidate climate adaptation genes identified, because they emerged from three stringent data analyses: 1) they were significant in a pan-European genome-climate association analysis (see GEA **Chapter 2**); 2) they had long haplotypes in tests for selection consistent with positive selection in at least one UK population (see **Appendix 4 Fig. S4.1, S4.3 and S4.5** for plots of extended haplotype homozygosity around each of the three focal SNPs); 3) they showed an unusually strong temporal trend in the UK over the last century (from allele frequency temporal trends detected in the museum record (described in this chapter). However, further inspection of plots of extended haplotype lengths and haplotype bifurcation of the ancestral and derived alleles around each of the three candidate SNPs revealed long-distance haplotypes in the core populations, rather than the outlier UK populations (see **Fig. S4.2, S4.4 and S4.6**).

4.5.1. Further Investigation of temporal trends at candidate adaptation genes

To further investigate these trends in allele frequencies at these SNPs, it would be important to further sample time-points 1980-2010, to compare with the HapMap samples (approx. 2010-2015) and contemporary population sequencing (post-2020), to aid drawing conclusions over longer time periods and for more individuals and more populations. One important consideration is the only significant SNP detected with allele frequency changes in the expected direction over the time-series was identified in the England population. Most of the museum sequenced birds were field collected in England, and could therefore be better suited to detecting temporal changes at SNP alleles where EHH tests revealed selection in the Oxford and Cambridge populations, rather than in the Scotland population. To validate SNPs that were thought to be under selection in Scotland, it may be necessary to increase the sampling effort of museum great tit specimens from across Scotland, including any collected from small Scottish islands. The inclusion of island populations in the museum record e.g. the Channel Islands and Isle of Man has helped to broaden the UK record to include more peripheral populations. Future studies could consider using great tit museum and/or contemporary samples from across Ireland, to look for genetic differentiation at climate adaptation genes between populations in the British Isles.

4.5.2. Sequencing from museum samples

We went into this work fully aware that DNA sequencing from low yield samples may influence sequencing quality for the data analysis. We applied strict bioinformatic filters (comparable to those used in other published studies) and inspected the sequenced samples for evidence of temporal and spatial bias in sequence quality. We observed high levels of heterozygosity (considerably greater than 0.5) among the samples with the lowest sequencing depth. This trend could possibly be explained by a phenomenon where, at low coverage, heterozygous genotypes are more readily called than homozygous genotypes, because of uncertainty whether a second allele is actually present but has yet to be seen.

One solution may be to use genotype probabilities rather than called genotypes (i.e. to propagate uncertainty in genotype calling through downstream analyses).

Here we demonstrate that DNA suitable for sequencing was recoverable from many UK museum bird specimens and could be used to infer allele frequency changes over the last century. Natural history collections can be used to investigate evolutionary responses to past climatic events (e.g. El Niño) and climate change including global warming (Raxworthy and Smith 2021; Lalueza-Fox 2022). Looking at the museum record over the last ~200 years is a way of validating signatures of selection that were originally identified from genome-environment association (GEA) analyses in contemporary populations. The approach of validating a GEA by looking at patterns of allele frequencies for potential climate adaptation genes in the museum record was also conducted for the historical and contemporary San Diego Willow flycatchers (Turbek et al. 2023), albeit with a focus on whether contemporary populations have experienced an increase in genetic diversity due to immigration.

4.5.3. Future directions

One future direction for working with this great tit temporal data set, relates to the fact that when the museum samples were collected, the bill lengths were also measured. The original motivation for sampling from the museum collections was to build on previous research showing the microevolution of longer bills in the UK population of great tits relative to European populations (Bosse et al. 2017). The temporal changes in allele frequencies of SNPs associated with bill length could address the suggestion that UK great tit populations have evolved longer bills as an adaptation to the relatively recent (i.e. last ~100-150 years) introduction of using bird feeders in the UK.

In summary, here we demonstrate that DNA suitable for sequencing was recoverable from museum bird samples. DNA degradation in older museum specimens may gradually accumulate over time due to damage by moisture or storage chemicals, which would

reduce the quality of the extracted DNA. Genotype calling using ANGSD and applying strict SNP calling filters is recommended for use on low coverage sequencing data from museum samples. Overall, across the bioinformatic data analyses of Chapters 2-4 I have identified three best candidate genomic regions for climate adaptation in UK great tit populations.

4.6. References

Andersson, L., André, C., Johannesson, K., & Pettersson, M. (2023). Ecological adaptation in cod and herring and possible consequences of future climate change in the Baltic Sea. *Front. Mar. Sci,*. **10**:1101855. doi: 10.3389/fmars.2023.1101855

Aronesty, E. (2011). *ea-utils* : "Command-line tools for processing biological sequencing data"; https://github.com/ExpressionAnalysis/ea-utils

Aronesty, E. (2013). *TOBioiJ* : "Comparison of Sequencing Utility Programs", doi:10.2174/1875036201307010001

Bosse ,M., Spurgin, L.G., Laine, V.N., Cole, E.F., Firth, J.A., Gienapp, P., Gosler, A.G, McMahon, K., Poissant, J., Verhagen, I., Groenen, M.A.M., van Oers, K., Sheldon, B.C., Visser, M.E., & Slate, J. (2017). Recent natural selection causes adaptive evolution of an avian polygenic trait. *Science*, **358**(6361):365-368. doi: 10.1126/science.aal3298

Bi, K., Linderoth, T., Singhal, S., Vanderpool, D., Patton, J.L., Nielsen, R., Moritz, C., & Good, J.M. (2019). Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. *PLOS Genetics*, **15**(5): e1008119. doi: 10.1371/journal.pgen.1008119

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**. doi: 10.1186/s13742-015-0047-8

Clark, R D., Catalano, K.A., Fitz, K.S., Garcia, E., Jaynes, K.E., Reid, B.N., Sawkins, A., Snead, A.A., Whalen, J.C., & Pinsky, M.L. (2023). The practice and promise of temporal genomics for measuring evolutionary responses to global change. *Mol. Ecol. Resour.*, **00**: 1–17. doi:10.1111/1755-0998.13789

Danecek, P., Bonfield, J.K., Liddle, J., Marshall, J., Ohan, V., Pollard, M.O., Whitwham, A., Keane, T., McCarthy, S.A., Davies, &R.M., Li., H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, **10**(2): giab008. doi: 10.1093/gigascience/giab008

da Silva, V.H., Laine, V.N., Bosse, M., Spurgin, L.G., Derks, M.F.L., van Oers, K., Dibbits, B., Slate, J., Crooijmans, R.P.M.A., Visser, M.E., & Groenen, M.A.M. (2019). The Genomic Complexity of a Large Inversion in Great Tits. *Genome Biol. Evol.*, **11**(7), 1870–1881. doi:/10.1093/gbe/evz106

Fong, J.J., Blom, M.P.K., Aowphol, A., McGuire, J.A., Sutcharit, C., & Soltis, P.S. (2023). Editorial: Recent advances in museomics: revolutionizing biodiversity research. *Front. Ecol. Evol.*, **11**:1188172. doi: 10.3389/fevo.2023.1188172

Groombridge, J.J., Jones, C.G., Bruford, M.W., & Nichols, R.A. (2000). 'Ghost' alleles of the Mauritius kestrel. *Nature*, **403**(616). doi: 10.1038/35001148

Han, F., Jamsandekar, M., Pettersson, M.E., Su, L., Fuentes-Pardo, A.P., Davis, B.W., Bekkevold, D., Berg, F., Casini, M., Dahle, G., Farrell, E.D., Folkvord, A., & Andersson, L. (2020). Ecological adaptation in Atlantic herring is associated with large shifts in allele frequencies at hundreds of loci. *eLife*, **9**:e61076. doi: 10.7554/eLife.61076

Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., & Nachman, M.W. (2016). Natural history collections as windows on evolutionary processes. *Mol. Ecol.*, **25**(4):864-81. doi: 10.1111/mec.13529

Irestedt, M., Thörn, F., Müller, I.A., Jønsson, K.A., Ericson, P.G., & Blom, M.P. (2022). A guide to avian museomics: Insights gained from resequencing hundreds of avian study skins. *Mol. Ecol. Resour.*, **22**(7): 2672-2684. doi: 10.1111/1755-0998.13660

Jones, C.G., Heck, W., Lewis, R.E., Mungroo, Y., Slade, G. and Cade, T. (1995). The restoration of the Mauritius Kestrel Falco punctatus population. *Ibis*, **137**:S173-S180. doi:10.1111/j.1474-919X.1995.tb08439.x

Joshi, N.A. & Fass, J.N. (2011). Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files (Version 1.33) [Software]. Available at https://github.com/najoshi/sickle.

Kedzierski, L., Tate, M.D., Hsu, A.C., Kolesnik, T.B., Linossi, E.M., Dagley, L., Dong, Z., Freeman, S., Infusini, G., Starkey, M.R., Bird, N.L., Chatfield, S.M., Babon, J.J., Huntington, N., Belz, G., Webb, A., Wark, P.A.B., Nicola, N.A., Xu, J., Kedzierska, K., Hansbro, P.M., & Nicholson, S.E. (2017). Suppressor of cytokine signaling (SOCS)5 ameliorates influenza infection via inhibition of EGFR signaling. *eLife***, 6**:e20444. doi: 10.7554/eLife.20444

Korneliussen, T.S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of Next Generation Sequencing Data. *BMC Bioinformatics*, **15**(1):356. doi: 10.1186/s12859-014-0356-4.

Kvist, L., Ruokonen, M., Lumme, J., & Orell, M. (1999). The colonization history and present-day population structure of the European great tit (Parus major major). *Heredity*, **82**(5): 495–502. doi: 10.1038/sj.hdy.6885130

Laine, V.N., Gossmann, T.I., Schachtschneider, K.M., Garroway, C.J., Madsen, O.,Verhoeven, K.J.F., de Jager, V., Megens, H.-J., Warren, W.C., Minx, P., Crooijmans, R.P.M.A., Corcoran, P., Great Tit HapMap Consortium, Sheldon, B.C., Slate, J., Zeng, K., van Oers, K., Visser, M.E., & Groenen, M.A.M. (2016). Evolutionary signals of selection on cognition from the great tit genome and methylome. *Nat. Commun.*, **7**(10474). doi: 10.1038/ncomms10474

Lalueza-Fox, C. (2022). Museomics. *Curr. Biol.*, ***32***(21): R1214-R1215. doi: 10.1016/j.cub.2022.09.019.

Lamichhaney, S., Fuentes-Pardo, A.P., Rafati, N., Ryman, N., McCracken, G.R., Bourne, C., Singh, R., Ruzzante, D.E., Andersson, L. (2017). Parallel adaptive evolution of geographically distant herring populations on both sides of the North Atlantic Ocean. *Proc. Natl. Acad. Sci.*, **114**(17):E3452–E3461. doi.org/10.1073/pnas.1617728114

Li, H. & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, **25**(14), 1754–1760. doi.org/10.1186/s13742-015-0047-8

Li, Y., Kim, R., Cho, Y.S., Song, W.S., Kim, D., Kim, K., Roh, J.D., Chung, C., Park, H., Yang, E., Kim, S-J., Ko, J., Kim, H., Kim, M-H., Bae, Y-C., Kim, E. (2018). Lrfn2-Mutant Mice Display Suppressed Synaptic Plasticity and Inhibitory Synapse Development and Abnormal Social Communication and Startle Response. *J. Neurosci.*, **38**(26):5872-5887. doi: 10.1523/JNEUROSCI.3321-17.2018

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* **17**(1): 10-12. doi: 10.14806/ej.17.1.200.

Mundy, N.I., Unitt, P., & Woodruff, D.S. (1997). Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. *The Auk,* **114**(1): 126-129. doi: 10.2307/4089075

Olkkonen, E. & Löytynoja, A. (2023). Analysis of population structure and genetic diversity in low-variance Saimaa ringed seals using low-coverage whole-genome sequence data. *STAR Protoc.*, **4**(4):102567. doi: 10.1016/j.xpro.2023.102567

Raxworthy, C.J. & Smith, B.T. (2021). Mining museums for historical DNA: advances and challenges in museomics. *Trends Ecol. Evol*., **36**(11): 1049-1060. doi: 10.1016/j.tree.2021.07.009

Sheldon, B. C., Kruuk, L. E. B., & Alberts, S. C. (2022). The expanding value of long-term studies of individuals in the wild. *Nature Ecol. Evol.*, ***6*(**12), 1799-1801. doi: 10.1038/s41559-022-01940-7

Swanson, D. L., Zhang, Y, & Jimenez, A. G. (2022). Skeletal muscle and metabolic flexibility in response to changing energy demands in wild birds. *Front Physiol.*, **22**(13):961392. doi: 10.3389/fphys.2022.961392

Turbek, S.P., Bossu, C., Rayne, C., Gruppi, C., Kus, B.E., Whitfield, M., Smith, T.B., Paxton, E.H., Bay, R.A., & Ruegg, K.C. (2023). Historical DNA reveals climate adaptation in an endangered songbird. *Nat. Clim. Chang.*, **13**: 735–741. doi: 10.1038/s41558-023-01696-3

Writing Committee for the Normal Tension Glaucoma Genetic Study Group of Japan Glaucoma Society (2010). Genome-wide association study of normal tension glaucoma: Common variants in *SRBD1* and *ELOVL5* contribute to disease susceptibility. *Ophthalmology*, **117**(7): 1331-1338.e5. doi: 10.1016/j.ophtha.2009.12.001

4.7. Appendix Chapter 4 digital files

4.7.1. Appendix 4 (supplementaryChapter4.pdf)

Including Figures S4.1-S4.6 showing signatures of selective sweeps in the UK museum record at SNPs: AX-100681217, AX-100879265 and AX-100459612.

4.7.2. Appendix 4.2 (sampleInfo.txt)

4.7.3. Appendix 4.3 (NEOF_application.docx)

4.7.4. Appendix 4.4 (sequencing report.pdf)

5. Discussion - Genomics of climate adaptation

- In Chapter 2, I performed a multiple-population GEA analysis and identified 36 candidate genes associated with adaptation to climate. The data showed that climate adaptation appears to be polygenic and genetically complex, involving many different genes and biological process pathways. The GEA analysis also suggested that substantial amounts of climate-associated genetic variation remains in the European populations. This is likely to be important for adaptation to future climate changes.

- In Chapter 3, I explored whether it was possible to identify the populations where climate adaptation at specific loci had taken place. Starting with the 36 candidate genes from the previous chapter, I identified 17 loci that had long haplotypes consistent with positive selection in single or geographically close populations. I found little evidence of signatures of parallel evolution in different populations. Instead, it seems likely that adaptation has involved different loci in different populations, perhaps reflecting the fact that each population has a unique set of climatic conditions.

- In Chapter 4, I followed up the findings from Chapters 2 and 3. Of the 17 loci identified in Chapter 3, 11 had haplotypes where the sweep might have occurred in the UK. Thus, I explored temporal trends in allele frequencies, using whole-genome sequenced great tit museum specimens from UK populations. Three candidate climate adaptation SNPs showed significant allele frequency changes over the last century. The closest genes to these three potential climate adaptation regions were *SOCS5*, *SRBD1* and *HERPUD2*, which were therefore the best climate adaptation candidate genes identified in this thesis. These loci passed significance thresholds from the stringent tests to detect selection (described in Chapters 2 and 3) and from an unusually strong temporal trend in the UK (Chapter 4).

Overall, the identification of candidate climate adaptation genes in Eurasian great tit populations (described across the thesis chapters) adds to the published putative climate adaptation genes described for vertebrates, which have typically been detected in domesticated animals such as chickens (Tian et al. 2020; Rachman et al. 2024) and cattle

(Ben-Jemaa et al. 2020; Flori et al. 2018; Saravanan et al. 2021), often with biological processes related to heat stress. Only one of the three identified candidate genes for adaptation to the UK climate displayed the allele frequency changes in the expected direction for the UK museum specimens – this was for candidate gene *HERPUD2*, which plays a biological role in skeletal muscle cells involved in flight and shivering (Swanson et al. 2022). Climate adaptation of vertebrates to the UK climate is unlikely to be for adaptation to heat-stress causing environmental conditions, so it is perhaps not surprising that the same candidate climate adaptation genes are not found in great tit and domesticated species.

5.1. Methods to validate candidate adaptive loci for climate adaptation

At this stage, it is uncertain whether the candidate genes have really been involved in climate adaptation, or how they have enabled it. Ideally, the adaptive loci would need to be validated for fitness effects under realistic conditions. Candidate genes can be validated in natural or artificially selected populations by recording the fitness of different genotypes in different conditions. Great tits could be reared in replicated field trial conditions across different field sites or in common garden experiments (described in Chapter 1) where individuals from differently adapted populations are reared in a common environment e.g. in field, aviary or laboratory controlled conditions. However, common garden experiments do not test the extent to which selected alleles are locally adapted to the home environment in which they have been under selection. Alternatively, reciprocal transplant experiments (described in Chapter 1) may be a good way to test how well individuals from a population are adapted to their home environment. However, reciprocal transplant of individuals from differently adapted populations may only be feasible on artificially selected aviary populations rather than in wild populations.

An alternative approach to observing adaptation in the wild, is to observe a population after a climatic event and investigate genomic responses and fitness differences between

genotypes at adaptive loci. Monitoring long-term wild populations, like the many decades of consecutive field work carried out by Grant and Grant on the finches of the Galapagos Islands (Grant and Grant 2003), coupled with quantitative genetics can be informative about genetic responses to the climatic conditions experienced within study populations, e.g. for the Wytham population (Charmantier et al. 2008). The ideal set up to investigate evolutionary responses to climate changes would be to monitor allele frequency changes at SNPs under selection before and after an extreme weather event (e.g. a summer heatwave), ideally in replicated natural populations. A prediction that could be tested, is that the frequency of alleles associated with adaptation to intense heat events increase after a heatwave year. Investigating if climate adaptation genes have evolved (in the predicted way) after a measured climate event(s), will aid the understanding of the genomic basis of local adaptation in response to climate changes.

5.2. Gene editing for functional validation of candidate adaptation genes

Another approach that could be used to directly test the functional role of candidate SNPs associated with adaptation is by rearing individuals that were the subject of gene editing techniques e.g. gene knockouts or gene expression differences, so that their fitness can be compared to that of wildtype individuals. Although few studies have performed functional validation of candidate adaptation genes using gene editing techniques, even on model species including *Drosophila*. See Ansai and Kitano (2022) for a review of gene editing technologies based on CRISPR and site-specific recombinases that can be used to validate candidate adaptation genes.

5.3. Investigating future genomic responses for climate adaptation

Experimental evolution (described in Chapter 1) is a method to study evolution in action across hundreds of generations. However, genomic adaptation identified in controlled

environments may not be that relevant to the situation in the wild, as species housed in experimental settings are in a completely different environment (usually with much better resources, food, warmth etc.). In an annually breeding species, like great tits, an experiment of this kind would take much longer than the duration of an individual researcher's career.

5.3.1. Genomic selection experiments to investigate adaptation to future climate changes

Genomic selection experiments are widely used for breeding of crop plants and livestock, with the aim of artificial selection for improved polygenic traits (see R2D2 Consortium et al. 2021 for a review of use in animal and plant species). Marker-based prediction of breeding values i.e. genomic selection / prediction (Meuwissen et al. 2001), considers the effects of all SNPs rather than just the ones that are deemed significant (R2D2 Consortium et al. 2021). Genomic selection methods require: 1) a reference population of genotyped and phenotyped individuals, with a statistical model built to estimate SNP effects on phenotypes and develop prediction equations. 2) that new candidates for selection are always genotyped and predicted using phenotypes where available (Meuwissen et al. 2001). Genomic estimated breeding values (GEBV) have been predicted for natural populations of mammals e.g. Soay sheep (Ashraf et al. 2022) and birds including a study of genomic selection for breeding time in a wild great tit population (Gienapp et al. 2019). This method of genomic selection has applications for testing evolutionary predictions for future scenarios including climate change, by speeding up the process of observing genomic responses across generations (Gienapp et al. 2019; R2D2 Consortium et al. 2021) or by enabling researchers to study genetic changes in the absence of pedigrees, which are required for conventional quantitative genetics.

Using the approach of a genomic selection experiment on aviary lines of great tits, Lindner et al. (2023) sped up evolution over three-years to select for early and late egg laying respectively (**Figure 5.1**). In total, 475 early laid eggs and 461 late laid eggs were translocated to the forest (De Hoge Veluwe, NL) to see how the offspring's lay dates compared to >2000 forest individuals. They found the early selected birds didn't lay their

eggs significantly earlier than the forest birds, while the later selected birds did have a significantly later lay date. The reproductive success of individuals selected for early or late lay date was no greater than for forest individuals. In the study years the egg lay date and winter moth caterpillar peak generally matched for the wild individuals. The authors of this study discuss that, at present, the real impact of climate change is being camouflaged, as the population effects are buffered by density-dependent processes i.e. if three chicks (out of ten) die before fledging because of climate change, the survival prospects of the remaining chicks increases as there is more food and less competition. However, there is expected to be a limit to this buffering of population effects, for example, when many chicks die from climate change, there is no further advantage to the remaining chicks (Lindner et al. 2023). Overall, the study demonstrated that genetic adaptation towards early lay dates in great tits (to meet earlier peak caterpillar abundance) is an extremely slow process over many generations. This experiment exemplifies how genomic selection responses can be tested for functional effects on fitness in the wild. However, it is worth recognizing that the intensive efforts required to rear and artificially select the aviary great tit populations and to perform field work to test the fitness of the selected lines in the wild, would potentially limit replication in other wild populations.

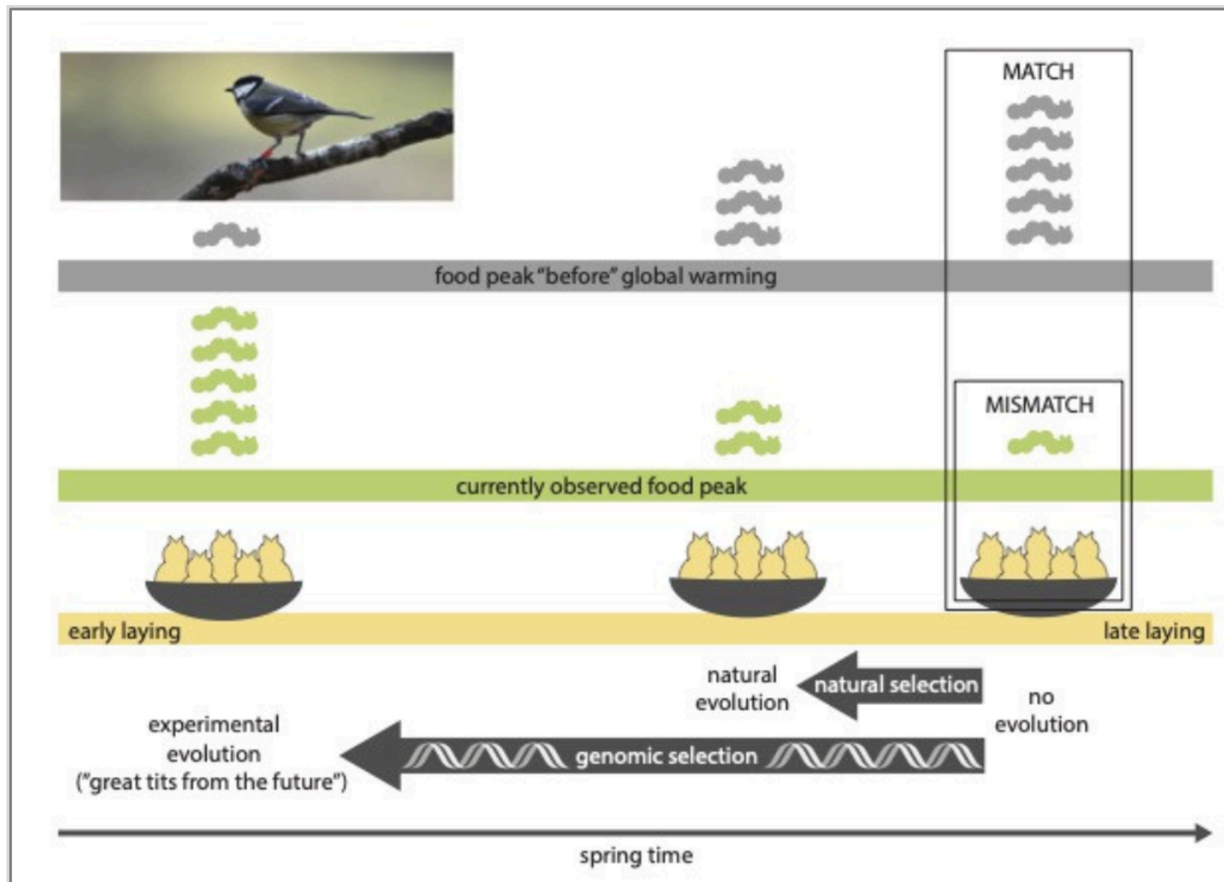


Figure 5.1: Representation of artificial genomic selection experiment in great tits by Lindner et al. 2023 to select for early lay date individuals, to match an earlier caterpillar moth peak, as predicted under global warming in the coming decade. If climate change occurs too fast, individuals won't be able to adapt sufficiently. Figure from Lindner et al. 2023.

5.3.2. Genomics of climate adaptation

The identification of climate adaptation genes in great tit populations could be aided by longer-term monitoring of the sampled wild populations, especially if fitness and climate were recorded over many years, rather than a snapshot of genomes in a given year (without fitness data). Furthermore, most of the data presented in this thesis relied on SNP chips rather than whole-genome sequencing, which has now become considerably cheaper. The combination of bigger and richer data, with ongoing developments in algorithms and methods to detect genomic responses and climate adaptation would undoubtedly improve power. Of course, most populations are not the focus of long-term studies, and maintaining

such studies can be complicated by issues such as a lack of collaboration, difficulties obtaining funding, a huge field effort, or war. Further understanding of the species and sub-species division of *Parus major* and of population connectivity, will likely be informative for population differentiation comparisons. More candidate climate adaptation genes functionally validated in vertebrates will give better insight into the comparative genomics of climate adaptation.

In this thesis the genomics of climate adaptation has been investigated in great tit populations, which have been used to consider more generally how vertebrates will genetically respond to projected climate changes. In a recent paper, Urban et al. 2024 describe how and when adaptive responses to climate changes can be predicted, considering sources of genetic variation, population sizes and generation times. It remains to be seen if adaptation in vertebrate populations can act fast enough to keep up with climate changes? Or if population tipping points will be reached?

5.4. Future research directions

Further understanding of the genomics of adaptation to climate in the great tit and other wild vertebrate systems will be aided by advances in monitoring of local climate variables and in predictive ecology approaches.

5.4.1. Monitoring Technology

More detailed monitoring of local climate conditions and physiological responses could be recorded in the nest box e.g. using biosensors of body temperature and metabolic rate in great tits (Andreasson et al. 2023). Alternatively, experienced conditions inside and outside the nest box could be real-time monitored e.g. temperature and humidity were recorded for flycatchers (Riggio et al. 2023), and even oxygen levels, or wind, rain or other climate variables could be directly recorded.

Studies of foraging behaviour and social networks at Wytham Woods have used radio frequency tracking to monitor feeding stations (McMahon et al. 2024). This approach could be used to monitor foraging or parental care behaviour in populations that experience intense summer heatwave events, where parental care activity is expected to be reduced in heat-stressed conditions. For example, reduced visits to the nest box were correlated with high daytime temperature in wild zebra finches, *Taeniopygia castanotis*, suggesting nestlings receive less food under thermally challenging conditions (Wheeler et al. 2023).

Some of the great tit population field sites are located in the Mediterranean region, which is predicted to have more frequent heatwaves in the future (IPCC 2023; see **Figure 1.8**). For these Mediterranean populations, it might be important to consider measures that reduce heat stress in nest boxes. For example, the experimental placement of nest boxes in shade compared to control (non-shade) conditions has been shown to reduce hatching failure and nestling mortality in the Mediterranean lesser kestrel, *Falco naumanni*, across the breeding season 2021-22 (Corregidor-Castro et al. 2023).

The use of drones or satellite images could be useful to observe the impacts of climate change on field sites (as is due to be investigated at Wytham Woods on the great tit-winter moth-oak tree system), by measuring the temporal changes in ecological communities, habitats and land-use (Pettorelli et al. 2014). Temperature and not photoperiod has been experimentally demonstrated to be driving the seasonal timing of winter moth egg hatching (van Dis et al. 2024). Further, local monitoring of insect communities and their abundance may serve to indicate ecosystem health. Understanding biotic and abiotic stress tolerance in species in an ecosystem could be further investigated. Overall, incorporating monitoring of more local climate variables and understanding different parts of the ecosystem may aid future investigations of genomic adaptation to experienced climate conditions.

The increased use of citizen science recording, such as bird song apps on smartphones for species identification and wildlife photo identification (such as for trees, plants, birds and insects) could aid species monitoring efforts. The efforts of bird ringing and sighting

reporting across Europe, will help to map bird breeding and migration. Additionally, uploading field study data into standardised data formats on a database e.g. SPI-Birds network (https://spibirds.org/en), aids the usability of the collected data for comparisons of the different studied bird species, or to ask other research questions.

5.4.2. Predictive ecology

The modelling of climate data to predict future climate conditions is a process that contains uncertainties, from potential errors in the direct measurement or estimation of climate variables over regions, to the estimation of the likelihood of emissions targets being met in the future (Schoeman et al. 2023). Climate change is multidimensional and single metrics may neglect all but the central trajectory of change, especially for extreme weather events (González-Trujillo et al. 2023). More accurate future climate models will help in predicting genomic responses of wild populations to climate changes in the future. Developments in technology like deep learning may be used in population genetics to detect signatures of selection e.g. Korfmann et al. (2023) demonstrated, using simulations, the ability to detect balancing selection from temporal changes in haplotype data. Machine learning has also been applied to variant effect prediction in human genomes, to understand the effect of dysregulated genes on biological pathways for disease (Wong et al. 2021). Genomic data and artificial intelligence (AI) machine learning models may become more widely employed inference frameworks in evolutionary genomics and will start to play an increasing role in conservation genomics (van Oosterhout 2024). Machine learning, combined with genotypic or haplotypic data from higher throughput sequencing technologies, may be able to train population genetic approaches to predict how species will adapt to new climate conditions. For example, Gradients Forests (GF) is a machine learning approach used for incorporating environmental drivers of genomic variation into climate change impact assessments. Fitzpatrick et al. (2021) found experimental support using balsam poplar trees for the genomic prediction of climate maladaption (from genomic offset in the GF method).

Online simulations of ecosystems, like the planned digital simulation of Hoge Veluwe National Park (Long-term Ecosystem Research-LIFE: https://lter-life.nl/en/large-scale-

research-infrastructure), aim to be a resource to compare and predict how different ecosystems will respond to climate changes. The digitisation of a studied ecosystem requires long-term ecology monitoring data, interaction models and climate scenarios, and that the information can be retrieved and queried on a website interface via cloud computing. In addition, the digitisation of natural history collections could also provide more open access of genome sequence and phenotypic measurement data of museum specimens. This would give further insight into the impact of global change on phenotypes, phenology, genetic diversity, population size or range, and interactions among species (Sanders et al. 2023). Digitisation of natural history collections may help provide a historical record of specimens, which could allow investigation of genomic adaptation responses to past climate changes (Holmes et al. 2016; Lopez et al. 2020).

The collection of large genomic data sets to investigate adaptation to climate, often including genome-wide and/or whole-genome sequences from many sampled individuals, requires the ability to process large data files and to digitally store them. Best practice for genomic research now includes open access of data and code, to allow reproducibility of bioinformatic pipelines. The publication of Chapter 2 followed these open data practices for publication in Evolution Letters (doi: 10.1093/evlett/qrad043) with a Dryad data repository (https://doi.org/10.5061/dryad.vt4b8gtxd). Open data practices are important for future research in the field. Possible uses include: 1) Combining the data with data from other studies in a meta-analysis. 2) Genotypes could be imputed across the whole-genome from the SNP chip data for further analysis and comparison. 3) A follow up study could compare allele frequencies observed in the future with the data generated here.

In summary, it is expected that in the future more advanced monitoring technology will be incorporated into field studies, and that will be able to measure many climate and physiological variables, which will help in understanding local environmental/climate changes and resulting gene adaptation responses towards them. Better predictive approaches for identifying candidate adaptive loci would reduce the false positive results, and more accurately identify the best candidate adaptation genes. At present there are only very few studies that functionally validate the fitness differences of identified candidate

climate adaptation genes, and these studies have not typically been conducted in vertebrates. The next step for future studies would be to experimentally validate the candidate climate adaptation genes identified in vertebrate species, by observing fitness differences for carrying candidate adaptation genes in the home environment or via gene editing methods, to truly implicate these genes and their biological function in the process of climate adaptation. Understanding when and how vertebrate populations will adapt to future climate change is an important topic to consider.

5.5. References

Andreasson, F., Rostedt, E., & Nord, A. (2023). Measuring body temperature in birds - the effects of sensor type and placement on estimated temperature and metabolic rate. *J. Exp. Biol.*, 226(24):jeb246321. doi: 10.1242/jeb.246321

Ansai, S., & Kitano, J. (2022). Speciation and adaptation research meets genome editing. *Phil. Trans. R. Soc. B.*, **377**:e20200516. doi: 10.1098/rstb.2020.0516

Ashraf, B., Hunter, D.C., Bérénos, C., Ellis, P.A., Johnston, S.E., Pilkington, J.G., Pemberton, J.M., & Slate, J. (2022). Genomic prediction in the wild: A case study in Soay sheep. *Mol. Ecol.*, **31**:6541–6555. doi:10.1111/mec.16262

Ben-Jemaa, S., Mastrangelo, S., Lee, S-H., Lee, J.H., & Boussaha, M. (2020). Genome-wide scan for selection signatures reveals novel insights into the adaptive capacity in local North African cattle. *Sci. Rep.*, **10**:e19466. doi:10.1038/s41598-020-76576-3

Charmantier, A., McCleery, R.H., Cole, L.R., Perrins, C., Kruuk, L.E.B., & Sheldon, B.C. (2008). Adaptive Phenotypic Plasticity in Response to Climate Change in a Wild Bird Population. *Science*, **320**:800-803. doi:10.1126/science.1157174

Corregidor-Castro, A., Morinay, J., McKinlay, S.E., Ramellini, S., Assandri, G., Bazzi, G., Glavaschi, A., De Capua, E.L., Grapputo, A., Romano, A., Morganti, M., Cecere, J.G., Pilastro, A., & Rubolini, D. (2023). Experimental nest cooling reveals dramatic effects of heatwaves on reproduction in a Mediterranean bird of prey. *Glob. Change Biol.*, **29**:5552–5567. doi:10.1111/gcb.16888

Fitzpatrick, M.C., Chhatre, V.E., Soolanayakanahally, R.Y. & Keller, S.R. (2021), Experimental support for genomic prediction of climate maladaptation using the machine learning approach Gradient Forests*. Mol Ecol. Resour.*, **21**: 2749-2765. Doi:10.1111/1755-0998.13374

Flori, L., Moazami-Goudarzi, K., Alary, V., Araba, A., Boujenane, I., Boushaba, N., Casabianca, F., Casu, S., Ciampolini, R., D'Acier, A.C., Coquelle, C., Delgado, J-V., El-Beltagi, A., Hadjipavlou, G., Jousselin, E., Landi, V., Lauvie, A., Lecomte, P., Ligda, C., ...Gautier, M. (2019). A Genomic Map of Climate Adaptation in Mediterranean Cattle Breeds. Mol. Ecol., 28(5):1009–1029. doi:10.1111/mec.15004

Gienapp, P., Calus, M.P.L., Laine, V.N., & Visser, M.E. (2019). Genomic selection on breeding time in a wild bird population. *Evol. Lett.*, **3**:142-151. doi:10.1002/evl3.103

González-Trujillo, J.D., Román-Cuesta, R.M., Muñiz-Castillo, A.I., Amaral, C.H., & Araújo, M.B. (2023). Multiple dimensions of extreme weather events and their impacts on biodiversity. *Climatic Change*, **176**:155. doi:10.1007/s10584-023-03622-0

Grant, R.B., & Grant, P.R. (2003). What Darwin's Finches Can Teach Us about the Evolutionary Origin and Regulation of Biodiversity, BioScience, 53(10):965–975. doi:10.1641/0006-3568(2003)053[0965:WDFCTU]2.0.CO;2

Holmes, M.W., Hammond, T.T., Wogan, G.O., Walsh, R.E., LaBarbera, K., Wommack, E.A., Martins, F.M., Crawford, J.C., Mack, K.L., Bloch, L.M., & Nachman, M.W. (2016). Natural history collections as windows on evolutionary processes. *Mol. Ecol.*, **25**(4):864-81. doi: 10.1111/mec .13529

IPCC 2023: Sections. In: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland, pp. 35-115. doi:10.59327/IPCC/AR6-9789291691647

Korfmann, K., Gaggiotti, O.E., & Fumagalli, M. (2023). Deep Learning in Population Genetics, *Genome Biol. Evol.*, **15**(2):evad008. doi:10.1093/gbe/evad008

Lindner, M., Ramakers, J.J.C., Verhagen, I., Tomotani, B.M., Mateman, A.C., Gienapp, P., & Visser, M.E. (2023). Genotypes selected for early and late avian lay date differ in their phenotype, but not fitness, in the wild. *Sci. Adv.*, **9**:eade6350. doi:10.1126/sciadv.ade6350

Lopez, L., Turnerm, K.G., Bellis, E.S., & Lasky, J.R. (2020). Genomics of natural history collections for understanding evolution in the wild. *Mol. Ecol. Resour.*, **20**:1153–1160. doi:10.1111/1755-0998.13245

McMahon, K., Marples, N.M., Spurgin, L.G., Rowland, H.M., Sheldon, B.C., & Firth, J.A. (2024). Social network centrality predicts dietary decisions in a wild bird population. *iScience*, 00. doi:10.1016/j.isci.2024.109581

Meuwissen, T., Hayes, B., & Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**:1819–1829. doi: 10.1093/genetics/157.4.1819

Pettorelli, N., Laurance, W.F., O'Brien, T.G., Wegmann, M., Nagendra, H., & Turner, W. (2014). Satellite remote sensing for applied ecologists: opportunities and challenges. *J. Appl. Ecol.*, **51**(4):839–848. doi:10.1111/1365-2664.12261

Rachman, M.P., Bamidele, O., Dessie, T., Smith, J., Hanotte, O., & Gheyas, A.A. (2024). Genomic analysis of Nigerian indigenous chickens reveals their genetic diversity and adaptation to heat-stress. *Sci. Rep.*, **14**:e2209. doi:10.1038/s41598-024-52569-4

Riggio, J., Engilis, A., Cook,H., de Greef, E., Karp, D.S., Truan, M.L. (2023). Long-term monitoring reveals the impact of changing climate and habitat on the fitness of cavity-nesting songbirds. *Biol. Conserv*., 278:e109885. doi:10.1016/j.biocon.2022.109885

R2D2 Consortium, Fugeray-Scarbel, A., Bastien, C., Dupont-Nivet, M., & Lemarié, S. (2021) Why and How to Switch to Genomic Selection: Lessons From Plant and Animal Breeding Experience. *Front. Genet.*, **12**:e629737. doi: 10.3389/fgene.2021.629737

Sanders, N.J., Cooper, N., Davis Rabosky, A.R., & Gibson, D.J. (2023). Leveraging natural history collections to understand the impacts of global change. *J. Anim. Ecol.*, **92**:232–236. doi:10.1111/1365-2656.13882

Saravanan, K.A., Panigrahi, M., Kumar, H., Parida, S., Bhushan, B., Gaur, G.K., Dutt, T., Mishra,B.P., & Singh, R.K. (2021). Genomic scans for selection signatures revealed candidate genes for adaptation and production traits in a variety of cattle breeds. *Genomics*, **113**(3):955-963. doi:10.1016/j.ygeno.2021.02.009

Schoeman, D.S., Sen Gupta, A., Harrison, C.S., Everett, J.D., Brito-Morales, I., Hannah, L., Bopp, L., Roehrdanz, P.R., Richardson, A.J. (2023). Demystifying global climate models for use in the life sciences. *Trends Ecol. Evol.*, **38**(9):843-858. doi:10.1016/j.tree.2023.04.005

Swanson, D. L., Zhang, Y, & Jimenez, A. G. (2022). Skeletal muscle and metabolic flexibility in response to changing energy demands in wild birds. *Front Physiol.*, **22**(13):961392. doi: 10.3389/fphys.2022.961392

Tian, S., Zhou, X., Phuntsok, T., Zhao, N., Zhang, D., Ning, C., Li, D., & Zhao, H. (2020). Genomic Analyses Reveal Genetic Adaptations to Tropical Climates in Chickens. *iScience*, **23**(11):e101644. doi:10.1016/j.isci.2020.101644

Urban, M.C., Swaegers, J., Stoks, R., Snook, R.R., Otto, S.P., Noble, D.W.A., Moiron, M., Hällfors, M.H., Gómez-Llano, M., Fior, S., Cote, J., Charmantier, A., Bestion, E., Berger, D., Baur, J., Alexander, J.M., Saastamoinen, M., Edelsparre, A.H., & Teplitsky, C. (2024). When and how can we predict adaptive responses to climate change? *Evol. Lett.*, **8**(1):172–187. doi:10.1093/evlett/qrad038

van Dis, N.E., Salis, L. & Visser, M.E. (2024). Temperature has an overriding role compared to photoperiod in regulating the seasonal timing of winter moth egg hatching. *Oecologia*, 00. doi:10.1007/s00442-024-05535-w

van Oosterhout, C. (2024). AI-informed conservation genomics. *Heredity*, **132**:1–4. doi:10.1038/s41437-023-00666-x

Wheeler, M., Ton, R., Brandl, H.B., Schuett, W., & Griffith, S.C. (2023). Reduced parental brood visit rate in wild zebra finches Taeniopygia castanotis is correlated with high maximum daily ambient temperature. *J. Avian Biol.*, 00:e03152. doi:10.1111/jav.03152

Wong, A.K., Sealfon, R.S.G., Theesfeld, C.L., & Troyanskaya, O.G. (2021). Decoding disease: from genomes to networks to phenotypes. *Nat. Rev. Genet.*, **22**:774–790 (2021). doi:10.1038/s41576-021-00389-x