

**Patterns of genomic differentiation and hybridization between two merging Emberizidae  
species implicate a sex chromosome inversion in plumage variation**

by

Ellen Glyn Marshall Nikelski

B.Sc., The University of British Columbia, 2017

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF  
THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Zoology)

THE UNIVERSITY OF BRITISH COLUMBIA

(Vancouver)

April 2021

© Ellen Glyn Marshall Nikelski, 2021

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, a thesis entitled:

**Patterns of genomic differentiation and hybridization between two merging Emberizidae species implicate a sex chromosome inversion in plumage variation**

---

submitted by Ellen Glyn Marshall Nikelski in partial fulfillment of the requirements for

the degree of Master of Science

in Zoology

---

**Examining Committee:**

Dr. Darren Irwin, Professor, Department of Zoology, UBC

Supervisor

Dr. Dolph Schluter, Professor, Department of Zoology, UBC

Supervisory Committee Member

Dr. Eric Taylor, Professor, Department of Zoology, UBC

Supervisory Committee Member

Dr. Judith Mank, Professor, Department of Zoology, UBC

Additional Examiner

## Abstract

Hybrid zones offer researchers the opportunity to investigate how evolutionary processes interact to drive speciation forward. However, in these areas where genetic divergence competes against gene flow, speciation and population merging are both possible with the outcome dependent on the strength of reproductive barriers between groups. The yellowhammer (*Emberiza citrinella*) and pine bunting (*Emberiza leucocephalos*) are Palearctic songbirds with highly divergent plumage patterns. Despite their differences, these taxa hybridize extensively and show negligible differentiation in their mitochondrial genomes. These observations create a conflicting picture of the state of reproductive barriers between groups, raising the question whether yellowhammers and pine buntings are actually separate species. In this thesis, I examine patterns of genetic variation among phenotypically pure and hybrid individuals to assess the strength of reproductive isolation between taxa. I hypothesize that, unlike mitochondrial differentiation, nuclear differentiation will be moderate and that patterns of divergence will reflect some weak reproductive isolation between groups. I find that, in allopatry, yellowhammers and pine buntings separate into distinct genetic clusters based on an island of differentiation on the Z chromosome. Yet, in other parts of the genome, I find evidence of past mitonuclear gene introgression. In sympatry, I report a breakdown of allopatric genetic clusters driven by extensive interbreeding. These findings combined with the high number of late generation hybrids identified within the sympatric zone suggest that reproductive barriers are weak between taxa. Interestingly, I further find low recombination within the island of differentiation identified between allopatric populations implying that this region may house a chromosomal inversion. The inversion is highly associated with plumage variation and may be

responsible for the maintenance of parental phenotypes within the sympatric zone. Because reproductive barriers are weak, it is likely that hybridization will continue between yellowhammers and pine buntings potentially leading to the merging of these groups, but that the putative inversion could preserve parental plumage phenotypes within this single species. Retention of such variation would increase the evolvability of the system such that the population could be safeguarded from extinction or, if evolutionary pressures change, could diverge again and move towards speciation.

## Lay Summary

The yellowhammer (*Emberiza citrinella*) and pine bunting (*Emberiza leucocephalos*) are songbirds that differ greatly in plumage, but that hybridize where they co-occur and differ little in their mitochondrial DNA. These observations had led scientists to question whether these groups are separate species. Here, I examine genetic variation across the system to assess differentiation between groups. I find that yellowhammers and pine buntings differ genetically where they do not overlap, but that, in the hybrid zone, “hybrid” and “pure” individuals often possess genetic material from both groups. This suggests that yellowhammers and pine buntings are not genetically distinct and could merge into one population. Yet, one genetic region important to plumage variation is inherited as a genetic block allowing “pure” plumage patterns to be preserved in the hybrid zone. These findings highlight the balance between divergence and merging during species formation and suggest how trait variation is maintained within groups.

## **Preface**

I devised the research questions and experimental design for this project in collaboration with Dr. Darren Irwin. The majority of blood and tissue samples used in this thesis were collected by Dr. Alexander Rubtsov who is a collaborator on this project. Blood and tissue samples that were not collected in the field were obtained from The Bell Museum, The Burke Museum of Natural History and Culture, The Field Museum, The State Darwin Museum, The Swedish Museum of Natural History, The Zoological Museum of the Zoological Institute of the Russian Academy of Sciences and the Zoological Museum of the University of Copenhagen. A full description of the samples used in this thesis and their origins can be found in the Appendix. DNA from a subset of blood and tissue samples used in this research was extracted by Dr. Alexander Rubtsov as part of a previous project on the yellowhammer and pine bunting system. I conducted all remaining DNA extractions as well as the molecular genetic procedures presented in this study. Statistical analyses of DNA sequence data were designed and performed by me in collaboration with Dr. Darren Irwin. Finally, I wrote all parts of this thesis with feedback from Dr. Darren Irwin.

# Table of Contents

<b>Abstract.....</b>	<b>iii</b>
<b>Lay Summary .....</b>	<b>v</b>
<b>Preface.....</b>	<b>vi</b>
<b>Table of Contents .....</b>	<b>vii</b>
<b>List of Tables .....</b>	<b>xi</b>
<b>List of Figures.....</b>	<b>xiv</b>
<b>List of Abbreviations .....</b>	<b>xviii</b>
<b>Acknowledgements .....</b>	<b>xix</b>
<b>Chapter 1: Introduction .....</b>	<b>1</b>
1.1    Speciation as a process.....	1
1.2    Genomic differentiation during speciation .....	5
1.3    Reproductive barriers.....	8
1.4    The yellowhammer and pine bunting system .....	10
1.5    Thesis objectives and research implications.....	13

## Chapter 2: Mitonuclear co-introgression opposes speciation between two hybridizing

<b>songbirds</b> .....	<b>17</b>
2.1 Introduction.....	17
2.2 Methods.....	26
2.2.1 Sampling .....	26
2.2.2 DNA sequencing and identification of single nucleotide polymorphisms .....	28
2.2.2.1 DNA extraction and genotyping-by-sequencing .....	28
2.2.2.2 Genotyping-by-sequencing data filtering .....	28
2.2.3 Statistical Analysis.....	30
2.2.3.1 Variant site analyses .....	30
2.2.3.2 Differentiation across the genome .....	31
2.2.3.3 Phylogenetic comparison with outgroups.....	32
2.2.3.4 Signals of mitonuclear co-introgression .....	32
2.3 Results.....	35
2.3.1 Phylogenetic comparison with outgroups.....	35
2.3.2 Overall genetic differentiation .....	36
2.3.3 Differentiation across the genome .....	37
2.3.4 Signals of mitonuclear co-introgression .....	39

2.4	Discussion .....	41
<b>Chapter 3: A chromosomal inversion maintains divergent plumage phenotypes as two avian species merge into one .....</b>		
<b>67</b>		
3.1	Introduction.....	67
3.2	Methods.....	76
3.2.1	Sampling .....	76
3.2.2	DNA sequencing and identification of single nucleotide polymorphisms .....	77
3.2.3	Statistical analysis.....	78
3.2.3.1	Variant site analyses .....	78
3.2.3.2	Plumage trait admixture mapping.....	80
3.2.3.3	Investigation of plumage trait dominance patterns.....	81
3.3	Results.....	82
3.3.1	Population structure inside and outside the hybrid zone .....	82
3.3.2	Differentiation on the Z chromosome.....	86
3.3.3	Demographic patterns across the hybrid zone .....	88
3.3.4	Genetic underpinnings of plumage traits .....	90
3.4	Discussion.....	95
<b>Chapter 4: Conclusions .....</b>		
<b>125</b>		

<b>Bibliography .....</b>	<b>141</b>
<b>Appendices.....</b>	<b>163</b>
Appendix A Supplementary Tables and Figures .....	163
A.1 Supplementary Tables.....	163
A.2 Supplementary Figures .....	196

## List of Tables

- Table 2.1.** Geographical location, sampling size and sampling break-down for each of the sites included in this thesis. Sampling locations may include multiple sites that appeared too close together to be shown in detail in Figure 2.1. Full details for the sites included in each sampling locations can be found in Supplementary Table 2.1. The sampling location numbers that appear in the “Sampling Location” column correspond to those that appear in red in Figure 2.1. The “Sample Size” columns describes the total number of samples collected from a particular site. Columns “Allopatric E. cit” – “Allopatric E. leuc” describe the demographic breakdown of samples within each sampling location. “E. cit.” represents *Emberiza citrinella* or yellowhammers and “E. leuc.” represents *Emberiza leucocephalos* or pine buntings. .... 52
- Table 2.2.** Summary statistics calculated while conducting mitonuclear co-introgression analysis. A total of 7187 2000bp windows were considered when determining introgression windows. A total of 134 mitonuclear genes were investigated for signals of co-introgression. “\*\*\*” indicates a significant p-value..... 55
- Table 2.3.** Identities, chromosomal locations, window rankings and functions of mitonuclear genes that appeared within 244 yellowhammer introgression windows. Window rankings span values of 1 to 244 and are based on the average Tajima’s D value of each introgression window. The introgression window with the lowest average Tajima’s D value was given the ranking of 1 and the introgression window with the highest Tajima’s D value was given a ranking of 244. In the “Mitonuclear Gene Function” column, ETC stands for “Electron Transport Chain”. Mitonuclear gene names are written as they appear in Hill (2019). .... 56
- Table 2.4.** Identities, chromosomal locations, window rankings and functions of mitonuclear genes that appeared within 222 pine bunting introgression windows. Window rankings span values of 1 to 222 and are based on the average Tajima’s D value of each introgression window. The introgression window with the lowest average Tajima’s D value was given the ranking of 1 and the introgression window with the highest Tajima’s D value was given a ranking of 222. In the “Mitonuclear Gene Function” column, ETC stands for “Electron Transport Chain”. Mitonuclear gene names are written as they appear in Hill (2019). .... 57
- Table 3.1.** Identities of SNPs showing significant association with phenotypic variation in the background colour of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form -log(p-value). Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29811. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found within an

annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Brow” indicates the amount of chestnut plumage at the brow and “Throat” indicates the amount of chestnut at the throat. “NA” indicates that a particular significant SNP was unique to the background plumage trait. .... 110

**Table 3.2.** Identities of SNPs that showed a significant association with phenotypic variation in the amount of chestnut plumage at the brow of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form  $-\log(p\text{-value})$ . Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29768. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found in an annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Background” indicates the colour of the background plumage and “Throat” indicates the amount of chestnut plumage at the throat. “NA” indicates that a particular significant SNP was unique to the brow plumage trait. 111

**Table 3.3.** Identities of SNPs that showed a significant association with phenotypic variation in the amount of chestnut plumage at the throat of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form  $-\log(p\text{-value})$ . Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29851. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found in an annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Background” indicates the colour of the background plumage and “Brow” indicates the amount of chestnut plumage at the brow. “NA” indicates that a particular significant SNP was unique to the throat plumage trait. 113

**Supplementary Table 2.1.** Detailed information on all the samples included in this study. Explanations for the abbreviations used in the “Phenotypic Class” and “Geographic Distribution” columns can be found in the methods sections of Chapter 2. In the “sex” column, “m” stands for male, “f” stands for female and “uk” stands for unknown. The “TH” column contains phenotypic scores for each individual for the throat plumage trait. The “BR” column contains phenotypic scores for each individual for the brow plumage trait. The “BG” column contains phenotypic scores for each individual for the background colour plumage trait. In the “Pheno Class” (Phenotypic Class) column, “OUT” stands for outgroup, “FML” stands for female and “UK” stands for unknown. The numbers in the “Sampling Location” column correspond to those that appear in Figure 2.1A. In all columns, a “NA” observation stands for “Not Applicable”. .... 163

**Supplementary Table. 2.2.** Detailed information on the genomic locations and functions of the 134 mitonuclear genes investigated in this study. In the “Function” column, “ETC” stand for electron transport chain. Mitonuclear gene names are written as they appear in Hill (2019). ... 188

## List of Figures

- Figure 2.1. A)** Map showing all the sampling locations included in this research. Red numbers accompanying each sampling location pie correspond to the sampling location numbers that appear in Table 1, which shows sample size and composition. Sampling locations may include multiple sites that appeared too close together to be shown in detail on the map. Full details for the sites included in each sampling locations can be found in Supplementary Table 2.1. Each sampling location pie is coloured and divided based on the proportion of each sample type that appeared within it. The sample types include: allopatric yellowhammers (Allo-Cit; yellow), near-sympatric yellowhammers (Near Sym-Cit; light orange), sympatric yellowhammers (Sym-Cit; red-orange), hybrids (Hybrid; green), sympatric pine buntings (Sym-Leuc; peach), near sympatric leucocephalos (Near Sym-Leuc; taupe) and allopatric pine buntings (Allo-Leuc; brown). The solid black line indicates the geographic range of the yellowhammer and the dashed black line indicates the geographic range of the pine bunting as described in Irwin et al. (2009).
- B)** Photos depicting phenotypic variation within the yellowhammer and pine bunting system. Each photo represents one of eight phenotypic classes that individuals are divided into based on variation at three plumage traits: background colour, amount of chestnut at the brow and amount of chestnut at the throat. The photos show one example of each classes, but are unable to capture the full variation within each phenotypic group. All photos are credited to Dr. Alexander Rubtsov. .... 59
- Figure 2.2.** Neighbour-joining tree of *Emberizidae* species created using average absolute between-population nucleotide diversity ( $\pi_B$ ). Sample sizes for each species are as follows: *E. aureola* = 1, *E. calandra* = 1, *E. cioides* = 1, *E. hortulana* = 1, *E. cirhus* = 6, *E. stewarti* = 4, *E. citrinella* = 53 and *E. leucocephalos* = 42. .... 60
- Figure 2.3.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 53) and allopatric pine buntings (brown; n = 42). PC1 explains 3.6% of the variation among individuals and PC2 explains 2.9% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 61
- Figure 2.4.** Relative differentiation ( $F_{ST}$ ) of 349,807 genome-wide SNPs identified between allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42). .... 62
- Figure 2.5.** Genome-wide patterns of genetic variation comparing allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42). Relative nucleotide differentiation ( $F_{ST}$ ), absolute between-population nucleotide diversity ( $\pi_B$ ), absolute within-population diversity ( $\pi_W$ ) and Tajima's D ( $Taj_D$ ) are shown as 2000 bp windowed averages across each chromosome.  $F_{ST}$  and  $\pi_B$  are shown as purple lines to indicate that values were calculated as a comparison between allopatric yellowhammers and pine buntings.  $\pi_W$  and  $Taj_D$  are shown as two separate lines (yellow = yellowhammers, brown = pine buntings) to indicate that values were calculated separately for each population. .... 64

**Figure 2.6.** The mean within-group absolute variation ( $\pi_W$ ) of allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42) plotted against between-group absolute differentiation ( $\pi_B$ ). Each dot represents the average value taken from a 2000 bp window of sequenced data across the nuclear genome. The black line indicates where mean within-group differentiation equals between-group differentiation. Increasing average values of  $F_{ST}$  calculated for each window are shown in darker shades of blue. .... 65

**Figure 2.7.** Association between relative differentiation ( $F_{ST}$ ) and between-group differentiation ( $\pi_B$ ) of allopatric yellowhammers (n=53) and allopatric pine buntings (n = 42). Each black dot represents average values calculated from a 2000bp window of sequenced data. A cubic spline fit between the variables is shown as purple line. .... 66

**Figure 3.1.** Genetic relationships between allopatric yellowhammers (n = 53), near-sympatric yellowhammers (n = 15), sympatric yellowhammers (n = 67), allopatric pine buntings (n = 42), near-sympatric pine buntings (n = 18), sympatric pine buntings (n = 52) and hybrids (n = 74). **A)** Whole-genome principal components analysis of allopatric yellowhammers (yellow), near-sympatric yellowhammers (light orange), sympatric yellowhammers (red-orange), allopatric pine buntings (brown), near-sympatric pine buntings (taupe), sympatric pine buntings (peach) and hybrids (green). PC1 explains 1.4% of the variation among individuals and PC2 explains 0.9% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. **B)** Ancestry proportions of allopatric yellowhammers (Allo- C), near-sympatric yellowhammers (N.Sym- C), sympatric yellowhammers (Sym- C), allopatric pine buntings (Allo- L), near-sympatric pine buntings (N.Sym- L), sympatric pine buntings (Sym- L) and hybrids (Hybrid) as predicted by an Admixture model with K=2. Information from 417,164 SNPs were included in this analysis. .... 115

**Figure 3.2.** Genetic differentiation across the Z chromosome among allopatric yellowhammers (n = 53), near-sympatric yellowhammers (n = 15), sympatric yellowhammers (n = 67), allopatric pine buntings (n = 42), near-sympatric pine buntings (n = 18), sympatric pine buntings (n = 52) and hybrids (n = 74). **A)** Z chromosome principal components analysis of allopatric yellowhammers (yellow), near-sympatric yellowhammers (light orange), sympatric yellowhammers (red-orange), allopatric pine buntings (brown), near-sympatric pine buntings (taupe), sympatric pine buntings (peach) and hybrids (green). PC1 explains 8.1% of the variation among individuals and PC2 explains 1.7% of the variation among individuals. Information from 11,147 SNPs was included in this analysis. Numbered boxes are used to designate each of 6 clusters within PC space. **B)** Genotype-by individuals plot of a subset of females (n = 32). **C)** Genotype-by individuals plot of a subset of males (n = 100) from the chromosome Z PCA. SNPs with  $F_{ST}$  greater or equal to 0.7 in comparisons of allopatric populations were included in this analysis. Boxes filled in with one colour indicate homozygosity at a locus and boxes split into different coloured triangles indicate heterozygosity. Light purple indicates alleles with putative pine bunting ancestry and dark purple indicates alleles with putative yellowhammer ancestry. Numbers along left side correlate to numbered clusters within the chromosome Z PCA. .... 117

**Figure 3.3.** Triangle plots of interclass heterozygosity (H) versus ancestry index (S) for allopatric yellowhammers (n = 53; yellow), near-sympatric yellowhammers (n = 15; light orange), sympatric yellowhammers (n = 67; red-orange), allopatric pine buntings (n = 42;

brown), near-sympatric pine buntings (n = 18; taupe), sympatric pine buntings (n = 52; peach) and hybrids (n = 74; green). **A)** Analysis including 145 SNPs that possess  $F_{ST}$  values greater than or equal to 0.6 when comparing allopatric yellowhammer and allopatric pine bunting samples. **B)** Analysis including 10 unlinked SNPs possessing  $F_{ST}$  values greater than or equal to 0.5 when comparing allopatric yellowhammer and allopatric pine bunting samples. .... 118

**Figure 3.4.** Associations between genome-wide SNPs and phenotypic variation in three plumage traits within the yellowhammer and pine bunting system. P-values for each SNP were determined using a likelihood ratio test calculated using the GEMMA program. Red lines indicate Bonferroni corrected significance thresholds. **A)** Associations between 220,220 genome-wide SNPs and variation in the background plumage colour of parental and hybrid individuals. **B)** Associations between 220,124 genome-wide SNPs and variation in the amount of chestnut plumage at the brow of parental and hybrid individuals. **C)** Associations between 220,307 genome-wide SNPs and variation in the amount of chestnut plumage at the throat of parental and hybrid individuals. .... 120

**Figure 3.5.** Investigation of potential dominance interactions between the alleles of SNPs that are significantly associated with plumage variation in the yellowhammer and pine bunting system as identified using GEMMA. **A)** Relative differentiation ( $F_{ST}$ ) of 111,47 SNPs on the Z chromosome calculated when comparing allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42). Four SNPs are highlighted in this panel: Z.4835388 (green), Z.18131016 (magenta), Z.26812248 (orange) and Z.59571043 (blue). **B)** Balloon plots illustrating the number of individuals possessing a particular phenotypic score at one three plumage traits with a particular genotype at one of the four SNPs highlighted in panel A. The plumage traits considered are: the background plumage colour (“Background”), the amount of chestnut at the brow (“Brow”) and the amount of chestnut at the throat (“Throat”). The SNP identities are indicated with the same colours as in panel A: Z.4835388 (green), Z.18131016 (magenta), Z.26812248 (orange) and Z.59571043 (blue). Dot size in each balloon plot indicates the number of individuals possessing a specific phenotypic score-genotype combination. Red stars indicate whether each of the four highlighted SNPs is significantly associated with phenotypic variation at the plumage trait of interest as determined using GEMMA. Blue labels “YH” (Yellowhammer) and “PB” (Pine Bunting) indicate phenotypic scores more commonly associated with “pure” members of each taxa. .... 122

**Figure 3.6.** Associations between SNPs located on chromosome Z and phenotypic variation in three plumage traits within the yellowhammer and pine bunting system. P-values for each SNP were determined using a likelihood ratio test calculated using the GEMMA program. Red lines indicate Bonferroni corrected significance thresholds. **A)** Associations between 6502 chromosome Z SNPs and variation in the background plumage colour of parental and hybrid individuals. **B)** Associations between 6509 chromosome Z SNPs and variation in the amount of chestnut plumage at the brow of parental and hybrid individuals. **C)** Associations between 6487 chromosome Z SNPs and variation in the amount of chestnut plumage at the throat of parental and hybrid individuals. .... 124

**Supplementary Figure 2.1.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 53) and allopatric pine buntings (brown; n = 41) following the

removal of sample “*Emberiza\_GBS1\_ASR05\_49*”. This sample was one of a pair of outliers that appeared in the principal components analysis shown in Figure 1. PC1 explains 8.6% of the variation among individuals and PC2 explains 3.0% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 196

**Supplementary Figure 2.2.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 52) and allopatric pine buntings (brown; n = 41) following the removal of sample “*Emberiza\_GBS1\_ASR05\_49*” and “*Emberiza\_GBS2\_BKS\_1609*”. “*Emberiza\_GBS1\_ASR05\_49*” was one of a pair of outliers that appeared in the principal components analysis shown in Figure 1. “*Emberiza\_GBS2\_BKS\_1609*” was an outlier that appeared in the principal components analysis shown in Supplementary Figure S1. PC1 explains 3.2% of the variation among individuals and PC2 explains 2.6% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 197

**Supplementary Figure 3.1.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 53), near-sympatric yellowhammers (light orange; n = 15), allopatric pine buntings (brown; n = 42) and near-sympatric pine buntings (taupe; n = 18). PC1 explains 2.9% of the variation among individuals and PC2 explains 2.5% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 198

**Supplementary Figure 3.2.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 53), near-sympatric yellowhammers (light orange; n = 15), sympatric yellowhammers (red-orange; n = 67), allopatric pine buntings (brown; n = 42), near-sympatric pine buntings (taupe; n = 18) and sympatric pine buntings (peach; n = 52). PC1 explains 1.7% of the variation among individuals and PC2 explains 1.5% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 199

**Supplementary Figure 3.3.** Ancestry proportions of hybrid individuals split into the phenotypic classes described in Rubtsov & Tarasov, 2017. Phenotypic classes include: citrinella hybrid (CH, n = 23), yellow hybrid (YH, n = 14), white hybrid (WH, n = 18), leucocephalos hybrid (LH, n = 12), female (FML, n = 4) and unknown class (UK, n = 3). Ancestry proportions were predicted by an Admixture model with K=2 and information from 417,164 SNPs. .... 200

**Supplementary Figure 3.4.** Whole-genome principal components analysis of allopatric yellowhammers (yellow; n = 53), near-sympatric yellowhammers (light orange; n = 15), sympatric yellowhammers (red-orange; n = 67), allopatric pine buntings (brown; n = 42), near-sympatric pine buntings (taupe; n = 18), sympatric pine buntings (peach; n = 49) and hybrids (green; n = 72) following the removal of five outliers identified in Figure 1A: “*Emberiza\_GBS4\_XD\_632*”, “*Emberiza\_GBS4\_XD\_636*”, “*Emberiza\_GBS4\_XD\_639*”, “*Emberiza\_GBS4\_XD\_972*” and “*Emberiza\_GBS5\_XD\_970*”. PC1 explains 1.4% of the variation among individuals and PC2 explains 0.8% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. .... 201

## List of Abbreviations

AFLP	Amplified fragment length polymorphism
ATP	Adenosine triphosphate
CH	Citrinella hybrid
ETC	Electron transport chain
GBS	Genotyping-by-sequencing
HDL	High-density lipoprotein
LH	Leucocephalos hybrid
mtDNA	Mitochondrial DNA
nucDNA	Nuclear DNA
PC	Pure citrinella
PCA	Principal components analysis
PL	Pure leucocephalos
SC	Almost citrinella
SL	Almost leucocephalos
SNP	Single nucleotide polymorphism
ULMM	Univariate linear mixed model
WH	White hybrid
WHAM	Wisconsin hypoalpha mutant
YH	Yellow hybrid

## Acknowledgements

This thesis would never have reached completion without the support of so many wonderful and amazing people. First and foremost, I would like to thank my supervisor—Dr. Darren Irwin—for his guidance and encouragement throughout this project. His patient instruction and insightful feedback have shaped me into the scientist that I am today. I would also like to thank my collaborator—Dr. Alexander Rubtsov—for giving me the opportunity to continue his work in the yellowhammer and pine bunting system, for providing me with endless assistance during sample collection and for taking me to see my birds. Another thank you must go to the Tazeev family in Unguday for welcoming me into their home during my field work and taking the time to teach me about Altai culture. Thank you to my committee members—Dr. Dolph Schluter and Dr. Eric Taylor—for direction and advice that helped make this project what it is. Thank you to the members of the Irwin lab—past and present—who provided me with both professional and emotional support during all my time with them and who put up with my constant shenanigans. In particular, I would like to thank Madelyn Ore who was willing to travel all the way to Siberia with me and who kept me sane while we were there. A huge thank you to my community at the BRC and spread across the Zoology and Botany Departments. Your never-ending curiosity and thirst for knowledge inspired me throughout my degree and continues to inspire me as I move forward on my academic journey. Most importantly, I would like to thank my family and friends who saw me through the ups and downs of this entire process. They celebrated my victories and they commiserated my defeats and, through it all, they never let me forget my dreams. I will be forever grateful for your support and only hope that one day I can repay it.

This thesis also benefitted from the contributions of many different academic organizations. For providing me with research and stipend funding during my degree, I am thankful to the organizations responsible for backing and dispensing the following grants: the NSERC CGSM, the Werner and Hildegard Hesse Research Award in Ornithology, the Werner and Hildegard Hesse Fellowship in Ornithology and NSERC Discovery Grants (RGPIN-2017-03919 and RGPAS- 2017-507830; awarded to Darren Irwin). For providing me with samples to analyze as part of my research, I would like to thank the following museums and their accompanying personnel: The Bell Museum, The Burke Museum of Natural History and Culture, The Field Museum, The State Darwin Museum, The Swedish Museum of Natural History, The Zoological Museum of the Zoological Institute of the Russian Academy of Sciences and the Zoological Museum of the University of Copenhagen.

Last, but not least, I send out one final thank you to the birds. You have captivated the minds of scientists for generations and taught them so much about the world. Thank you for doing the same for me.

# Chapter 1: Introduction

## 1.1 Speciation as a process

Ernst Mayr described species as “groups of actually or potentially interbreeding natural populations that are reproductively isolated from other such groups” (Mayr, 1942). This definition—later dubbed the “Biological Species Concept”—has shaped the field of evolutionary biology and versions of it are still widely used today as scientists continue to investigate and classify the biodiversity of the world. In essence, the Biological Species Concept defines species based on the strength of their reproductive barriers. A reproductive barrier is any morphological, behavioural, physiological or genetic trait that impedes gene flow between two populations either by limiting their ability to interbreed or by causing low fitness of their hybrid offspring (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). Male colouration patterns in avian systems are potential examples of barriers that act by restricting interbreeding between taxa. In these cases, females may discriminate between males based on physical appearance and mate only with individuals who possess the colour patterns unique to their population (e.g. Saetre et al. 1997). Alternatively, adaptation of separate populations to local environments can cause reproductive barriers that function by directly impacting hybrid fitness. Here, hybrids inherit a mixture of parental traits such that they are maladapted to parental environments when compared to “pure” offspring. In this way, hybrids and interbreeding is selected against, potentially leading to increasing reproductive isolation between groups (e.g. Hatfield and Schluter, 1999). Different reproductive barriers can function together, building up over time until gene flow between taxa

ceases entirely and full species status is reached. Speciation researchers strive to understand how different deterministic and stochastic forces drive the evolution of such reproductive barriers as well as how different types of barriers contribute to overall reproductive isolation.

One of the fundamental questions regarding the speciation process is how reproductive barriers develop within a population of interbreeding and undifferentiated organisms. Though models of speciation differ immensely, many rely on an initial period of allopatric isolation where a large, panmictic population is divided into several smaller, isolated populations by a geographic event (Mayr, 1954; 1963; 1970). Examples of such events can include the formation of a mountain, the loss of a land bridge or—perhaps most commonly referenced—the expansion of ice sheets during a glacial maximum. During a period of allopatric isolation, genetic drift and selection cause the two populations to diverge without the homogenizing influence of gene flow (reviewed in Coyne & Orr, reviewed in Price, 2008). With more time spent apart, the isolated populations diverge at a greater number of traits increasing the probability that one of these divergent characteristics will act as a reproductive barrier between the populations upon meeting. Eventually, obstructions may disappear and the diverged populations meet again in a region of secondary contact where the strength of any evolved barriers is definitively tested.

Following the removal of any geographic barriers, the ranges of separated populations will expand and sometimes overlap in a zone of secondary contact. Depending on the amount of divergence that occurred between populations in allopatry, the degree of interbreeding and gene flow within this sympatric zone will vary considerably (reviewed in Coyne & Orr, 2004; Price, 2008). If the amount of divergence between populations is large, reproductive barriers will be strong and gene flow will be greatly reduced or nonexistent. In this situation, speciation has

reached or is near completion and the two taxa will likely endure as independent units well into the future. If the amount of divergence between populations is small, reproductive barriers will be weak and gene flow will be extensive. Here, any divergence between the two taxa may be lost as they merge back into one interbreeding population. Sitting between these two endpoints is a situation where divergence in allopatry produces moderate reproductive barriers such that gene flow occurs between populations, but not to such a degree that population merging is imminent (Barton & Hewitt, 1985; 1989; Hewitt, 1988). In this scenario, divergent populations persist alongside hybrids within regions of secondary contact while opposing evolutionary forces act to push taxa towards or away from complete reproductive isolation (Dobzhansky, 1937; 1940; Blair, 1955).

In the last of the three situations described above, moderately reproductively isolated taxa meet in an area of secondary contact and interbreed to some extent. These areas of range overlap are classified as hybrid zones. Often hailed as windows into the evolutionary process, hybrid zones offer snapshots of the many intermediate steps during speciation (Barton & Hewitt, 1985; 1989; Hewitt, 1988; Gompert et al. 2017). By investigating the population dynamics within these regions, researchers are able to not only evaluate the relative strength of reproductive barriers between taxa, but also assess the nature and importance of specific barriers during the early stages of speciation. For example, observations that females within a new hybrid zone do not discriminate between males of different groups, but that hybrid offspring have lower survival than pure offspring would suggest that early barriers to gene flow are based around low hybrid fitness rather than mating preferences. As such, the strength of hybrid zone studies lies in their

ability to illustrate specific details about the speciation process that are lost when considering taxa that are close to or have reached full reproductive isolation.

Over the past several decades, numerous important discoveries in speciation research have been derived from a wide variety of observational, experimental and modelling approaches. Yet, no single methodology has advanced the field more than the use of genomic sequencing (Toews et al. 2016a; Gompert et al. 2017; Campbell et al. 2018). With techniques such as reduced-representation and whole-genome sequencing, researchers have been able to use millions of genetic loci to objectively evaluate the amount of divergence and gene flow between taxa. Combining these genomic studies with surveys of hybrid zones has proved particularly powerful as it allows for the direct assessment of hybrid ancestry and, by extension, mating dynamics among individuals within a zone of overlap (e.g. Pulido-Santacruz et al. 2018). With this information, researchers can better understand the absolute strength and identity of important reproductive barriers. Additionally, the advent of sequencing technology has also highlighted the importance of analyzing genomic patterns not only in hybrids zones, but also between allopatric zones where the ranges of taxa do not overlap. Examination of these areas can reveal how extensively and in what way populations diverged in isolation and to what degree gene flow occurs across hybrid zones (e.g. Irwin et al. 2018). Thus, by comparing genomic sequences taken from both allopatric and hybrid zones, it is possible to gain a more complete understanding of the speciation process and potentially forecast the fate of hybridizing taxa.

## 1.2 Genomic differentiation during speciation

During a period of isolation, separated populations become increasingly differentiated in their genomes due to input from evolutionary forces such as genetic drift and selection (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). Following secondary contact, some amount of this genetic differentiation is preserved within allopatric zones where populations are more insulated from the homogenizing influence of gene flow. Information regarding the degree and pattern of differentiation between allopatric regions can be used to document trends in genetic divergence between the taxa of interest and also, when compared across systems, across the tree of life. Regarding this latter point, it is often observed that differentiation between closely related species is highly heterogeneous across the nuclear genome (Harr, 2006; Nadeau et al. 2012; Irwin et al. 2018): low overall differentiation across much of the genome is interspersed by areas of high differentiation often called “islands of differentiation”. General questions regarding the origins and drivers of “islands of differentiation” have caused much debate in the sphere of evolutionary research producing several different hypotheses (“divergence-with-gene-flow”: reviewed in Wu, 2001; “recurrent-selection”: Cruickshank & Hahn, 2014; “sweep-before-differentiation”: Irwin et al. 2018). At the same time, researchers have also postulated about regions of the genome that are expected to diverge more readily between taxa. Two genomic regions of particular note are sex chromosomes and mitonuclear genes.

In genomic comparisons of closely related species, it has been consistently observed that sex chromosomes diverge more quickly than autosomes (Thorton & Long, 2002; Borge et al. 2005; Lu & Wu, 2005; Harr, 2006; Ruegg et al. 2014; Sackton et al. 2014). To explain these differences, evolutionary researchers have proposed the “faster X effect” or, similarly, the “faster

Z effect”. The “faster X/Z effect” suggests that beneficial, recessive mutations are more likely to become fixed on the X and Z chromosomes because their beneficial effects are directly expressed within the heterogametic sex (reviewed in Meisel & Connallon, 2013; reviewed in Irwin, 2018). In this way, natural selection drives the accumulation of adaptive genetic differences on sex chromosomes more rapidly than on autosomes. Additionally, X and Z chromosomes are also thought to differentiate more quickly due to their lower effective populations size— $3/4$  the effective populations size of autosomes (Mank et al. 2010; reviewed in Irwin, 2018). A lower effective population size increases the likelihood that weakly deleterious mutations will become fixed due to the relaxation of selective forces and the increased influence of genetic drift. In both of the situations described above, differences between autosomal and X/Z chromosomal divergence are highly dependent on the rate of reproductive success among males and females such that a greater accumulation of mutations on sex chromosomes may not always be observed (reviewed in Irwin, 2018). Nevertheless, because high differentiation is seen so often on sex chromosomes, researchers have proposed that X and Z chromosomes likely contribute disproportionately towards reproductive isolation during the early stages of speciation (Sæther et al. 2007; Presgraves, 2008).

Nuclear genes that encode proteins that function in the mitochondria, termed “mitonuclear genes,” are also proposed to diverge particularly quickly between closely related taxa. In previous genomic work, geneticists have noted that mitochondrial DNA (mtDNA) accumulates genetic differences much faster than nuclear DNA likely as a result of the former’s higher mutation rate (Lynch et al. 2006). This characteristic along with the mitochondrial genome’s short coalescence time has allowed mtDNA markers to be used when tracking

historical relationships between species in phylogenetic studies (Moore, 1995; Zink & Barrowclough, 2008). Interestingly, recent work in the field of mitonuclear ecology suggests that differentiation in mtDNA may also drive similar differentiation of mitonuclear genes (reviewed in Hill, 2019) whose protein products associate closely with products of mtDNA (Calvo & Mootha, 2010; Lotz et al. 2014). Nuclear-encoded proteins play key roles alongside mitochondria-encoded proteins in oxidative phosphorylation and in mtDNA transcription, translation and replication. Close interaction between these protein types create conditions conducive with co-evolution between the mitochondrial and nuclear genomes. Verbal models relating to this topic suggest that divergence at mitochondrial genes could generate selective pressure for compatible divergence (mitonuclear co-divergence) at mitonuclear genes that maintains mitonuclear compatibility (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Evidence for fitness declines caused by mitonuclear incompatibility has been observed between ecologically divergent populations of copepods (Ellison & Burton, 2006; Burton et al. 2013) and hybrid cells derived from different lineages of great ape (Kenyon & Moraes, 1997; Barrientos et al. 1998). In both situations, mismatched combinations of mitochondrial and mitonuclear alleles limit the efficiency of the mitochondria, greatly impacting individual fitness (Kenyon & Moraes, 1997; Barrientos et al. 1998; Ellison & Burton, 2006; Burton et al. 2013). Taken together, rapid differentiation in response to mtDNA divergence as well as direct fitness consequences of mitonuclear incompatibility provide a strong argument for mitonuclear genes acting as potential reproductive barriers between diverging taxa. In this way and considering the ubiquity of mitochondria, differentiation at mitonuclear genes may act as a major driver of speciation across taxa.

### 1.3 Reproductive barriers

Hybrid zones are geographic regions where members of separate taxa have the opportunity to directly interact and interbreed. In this way, hybrid zones act as natural breeding experiments and allow for the assessment of reproductive isolation between groups (Barton & Hewitt, 1985; 1989; Hewitt, 1988; Gompert et al. 2017). Genetic analysis within these regions can be employed to identify regions of the genome that are resistant to gene flow and may mediate barriers between taxa. In addition to identifying the presence and genetic underpinnings of reproductive barriers, examination of breeding and population dynamics within hybrid zones can also hint at the identity of reproductive barriers and the ways by which they limit gene flow between taxa. In the most general sense, reproductive barriers may restrict gene flow between groups in two ways. On the one hand, a reproductive barrier can prevent the formation of a hybrid zygote (prezygotic barrier) or, on the other, it can reduce the fitness of a hybrid individual (postzygotic barrier) (reviewed in Coyne & Orr, 2004; Price, 2008).

Perhaps the most well-known and well-studied class of prezygotic barriers are those related to mate attraction and mate choice or, more generally, to sexual selection. Prezygotic barriers of this type are often male display traits that females use to distinguish between hetero- and con-specific mates and to determine the best quality sire for her offspring (West-Eberhard, 1983; Grant & Grant, 1997; Edwards et al. 2005). Divergence of male traits in conjunction with female preferences for male traits produces strong prezygotic barriers that prevent interbreeding between members of separate taxa even when they interact as sexually mature adults within sympatric zones. So far, researchers have focused much of their attention on how visual signals such as colouration and patterning act as prezygotic barriers (e.g. Saetre et al. 1997; Lukhtanov

et al. 2005; Seehausen et al. 2008; Uy et al. 2009), producing a rich body of literature regarding the evolution and genetic basis of pigment traits across taxa. Contact zones where hybrid individuals display patterning intermediate to parental populations provide particular insight into pigment genes by allowing researchers to correlate phenotypic variation within the system with genetic variation (e.g. Brelsford et al. 2017; Kirschel et al. 2020).

Postzygotic barriers limit gene flow between taxa by affecting fitness of hybrid offspring (reviewed in Coyne & Orr, 2004; Prince, 2008). According to Bateson-Dobzhansky-Mueller models, these barriers are caused by genetic incompatibilities that develop between taxa over the course of genetic divergence (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). In this scenario, genetic differentiation—often during allopatric isolation—produces a distinct set of compatible, co-evolved alleles within each separate population. Hybrids of these diverged taxa inherit a mixture of co-evolved alleles, and some combinations of these alleles may not function efficiently together. As a result of any genetic incompatibilities between divergent alleles, hybrid fitness is reduced contributing towards reproductive isolation between taxa. One excellent example of a potentially strong postzygotic barrier is mitonuclear incompatibility, which was discussed previously in the context of genetic differentiation. In this situation, co-divergence of mtDNA and mitonuclear genes produces a co-evolved collection of alleles within each separate population (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Mismatching of these alleles during hybridization impacts mitochondrial function and hampers the fitness of hybrid individuals (Kenyon & Moraes, 1997; Barrientos et al. 1998; Ellison & Burton, 2006; Burton et al. 2013).

Prezygotic and postzygotic barriers can be linked in that they can both be affected by recombination. Recombination involves the mixing of parental alleles during the production of gametes. Because many prezygotic and postzygotic barriers depend on multiple genes or even multiple traits to function, recombination has the potential to cause barrier breakdown by disassociating important co-evolved alleles (reviewed in Ortiz-Barrientos et al., 2016). Keeping these observations in mind, researchers postulate that the maintenance of reproductive barriers may be directly tied to genomic structures that limit recombination—specifically chromosomal inversions (Noor et al. 2001; Rieseberg, 2001). Fixation of alternative versions of a chromosomal inversion creates a linkage block within each diverged population that is relatively safe from recombination. Here, the genetic differences responsible for reproductive barriers can accumulate and important genetic associations can be maintained over the course of divergence and hybridization. Studies that investigate whether inversions harbour a greater number of genes important to reproductive barriers are limited, but the few that exist provide some evidence for this correlation (Lowry & Willis, 2010; Ayala et al. 2013; Todesco et al. 2020). Further work that examines the link between chromosomal inversions and reproductive isolation is an important avenue for future speciation research.

#### **1.4 The yellowhammer and pine bunting system**

In this thesis, I used a sister pair of avian species—the yellowhammer and the pine bunting—as a model to investigate patterns of genetic differentiation and the evolution of reproductive barriers between recently diverged taxa. The yellowhammer (*Emberiza citrinella*)

and the pine bunting (*Emberiza leucocephalos*) are songbirds of the family Emberizidae whose ranges occupy opposite sides of the Eurasian continent—the former in the west and the latter in the east (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Presumed to have evolved in isolation during the Pleistocene glaciations (Irwin et al. 2009), yellowhammers and pine buntings are divergent in a variety of traits. First, these species are highly divergent in their appearances with yellowhammers possessing bright yellow body plumage and very limited facial markings and pine buntings possessing white body plumage and strong chestnut markings at the brow and throat. As well, in terms of habitat, yellowhammers prefer shrubby open environments while pine buntings prefer coniferous forests (Panov et al. 2003). Finally, these two taxa also contrast somewhat in their songs; however, these comparisons are complicated by the fact that each species sings multiple dialects dependent on their geographic location (Panov et al. 2003; Rubtsov et al. 2017).

Despite their multitude of differences, yellowhammers and pine buntings hybridize extensively within a massive zone of secondary contact—currently measuring 2000 kilometres by 1000 kilometres—in central and western Siberia (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Even more surprising, these species are remarkably similar in their mitochondrial genomes. Genomic analysis found that yellowhammers and pine buntings across this system are only 0.00054% divergent in their mitochondrial DNA which stands in stark contrast with their highly divergent plumage phenotypes (Irwin et al. 2009). Further examination of a subset of nuclear markers recovered moderate nuclear differentiation suggesting that the low mitochondrial differentiation between these species is likely the result of past mitochondrial introgression during secondary contact.

Looking at the yellowhammer and pine bunting sympatric zone in more detail, long-term observation suggests that this area has expanded 1000 kilometres to the east in the last century and 350 kilometres to the west in the last twenty-five years (reviewed in Panov et al. 2003; reviewed in Rubtsov & Tarasov, 2017). Within this area, hybridization is widespread with hybrid proportions of almost 50% in some surveyed locations (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Hybrids are identified based on their intermediate plumage phenotypes at three particular traits of interest: body colour, amount of chestnut colouration at the brow and amount of chestnut colouration at the throat (Panov et al. 2003; Rubtsov & Tarasov, 2017). In extreme cases, a hybrid individual may possess the body colour of one parental species with the full facial markings of the other. The observed phenotypic hybrids likely represent a wide variety of hybrid types including F<sub>1</sub>'s, backcrosses, F<sub>2</sub>'s and further later generation hybrids. Hybrid males are able to hold territories in landscapes where pure individuals are found and there is also some evidence that to suggest that hybrid individuals are fertile and able to attract phenotypically pure mates (reviewed in Panov et al 2003).

Altogether, the above observations create a picture of limited reproductive isolation between yellowhammers and pine buntings that is at odds with the amount of divergence noted between these species' appearances, ecologies, and behaviours, but consistent with their low mitochondrial differentiation. To clarify these ideas, a genomic analysis of this system employing genome-wide genetic markers is warranted. This information can be used to examine patterns of genetic differentiation and hybridization between yellowhammers and pine buntings across allopatric and sympatric zones and create a more concrete picture of the relationship between species. As well, more extensive genetic information will allow for the evaluation of

reproductive barriers—both pre- and post-zygotic—between taxa that, at this point, appear relatively weak. Finally, with this data, it may be possible to connect presumed adaptive mitochondrial introgression with mitonuclear co-introgression which could represent an important driver of hybridization between these taxa.

## **1.5 Thesis objectives and research implications**

The yellowhammer and pine bunting system presents the opportunity to investigate an evolutionary contradiction where two putative species differ greatly in many morphological, ecological and behavioural traits, but differ minimally in mitochondrial DNA and also hybridize extensively (Panov et al. 2003; 2007; Rubtsov, 2007; Irwin et al. 2009; Rubtsov & Tarasov, 2017). In this thesis, I strive to clarify the relationship between yellowhammers and pine buntings and to determine whether their current species designations are justified. I further seek to understand how the relationship between yellowhammers and pine buntings will persist or change in the future. Will full reproductive isolation evolve between these taxa to prevent future hybridization or will they merge together into one group and lose their existing trait differences? I will investigate these questions first by examining genetic differentiation between allopatric populations and then by comparing these genomic patterns to those seen within the hybrid zone.

In chapter two of this thesis, I will examine the genetic relationship between allopatric populations of yellowhammers and pine buntings. By comparing sequences taken from individuals within these regions, I will evaluate the degree of genetic differentiation between groups away from the homogenizing influence of hybridization and gene flow. Previous genomic

work in this system (Irwin et al. 2009) reported negligible mitochondrial differentiation and only moderate nuclear differentiation between taxa in spite of their divergent ecologies, behaviours and appearances (Panov et al. 2003; Rubtsov and Tarasov, 2017). I therefore predict that allopatric yellowhammers and pine buntings will possess some genetic differentiation to account for trait differences, but that differentiation will be relatively low overall likely as a result of gene flow. I will further compare patterns of differentiation in the yellowhammer and pine bunting system to existing studies that similarly examined genomic divergence between closely related species. Consistent with such studies, I expect to observe heterogeneous genetic differentiation across the yellowhammer and pine bunting genomes with peaks of high differentiation occurring within a background of low differentiation. Finally, I will assess mitonuclear interactions between allopatric populations to determine whether interplay between the mitochondrial and nuclear genomes could be mediating gene flow within this system. As described earlier, yellowhammers and pine buntings differ little in their mtDNA likely as a result of mtDNA introgression (Irwin et al. 2009). Based on mitonuclear theory, low hybrid fitness as a result of mismatched mitonuclear and mitochondrial alleles should lead to similar introgression of compatible mitonuclear alleles as part of mitonuclear co-introgression (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). I therefore expect mitonuclear genes to display genetic signals—such as low absolute genetic differentiation between groups—consistent with introgression of mitonuclear alleles from one taxon into the other.

In chapter three, I will investigate patterns of genetic variation across the entire yellowhammer and pine bunting system—focusing primarily on the sympatric zone. Within the sympatric zone, I will assess levels of admixture and hybridization and use this information to

infer the strength of reproductive barriers between taxa. Previous work in this system presents conflicting pictures of the state of reproductive isolation between yellowhammers and pine buntings. Whereas moderate behavioural, ecological, morphological and genetic divergence (Panov et al. 2003; Irwin et al. 2009; Rubsov & Tarasov, 2017) implies that some amount of reproductive isolation exists between these groups, evidence of extensive hybridization within the sympatric zone (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) and of hybrid viability (Lohrl, 1967 cited in Panov et al. 2003) indicates that reproductive barriers may be limited. With these opposing viewpoints in mind, I predict that levels of admixture and hybridization within the sympatric zone will be intermediate suggesting there is some reproductive isolation between groups, but that isolation is far from complete. Furthermore, I will also take advantage of the plumage variation seen among individuals within the sympatric zone to examine the genetic underpinnings of colouration traits in yellowhammers, pine buntings and their associated hybrids. Plumage differences represent one of the strongest candidates for a reproductive barrier in this system (Panov et al. 2003; Rubtsov & Tarasov, 2017). If plumage is important to reproductive isolation between yellowhammers and pine buntings, I predict that the genes controlling plumage phenotypes will resist gene flow within the hybrid zone and potentially reside within regions of high genetic differentiation identified between allopatric populations.

Due to the highly dynamic nature of the yellowhammer and pine bunting system, genomic research comparing these hybridizing taxa has implications in many different aspects of evolutionary research. Broadly, examination of genetic variation within hybrid and allopatric zones will illustrate how opposition between divergent selection and gene flow mediates

differentiation and reproductive isolation between taxa. Results of this kind will add to the growing body of research that aims to understand general patterns of genomic divergence and speciation across the tree of life. More narrowly, the unique history of yellowhammers and pine buntings allows work in this system to further science's understanding of particular evolutionary forces that may greatly influence the progression of speciation. In particular, examination of the hypothesized mitonuclear co-introgression between yellowhammers and pine buntings will underline the importance of mitonuclear interactions in countering genetic differentiation and reproductive isolation instead of driving them as is normally highlighted in theory (e.g. Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Similarly, understanding the genetic underpinnings of plumage patterning in this system will facilitate continued study into the role of colouration traits as reproductive barriers by providing researchers with specific candidate regions on which to focus their research. Thus, when considered as a whole, this thesis has the potential to provide general and novel insight on a variety of evolutionary concepts and to contribute towards a more nuanced and holistic understanding of the speciation process.

## **Chapter 2: Mitonuclear co-introgression opposes speciation between two hybridizing songbirds**

### **2.1 Introduction**

Speciation research is, at its core, an investigation into how the astounding amount of biodiversity in this world came to be. What processes led to the creation and the loss of the different species we see in the past and present? Originally, knowledge about these species was based on what could be seen by the observer, for example distinguishing morphological characteristics, unique mating practices or differences in habitat and range (Coyne & Orr, 2004; Price, 2008). Yet, use of these subjective characteristics may fail to capture cryptic differences between reproductively isolated groups (Herbert et al. 2004; Toews & Irwin, 2008; Funk et al. 2012; Pulido-Santacruz et al. 2018) or may cause a single population to be incorrectly split into two species based on segregating polymorphisms (Mason & Taylor, 2015; Harris et al. 2018; Jones & Weisrock, 2018; Tonzo et al. 2018). Steady advancement paired with price reductions in sequencing technology has introduced genomic data into the realm of speciation research allowing for a closer, more objective examination of what makes a species.

For decades, mtDNA has been a cornerstone of phylogenetics, which is an integral part of speciation research. Due to its uniparental inheritance, mtDNA has one quarter the effective population size and coalescence time of most nuclear markers (Moore, 1995). These characteristics combined with mtDNA's relatively high mutation rate (Lynch et al. 2006) allow researchers to infer species relationships through time by examining the accumulation of genetic

changes in the mitochondrial genome (Moore, 1995; Zink & Barrowclough, 2008). By taking advantage of this information as well as the highly conserved nature and ubiquity of mitochondria, phylogeneticists have been able to confirm prior species designations (Herbert et al. 2004; Kerr et al. 2007) as well as uncover previously unknown biodiversity simply by sequencing the mitochondrial genome of a population in question (Voelker, 2002; McKay & Zink, 2010). For example, highly successful projects such as the “International Barcode of Life Consortium” (International Barcode of Life Consortium, 2021) have used mtDNA to determine the phylogenetic relationships between hundreds of thousands of animal species. Yet, as powerful as mtDNA is when uncovering the underlying relatedness between populations, its low gene content and inheritance as a single linkage block limit the insight it can provide on the speciation process. To tackle mechanistic speciation questions, it is necessary to exploit the more dynamic genetic variation of the nuclear genome.

Techniques such as reduced representation and whole genome sequencing produce genetic datasets that span the entirety of the nuclear genome (Toews et al. 2016a). Using this wealth of information, researchers can focus on the intricacies of the speciation process rather than species classification. In particular, comparisons of genetic differentiation between pairs of recently diverged species have consistently revealed patterns of heterogeneous genetic differentiation across the nuclear genome (Harr, 2006; Nadeau et al. 2012; Irwin et al. 2018) as well as high differentiation on sex chromosomes (Thorton & Long, 2002; Lu & Wu, 2005; Borge et al. 2005; Harr, 2006; Ruegg et al. 2014; Sackton et al. 2014). Identification of such consistent patterns of genomic evolution are important discoveries that allow researchers to test underlying mechanisms driving speciation.

In genomic studies, researchers have repeatedly observed heterogeneous patterns of differentiation when comparing the genomes of recently diverged species (Harr, 2006; Nadeau et al. 2012; Irwin et al. 2018). More specifically, they report “islands of differentiation” where peaks of high relative differentiation are found against a background of low relative differentiation. Explanations for this pattern and its connection to the speciation process are varied, but three of the most well-known hypotheses are: the “divergence-with-gene-flow” model, the “recurrent selection” model and the “sweep-before-differentiation” model. In the “divergence-with-gene-flow” model, peaks in relative differentiation are associated with the loci responsible for incompatibilities between divergent groups that often cause reduced hybrid fitness (reviewed in Wu, 2001). The model predicts that the genomes of interacting taxa will tend to homogenize because of hybridization, but that these loci will cause reproductive isolation and will therefore be resistant to gene flow allowing them to remain highly differentiated despite low differentiation elsewhere in the genome. In the “recurrent-selection” model, selection (either background or directional) at certain genomic locations, first in a common ancestor and then afterwards in its non-admixing daughter populations, reduces absolute nucleotide diversity both between and within species (Cruickshank & Hahn, 2014). These reductions in genetic diversity translate into corresponding increases in relative differentiation and create a heterogeneous landscape with differentiation peaks at selected regions. A similar hypothesis that also invokes multiple bouts of selection, the “sweep-before-differentiation” model, further stipulates that peaks in relative differentiation can also be produced by adaptive selective sweeps between admixing populations followed by further adaptive selection (either parallel or directional) at the same regions in local populations (Irwin et al. 2018). As in the “recurrent-selection” model,

multiple selective events lead to reductions in absolute genetic diversity both between and within populations creating peaks in relative differentiation. Evidence from several systems showing low absolute nucleotide distance associated with peaks in relative differentiation provides tentative support for models invoking repeated bouts of selection (“recurrent-selection”: Cruickshank & Hahn, 2014; “sweep-before-differentiation”: Irwin et al. 2018).

Sex chromosomes are one area of the nuclear genome that commonly shows high peaks in relative differentiation between closely related species. This observation has led many researchers to suggest that sex chromosomes play a disproportionately large role in the speciation process (Sæther et al. 2007; Presgraves, 2008). In particular, Z chromosomes (Borge et al. 2005; Ruegg et al. 2014; Sackton et al. 2014) and X chromosomes (Thorton & Long, 2002; Lu & Wu, 2005; Harr, 2006) show higher levels of divergence on a shorter time scale than autosomal chromosomes. To explain this phenomenon, researchers have proposed the “faster Z effect” or, similarly, the “faster X effect”. The “faster Z/X effect” postulates that, because a beneficial recessive mutation on the Z or X chromosome is immediately exposed to selective forces in the heterogametic sex, fixation of this mutation proceeds much faster than it would have had the mutation appeared on an autosome (reviewed in Meisel & Connallon, 2013; reviewed in Irwin, 2018). As well, the lower effective population sizes of the Z and X chromosomes allow for the fixation of a greater number of slightly deleterious mutations due to less effective purifying selection and a greater role of genetic drift (Mank et al. 2010; reviewed in Irwin, 2018). If any of these fixed mutations contribute to reproductive isolation between groups, their higher probability of fixation has the potential to drive speciation forward. Nevertheless, trends attributed to the “faster Z/X” effect must be treated with a certain amount of

caution as divergence on sex chromosomes is also dependent on the fraction of each sex that contributes to the next generation as well as many other factors (reviewed in Irwin, 2018).

The use of mtDNA and nuclear DNA (nucDNA) in speciation research has greatly advanced the field—the former being instrumental in the demarcation of distinct species and the latter helping to illuminate the processes important to creating these divergent groups. Yet, as interconnected as these two methodologies seem, an increasing number of studies have identified discordance between the signals of differentiation identified when analyzing mtDNA and when analyzing nucDNA (Irwin et al. 2009; Melo-Ferreira et al. 2009; Aboim et al. 2010; Bryson et al. 2012). Discordance between mitochondrial and nuclear genomes is defined as a disparity between the patterns of genetic differentiation found when using each marker type that produces conflicting views of the divergence history between groups (Toews & Brelsford, 2012). Some examples of discordance are attributable to the mitochondrial genome's lower effective population size increasing the speed of lineage sorting (Moore, 1995; Zink & Barrowclough, 2008), leading to more genetic structure in mtDNA than in nucDNA. However, the contrasting pattern, in which nuclear markers show greater amounts of differentiation than mitochondrial markers, requires a different explanation that calls into question previous claims that mtDNA is predominantly neutral in nature.

If we consider that the mitochondrial genome is subject to selective forces as has been suggested in several studies (Ballard & Melvin, 2010; Scott et al. 2010; reviewed in Hill, 2019), it is possible to explain mitonuclear discordance through adaptive introgression between species. In this situation, hybridization and backcrossing introduces an advantageous mutation in the mtDNA of one species into another species leading to introgression of mitochondrial haplotypes.

This results in a lack of genetic structure in the mitochondrial genome, but a retention of nuclear structure that separates the two groups. Further proposed explanations to rationalize mitonuclear discordance include demographic processes, sex-based disparities and hybrid zone movement (Toews & Brelsford, 2012; Sloan et al. 2017). Yet, in cases where the degree of discordance is large, few of these hypotheses are as widely invoked as adaptive introgression. As a result, the idea of selective forces acting on mtDNA has gained traction as an important aspect of speciation especially when it is examined in conjunction with selection acting on nucDNA as part of the emerging field of mitonuclear ecology.

Mitonuclear ecology is the study of how forces acting on the mitochondrial and nuclear genomes interact to influence ecological and evolutionary processes (Hill, 2019). Best known for aerobic respiration, the mitochondrion is the site of the electron transport chain responsible for oxidative phosphorylation. Due to the mitochondrial genome's reduced size of only 37 genes, proper functioning of the electron transport chain as well as transcription, translation and replication of mtDNA is reliant on about 1500 proteins encoded by the nuclear genome (Calvo & Mootha, 2010; Lotz et al. 2014). This interplay between the mtDNA and nucDNA implies co-evolution between the two genomes such that changes in one—by mutation, selection or another evolutionary force—must be balanced by changes in the other to prevent genetic incompatibility and to maintain mitochondrial function (Gershoni et al. 2009; Hill, 2019). Because mitonuclear incompatibilities can act as strong Bateson-Dobzhansky-Mueller incompatibilities, evolutionary biologists have begun relating them to the process of divergence between taxa and to speciation more generally (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Briefly, it has been suggested that mtDNA's high mutation and fixation rate will lead to increased divergence in the

mitochondrial genome between isolated populations and that this divergence will drive similar compensatory divergence at nuclear genes whose products function in the mitochondria (mitonuclear genes). Secondary contact between diverged populations would result in hybrid breakdown due to mismatches between co-evolved combinations of mtDNA and nucDNA leading to reproductive isolation and eventual speciation of the groups. On the flip side, the appearance of a highly adaptive mitochondrial mutation in one population could lead to adaptive introgression between diverging populations as has been suggested in many cases of mitonuclear discordance (Toews & Brelsford, 2012). In this situation, mitochondrial introgression could lead to co-introgression at mitonuclear genes (Beck et al. 2015; Sloan et al. 2017; Morales et al. 2018) due to selection against incompatible combinations of mtDNA haplotypes and nuclear alleles. This co-introgression has the potential to oppose speciation between groups. Interestingly, suggestions of mitonuclear co-introgression between diverging groups align well with the “sweep-before-differentiation” model discussed earlier when describing islands of differentiation (Irwin et al. 2018). In this situation, mitonuclear co-introgression would represent the shared selective sweep between admixing populations and subsequent selection at mitonuclear genes in later isolated populations would create differentiation peaks at these genomic regions. Whether leading to co-divergence or co-introgression, interactions between mtDNA and nucDNA have the potential to greatly impact the speciation process. Additional research is needed to provide further empirical support or a lack thereof for the link between mitonuclear interactions and speciation as well as to investigate the generality of this link throughout the tree of life.

Here, I focus my attention on the yellowhammer (*Emberiza citrinella*) and the pine bunting (*Emberiza leucocephalos*)—an avian species pair that sits at the nexus of speciation and

mitonuclear theory. Members of the family Emberizidae, yellowhammers and pine buntings are thought to have diverged during the Pleistocene glaciations when populations were separated on either side of Eurasia by a large area of unsuitable habitat (Irwin et al. 2009). Currently, these taxa are moderately divergent in song and ecology, but highly divergent in plumage traits (Panov et al. 2003; Rubtsov & Tarasov, 2017). Despite their differences, yellowhammers and pine buntings hybridize extensively in a large, dynamic contact zone in central and western Siberia where they produce offspring with variable plumage phenotypes (Panov et al. 2003; 2007; Rubtsov, 2007; Irwin et al. 2009; Rubtsov & Tarasov, 2017). This high degree of hybridization calls into question the species designations held by these taxa for decades and has prompted genetic research of the system to further investigate the relationship between these groups. Genomic work has identified mitonuclear discordance between allopatric yellowhammers and pine buntings (Irwin et al. 2009), with almost no mtDNA divergence but moderate divergence in AFLP markers seen between these groups. To explain these results, Irwin et al. (2009) suggested that the mtDNA of one species introgressed into the other during a previous selective sweep, and this hypothesis was supported by several statistical tests performed on the mtDNA haplotype network. Based on mitonuclear theory, such mtDNA introgression could select for co-introgression at mitonuclear genes if sizeable genetic differentiation had developed between yellowhammers and pine buntings (Sloan et al. 2017; Hill, 2019) which is implied by their divergent behaviour, ecology and appearance. Mitonuclear co-introgression and the resulting lack of mitonuclear incompatibility could facilitate the continued hybridization seen between yellowhammers and pine buntings and prevent the build-up of reproductive isolation between these taxa (Hill, 2019). However, further genomic research is needed to address this idea.

In this study, I performed reduced representation genome sequencing on DNA samples taken from allopatric yellowhammers and pine buntings to address questions surrounding the degree of genetic differentiation between taxa as well as the potential influence of mitonuclear co-introgression on the process of divergence. First, I tested whether genetic differentiation across the nuclear genome resembled what was seen in mtDNA (virtually none) or what was seen in the analysis of nuclear AFLP markers (moderate; Irwin et al. 2009). Highly differentiated allopatric populations would provide support for present yellowhammer and pine bunting species designations and offer a counterpoint for the extensive hybridization observed within the sympatric zone. Second, I characterized patterns of differentiation across the nuclear genome and compared them to patterns of differentiation between other pairs of sister species (e.g. “islands of differentiation”). With this information, I drew upon evolutionary theory to identify potential forces and processes that could be driving or hampering divergence between these taxa. Finally, I identified windows of nuclear introgression between yellowhammers and pine buntings and tested whether these introgression windows corresponded to mitonuclear gene locations thus providing indirect support for mitonuclear co-introgression. Evidence of co-introgression could offer a possible explanation for the high degree of hybridization seen between yellowhammers and pine buntings as well as potentially implicate this process as a homogenizing force that can counter divergence and speciation.

## 2.2 Methods

### 2.2.1 Sampling

I included a total of 335 blood and tissue samples spanning the breadth of Palearctic Eurasia in this study. Of these samples, 135 were phenotypic yellowhammers, 112 were phenotypic pine buntings, 74 were phenotypic hybrids and 14 were other members of the Emberizidae family (one yellow-breasted bunting [*Emberiza aureola*], one corn bunting [*Emberiza calandra*], one meadow bunting [*Emberiza cioides*], one ortolan bunting [*Emberiza hortulana*], four white-capped buntings [*Emberiza stewarti*] and six ciril buntings [*Emberiza cirilus*]) acting as outgroups (Figure 2.1A; Table 2.1; Supplementary Table 2.1). An additional 34 samples (nineteen phenotypic yellowhammers, five phenotypic pine buntings, eight phenotypic hybrids, one ortolan bunting and one black-faced bunting [*Emberiza spondocephala*]) were obtained as part of this research, but sequences from these individuals did not meet quality standards and were discarded before completing the statistical analysis. Genetic information from about half of the samples described here was previously evaluated in Irwin et al. (2009) using an ALFP analysis and minimum spanning haplotype networks of the mitochondrial gene ND2 and the sex-linked gene CH1DZ. A total of 166 yellowhammer, pine bunting, hybrid and outgroup samples were sequenced and analyzed for the first time as part of this study.

When possible, body measurements and photographs were taken of live birds and museum skins. At the same time, yellowhammer, pine bunting and hybrid males were scored phenotypically and sorted into phenotypic classes based on the protocols presented in Panov et al. (2003) and Rubtsov & Tarasov (2017). Briefly, each individual was given a score from 0-7

for background plumage colour, the amount of chestnut (vs. yellow or white) at the brow and the amount of chestnut (vs. yellow or white) at the throat. For background colour, birds were assessed on the strength of yellow—ranging from bright yellow to pale yellow to pure white—in head and body plumage that did not show brown or black streaking. Phenotypic scores of 0 are consistent with a phenotypically pure yellowhammer and scores of 7 are consistent with a phenotypically pure pine bunting. Phenotypic classes include: pure citrinella (PC), almost citrinella (SC), citrinella hybrid (CH), yellow hybrid (YH), white hybrid (WH), leucocephalos hybrid (LH), almost leucocephalos (SL) and pure leucocephalos (PL) (Rubtsov & Tarasov, 2017; Figure 2.1B). Unless stated otherwise, I grouped the PC and SC phenotypic classes and the SL and PL phenotypic classes together and treated them as phenotypic yellowhammers and phenotypic pine buntings respectively in subsequent analyses. Similarly, I grouped the CH, YH, WH and LH phenotypic classes together as phenotypic hybrids. I further divided phenotypic yellowhammers and phenotypic pine buntings into three geographic classes: sympatric, near-sympatric and allopatric. A sympatric designation indicates that the individual was sampled within the yellowhammer and pine bunting sympatric zone shown in Irwin et al. (2009) and pictured in Figure 2.1A. Near-sympatric individuals were sampled outside of, but close to the border of the sympatric zone such that a recent expansion of the sympatric zone might capture these areas. The possibility of an expansion is supported by the highly dynamic nature of this sympatric zone (Irwin et al. 2009; Rubtsov & Tarasov, 2017) and by a small number of hybrids in my data set that occurred within the near-sympatric geographic area. “Allopatric” individuals were sampled far (roughly more than 400 kilometres) from the borders of the sympatric zone.

## **2.2.2 DNA sequencing and identification of single nucleotide polymorphisms**

### **2.2.2.1 DNA extraction and genotyping-by-sequencing**

DNA was extracted from blood and tissue samples using a standard phenol-chloroform method. The resulting DNA pellets were then resuspended in 35-50  $\mu$ l of 1 $\times$  TE buffer and stored at 4°C until sequencing. I quantified the DNA concentration of each extract using a Qubit ssDNA Broad Range assay kit and a Qubit fluorometer (Invitrogen) according to the manufacturer's directions.

Following extraction, I divided the DNA samples into four genotyping-by-sequencing (GBS) libraries (Elshire et al. 2011). The GBS libraries were prepared as per the protocol described by Alcaide et al. (2014) and the modifications specified by Geraldès et al. (2019). Unlike in Geraldès et al. (2019), I maintained the fragment size at 300-400 bp during the size selection step of library preparation to be consistent with the original protocol (Alcaide et al. 2014). Paired-end sequencing of GBS libraries was completed by Genome Québec using the Illumina HiSeq 4000 system producing more than 1.2 billion 150 bp reads across the four libraries.

### **2.2.2.2 Genotyping-by-sequencing data filtering**

I processed my GBS reads using the scripts described in Irwin et al. (2016; 2018). To summarize, GBS reads were demultiplexed using a custom perl script designed by Baute et al. (2016). Demultiplexing involved sorting GBS reads based on barcode sequence, removing the barcode and adaptor sequences from reads and discarding any reads shorter than 30 bp in length. Next, GBS reads were trimmed for quality using Trimmomatic version 0.36 (Bolger et al. 2014)

with the parameters: TRAILING:3, SLIDINGWINDOW:4:10, MINLEN:30. Trimmed reads were aligned to the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010) using the program BWA-MEM (Li & Durbin, 2009) and a BAM file of this information was created for each individual using a combination of Picard (<http://broadinstitute.github.io/picard/>) and SAMtools (Li et al. 2009). Following alignment, each individual's BAM file was converted into a GVCF file using the HaplotypeCaller command as part of GATK version 3.8 (McKenna et al. 2010). I processed the resulting GVCF files in two separate ways to create a single genome-wide “variant site” VCF file and a series of chromosome-specific “info site” VCF files.

To create the genome-wide “variant site” VCF file, I used the GenotypeGVCFs command in GATK version 3.8 to identify and isolate single nucleotide polymorphisms (SNPs) among individuals. This command also converted all the variant site information into a single VCF file encompassing the entire genome. Using a combination of VCFtools (Danecek et al. 2011) and GATK version 3.8, I filtered the variant site VCF file to remove indels and non-biallelic SNPs. As well, I also applied the following quality filters— $QD < 2.0$ ,  $MQ < 40.0$ ,  $FS > 60.0$ ,  $SOR > 3.0$ ,  $ReadPosRankSum < -8.0$ —where loci with quality values below these thresholds were discarded. Finally, loci with more than 60% missing genomic data were removed and the VCF file was converted into an allele frequency format. The average coverage of variable sites in this VCF file was 16.59.

To convert GVCF files into “info site” VCF files, I once again employed the GenotypeGVCFs command in GATK version 3.8 with the addition of the -allSites and -L flags to retain invariant sites and split the information into chromosome-specific files respectively.

Info site VCF files were then filtered using VCFtools and GATK version 3.8 to remove indels, sites with more than two alleles and sites with more than 60% missing genomic data. Use of these filters simplified calculations in downstream analyses and also ensured that these analyses were restricted to sites with sufficient data. Custom perl scripts (Irwin et al. 2016) were further used to remove loci with a MQ lower than 20 and sites with a heterozygosity greater than 60% to avoid potential paralogs. Finally, I converted each chromosome specific VCF file into allele frequency format.

### **2.2.3 Statistical Analysis**

The statistical analyses in this chapter focus exclusively on genomic information obtained from allopatric yellowhammers, allopatric pine buntings and outgroups. Genomic information from yellowhammers, pine buntings and hybrids sampled in the near-sympatric and sympatric geographic zones will be analyzed separately in Chapter 3 of this thesis. All analyses were completed using R version 3.6.2 (R Core Team, 2014).

#### **2.2.3.1 Variant site analyses**

I loaded the genome-wide “variant site” VCF file into R where it was processed and analyzed using modified versions of the R scripts described in Irwin et al. (2018). In total, 882,227 SNPs that were variable among *Emberizidae* individuals passed filtering thresholds with 374,780 of those SNPs being variable between allopatric yellowhammers and pine buntings.

Initially, I calculated a value for sample size, allele frequency and Weir and Cockerham’s  $F_{ST}$  (Weir & Cockerham, 1984) comparing allopatric yellowhammers (n = 53) with allopatric pine buntings (n = 42) for each of the 374,780 SNPs identified between them. Genetic

differentiation between groups was then visualized using a principal components analysis (PCA) generated with the `pca` command and the `svdImpute` method to account for any missing genomic data in the `pcaMethods` package (Stacklies et al. 2007). To examine the spread of variant sites across the genome and identify areas of high genetic differentiation, I also graphed the  $F_{ST}$  values of 349,807 SNPs—each assigned to one of thirty autosomes or one sex chromosome—as a Manhattan plot using the package `qqman` (Turner, 2018). The remaining SNPs that varied between these allopatric populations did not possess known genomic locations and, therefore, could not be included in the Manhattan plot.

### 2.2.3.2 Differentiation across the genome

To more thoroughly investigate genomic differentiation between allopatric yellowhammers ( $n = 53$ ) and allopatric pine buntings ( $n = 42$ ), I performed further analysis on both variant and invariant loci contained within “info site” VCF files using R scripts described in Irwin et al. (2018). Any filtering or processing steps were kept constant with previous variant site analyses to ensure result consistency.

I calculated the average sample size, allele frequency, Weir and Cockerham’s  $F_{ST}$ , between-group nucleotide diversity ( $\pi_B$ ) and within-group nucleotide diversity ( $\pi_W$ ) in nonoverlapping windows of available GBS sequence data across each chromosome. The first window was positioned at the “start” of each chromosome as described in the zebra finch reference genome. Unlike in Irwin et al. (2018), I used a window size of 2000 bp of sequenced data rather than 10,000 bp of sequenced data to visualize narrow peaks in relative and absolute differentiation within our dataset. For this study, a new R script was created to calculate Tajima’s

D (Tajima, 1989) in nonoverlapping 2000 bp windows consistent with the process previously mentioned. Values of Tajima's D were used to identify areas of the genome where patterns of variation in allopatric yellowhammers and allopatric pine buntings deviated from models of neutrality. Significantly negative Tajima's D implies that there are more rare alleles in a population than expected under neutrality likely as a result of a selective sweep or a population expansion following a bottleneck. Significantly positive Tajima's D suggests that there are fewer rare alleles in a population than expected under neutrality potentially stemming from balancing selection or a rapid population contraction.

### **2.2.3.3 Phylogenetic comparison with outgroups**

Using the "info site" VCF files and the protocol described previously, I calculated 2000 bp windowed averages of  $\pi_B$  between allopatric yellowhammers and pine buntings as well as among these focal species and six outgroups (*Emberiza aureola*, *Emberiza calandra*, *Emberiza cioides*, *Emberiza circlus*, *Emberiza hortulana* and *Emberiza stewarti*). A list of average  $\pi_B$  values for each taxon pair was converted into a distance matrix and used to create an unrooted neighbour-joining tree. This tree was constructed using the ape R package (Paradis & Schliep, 2019) and the BioNJ algorithm (Gascuel, 1997) with *Emberiza aureola* set as the outgroup (Alström et al. 2008). A bootstrapping approach was not applied in this analysis because the purpose of the tree was just to get a rough estimate of species relationships.

### **2.2.3.4 Signals of mitonuclear co-introgression**

To test for signals of introgression between yellowhammers and pine buntings at mitonuclear genes, I identified putative windows of introgression across the genome and

correlated them with mitonuclear gene locations. If mitonuclear genes occurred within introgression windows statistically more often than would be expected due to chance, it would provide some support for co-introgression of mitochondrial DNA and mitonuclear genes as suggested by mitonuclear theory (Hill, 2019).

The first step in this process involved creating a list of mitonuclear genes to analyze for signals of introgression. To focus on strong, detectable signals of introgression, I chose mitonuclear genes that were protein-coding and whose protein products interacted directly with mitochondrial DNA or an immediate product of the mitochondrial genome (ie. protein, mRNA, tRNA and rRNA). In this way, any change in mitochondrial DNA including those caused by introgression would have strong, direct effects on the chosen mitonuclear genes potentially increasing selective pressure for mitonuclear co-introgression of compatible alleles (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Mitonuclear genes that met these conditions included genes with protein products that were part of ATP synthase or the first, third and fourth complex of the electron transport chain, that were assembly and ancillary proteins involved in the formation of the electron transport chain, or that were part of the specific transcription, translation or DNA replication machinery within the mitochondria. I created this list of genes using information from Figure 2.3 and Table 2.1 in Hill (2019). After removing any genes that were not annotated in the zebra finch reference genome or that lacked a specific location on the reference genome, a total of 134 mitonuclear genes remained for analysis (Supplementary Table 2.2).

Next, I identified putative windows of introgression across the genome using the “info sites” VCF files and the 2000 bp windowed statistics calculated previously. In this study,

windows of introgression were classified as those possessing both low average Tajima's  $D$  and low average  $\pi_B$  where the Tajima's  $D$  statistic was used to distinguish between DNA sequences evolving neutrally and those evolving under non-neutral processes such as selection (Tajima, 1989). For this analysis, the quantitative criteria for a putative introgression window (hereafter referred to simply as "introgression windows" although they should be considered putative) were an average Tajima's  $D$  within the lowest 5% of the available windowed averages and an average  $\pi_B$  within the lowest 30% of the available windowed averages. Out of the 7187 windows described across the genome, 244 introgression windows were identified for yellowhammers, 222 introgression windows were identified for pine buntings and 71 windows were shared between the two species.

Following window categorization, I used a custom R script to identify how many mitonuclear genes occurred within introgression windows. To do this, I determined the genomic centre position of each mitonuclear gene as well as the average genomic position of each of the 7187 available windows. I then calculated the absolute difference between mitonuclear gene centres and the average window positions. Mitonuclear genes were assumed to occur within whatever window minimized this difference. With this information, I was able to determine the number of mitonuclear genes occurring within the introgression windows of each species.

I compared the number of mitonuclear genes within introgression windows to what would be expected if these genes were distributed randomly across the genome using two separate two tailed binomial tests—one for yellowhammers and one for pine buntings. Because genes are often not distributed equally across the genome (International Chicken Genome Sequencing Consortium, 2004), I further conducted a Fisher's Exact test for both

yellowhammers and pine buntings to compare the number of mitonuclear genes found in introgression windows to the total number of protein-coding genes found in these windows.

## 2.3 Results

Following quality filtering, I identified 882,227 variant sites across the entire nuclear genome and 13,703,455 invariant sites across thirty autosomes and one sex chromosome within my genotyping-by-sequencing (GBS) dataset of 335 *Emberizidae* individuals. Variant sites were retained as a separate dataset to be analyzed singly while invariant sites were grouped together with variant sites in 2000 base pair windows across the nuclear genome for analysis. A total of 7187 windows were classified across 30 autosomes and 1 sex chromosome with each window covering an average distance of about 139 kilobases of the genome.

### 2.3.1 Phylogenetic comparison with outgroups

A neighbour-joining tree of average  $\pi_B$  values between allopatric yellowhammers, allopatric pine buntings and six outgroups (Figure 2.2) depicted similar species relationships as were identified in other phylogenies of *Emberizidae* using a much smaller subset of genetic loci (Alström et al. 2008; Irwin et al. 2009). The yellowhammers and the pine bunting were shown as sister species and, as a pair, were most closely related to the white-capped bunting (*Emberiza stewarti*). This relationship was well supported as indicated by the long branch length between the white capped bunting node and the yellowhammer/pine bunting node. Continuing outward, the ciril bunting (*Emberiza cirilus*) was the next most closely related species to the yellowhammer/pine bunting pair, followed by the ortolan bunting (*Emberiza hortulana*), the

meadow bunting (*Emberiza cioides*) and the corn bunting (*Emberiza calandra*). The yellow-breasted bunting (*Emberiza aureola*) was the most divergent group as it has a long branch from the clade containing all the other species considered.

### 2.3.2 Overall genetic differentiation

Out of the 882,227 single nucleotide polymorphisms (SNPs) documented among all *Emberizidae* individuals, 374,780 of them were variable within the yellowhammer and pine bunting system. Average genome-wide  $F_{ST}$  between allopatric groups based on these loci was very low at 0.0232.

Despite the low average  $F_{ST}$ , a principal components analysis (PCA) of variant site information separated allopatric yellowhammers and pine buntings into tight genetic clusters (Figure 2.3). In this analysis, PC1 explained 3.6% of the variation seen among individuals while PC2 explained 2.9% of the variation. There were two pine bunting outliers along PC1, while allopatric yellowhammers and pine buntings separated into distinct clusters mainly along PC2. Further investigation into these pine bunting outliers revealed that they were both males from the same location. It is possible that a familial relationship between these two birds is responsible for their position as outliers. Removal of one outlier individual from this PCA caused the other to fall into the pine bunting cluster, but did reveal a further yellowhammer outlier (Supplementary Figure 2.1). Removal of this yellowhammer outlier in addition to one member of the pine bunting outlier pair in turn revealed another yellowhammer outlier (Supplementary Figure 2.2). It is likely that there are several birds with familial relationships in my dataset due to the nature of sample collection and these relationships are responsible for the presence of the different

outliers. Nevertheless, despite these observations and outliers, the distinct genetic clusters of allopatric yellowhammers and pine buntings remained intact in all the PCAs considered.

### 2.3.3 Differentiation across the genome

To characterize patterns of differentiation across the nuclear genome, I plotted  $F_{ST}$  value against genomic location for each of 349,807 SNPs that had known genomic positions and were variable between allopatric groups (Figure 2.4). Patterns of relative differentiation were highly heterogeneous with peaks in  $F_{ST}$  seen on most of the larger autosomes. More specifically, I found large peaks in  $F_{ST}$  on chromosomes 1, 2, 3, 4, 5, 7, 10, 12, 14, 15 and 20. Chromosome Z also showed a very large peak in  $F_{ST}$  with several SNPs possessing values close to one. In fact, average  $F_{ST}$  for the Z chromosome was 0.1246—more than five times larger than what was seen genome-wide.

To more thoroughly investigate these peaks in relative differentiation as well as delve deeper into the mechanisms controlling differentiation among and within groups, I combined variant and invariant sites into genomic containing 2000 bp of obtained sequence. A plot showing windowed averages of  $F_{ST}$  across each chromosome depicted the  $F_{ST}$  peaks identified earlier in a background of extremely low  $F_{ST}$  (Figure 2.5). High relative differentiation on the Z chromosome was separated into several peaks of varying size across the chromosome.

Relative differentiation ( $F_{ST}$ ) can be further broken down into its component parts—between-group ( $\pi_B$ ) and within-group ( $\pi_W$ ) nucleotide differentiation. By examining windowed averages of  $\pi_B$  and  $\pi_W$  across the genome, I can begin to unravel the forces generating the observed islands of differentiation. Patterns of between-group and within-group differentiation

were heterogenous across the genome and comparable in magnitude (Figure 2.5). Corresponding peaks in  $\pi_B$  and  $\pi_W$  were seen in several locations—particularly on chromosomes 6, 14 and 28. I found that values of absolute differentiation were low overall with an average genome-wide  $\pi_B$  of 0.0041 and an average genome-wide  $\pi_W$  of only slightly less, 0.0040, for both allopatric yellowhammers and pine buntings.

Because between-group and within-group nucleotide differentiation are so intimately related in their evolution and calculation, it is expected that windowed averages of these two statistics will show a highly positive relationship. In accordance with this expectation,  $\pi_B$  and  $\pi_W$  averages followed a 1:1 association line in the majority of windows across the nuclear genome (Figure 2.6). However, some genomic windows showed particularly low  $\pi_W$  compared to  $\pi_B$ ; these were the regions with high  $F_{ST}$  (Figure 2.6). Additionally, when distinguishing between the prevailing hypotheses used to explain high  $F_{ST}$  regions, I found a weak negative correlation between windowed averages of  $F_{ST}$  and  $\pi_B$  (Spearman's Rank Correlation: -0.1196,  $p < 2.2 \times 10^{-16}$ ; Figure 2.7). This result is consistent with models that invoke repeated selective events to explain islands of differentiation (Cruickshank & Hahn 2014; Irwin et al. 2018) as a divergence-with-gene flow model would predict a positive correlation between windowed averages of  $F_{ST}$  and  $\pi_B$  (reviewed in Wu, 2001).

Finally, when looking at patterns of Tajima's D, I found that averages varied immensely between windows across the genome, but that values were usually negative (Figure 2.5). Tajima's D is often used to indicate when patterns of differentiation differ from neutral expectations. A significantly negative value may indicate adaptive processes such as a recent

selective sweep or demographic processes such as a population expansion following a bottleneck (Tajima, 1989). The average genome-wide Tajima's  $D$  was similar between populations: -1.377 for allopatric yellowhammers and -1.335 for allopatric pine buntings. Peaks in Tajima's  $D$  where values were greater than one tended to be narrow and infrequent on autosomes. However, I saw a slight increase in both peak frequency and width on chromosome  $Z$ . These positive peaks may represent areas of the genome that are subject to balancing selection of multiple allele variants (Tajima, 1989).

### **2.3.4 Signals of mitonuclear co-introgression**

A total of 7187 windows consisting of 2000 bp of sequenced sites were classified across the nuclear genomes of allopatric yellowhammers and pine buntings. I identified a subset of introgression windows for each species that possessed an average  $\pi_B$  within the lowest 30% of the available windowed averages and an average Tajima's  $D$  within the lowest 5% of the available windowed averages. For yellowhammers, I identified 244 introgression windows and for pine buntings I identified 222 introgression windows (Table 2.2). A total of 71 introgression windows were shared between the two groups. Average values of  $\pi_B$  and Tajima's  $D$  in yellowhammer and pine bunting introgression windows were 0.0016 and -2.3751 and 0.0019 and -2.3369 respectively.

To examine signals of introgression at mitonuclear genes, I associated the introgression windows of each species with the genomic locations of a curated list of 134 mitonuclear genes. Nine mitonuclear genes appeared within yellowhammer introgression windows—6.7% of the genes considered (Table 2.2). This finding was significant in a two-tailed binomial test ( $p =$

0.04952) suggesting that mitonuclear genes appeared in yellowhammer introgression windows more often than would be expected if they were assigned to windows randomly. Conversely, this finding was not statistically significant in a Fisher's Exact test ( $p = 0.1311$ ) which takes into account the differing densities of genes across the genome. Four mitonuclear genes were located within pine bunting introgression windows—3.0% of the genes considered. This result was statistically insignificant in both a two-tailed binomial test ( $p = 1$ ) and a Fisher's Exact test ( $p = 1$ ) indicating that mitonuclear genes did not appear in pine bunting introgression windows more often than would be expected due to chance. Overall, the significant signal of introgression in yellowhammers and insignificant signal of introgression in pine buntings suggests that mitonuclear introgression could be biased in the direction of pine buntings into yellowhammers if it is present.

To more thoroughly examine the dynamics and drivers of mitonuclear co-introgression in the yellowhammer and pine bunting system, I summarized the functions of the mitonuclear genes appearing in each species' introgression windows. The nine mitonuclear genes appearing in yellowhammer introgression windows were: APOPT1, COX5A, COX17, MRPL1, MRPL27, MRPL32, NDUFC1, mtSSB, UQCR11 (Table 2.3). Three of these genes code for protein subunits in the mitochondrial ribosome, three code for structural subunits of the electron transport chain (ETC), two code for assembly factors of the ETC and one codes for a single-stranded DNA-binding protein involved in mtDNA replication. All putatively introgressed genes appear on separate autosomes except for two genes that appear on chromosome 4. Interestingly, three of the five putatively introgressed genes associated with the ETC are specifically associated with complex IV or cytochrome c oxidase. As well, after ranking each of the 244 yellowhammer

introgression windows based on windowed Tajima's *D*—where a ranking of one indicates the window with the lowest Tajima's *D* value out of all the introgression windows—I noted that mitonuclear genes coding for structural components of the ETC appeared in lower ranking windows than other putatively introgressing genes. One exception is COX17—an assembly factor for ETC complex IV—which appeared in the lowest ranking window of the nine mitonuclear genes that putatively introgressed in yellowhammers.

The four mitonuclear genes appearing in pine bunting introgression windows were: ATP5H5I, COX5A, MRPL2 and NDUFB4 (Table 2.4). All four genes appear on separate autosomes with three of these genes coding for structural subunits of the ETC and one coding for a protein subunit of the mitochondrial ribosome. The COX5A gene, which codes for a structural subunit of ETC complex IV, was found in both yellowhammer and pine bunting introgression windows. As was seen with the yellowhammers, ranking of pine bunting introgression windows based on windowed Tajima's *D* found that mitonuclear genes coding for structural components of the ETC were in lower ranked windows than the gene coding for a ribosomal subunit. The ATP5H5I gene which codes for a subunit of ETC complex V or ATP synthase was found in the lowest ranked window of the mitonuclear genes thought to have introgressed into pine buntings from another source.

## **2.4 Discussion**

In this study, I investigated the evolutionary relationship between yellowhammers and pine buntings by measuring genetic differentiation between allopatric populations of these

species. An interesting aspect of this system is that separate taxa are highly divergent in plumage traits and moderately so in ecological and song traits (Panov et al. 2003; Rubtsov & Tarasov, 2017), but show little to no mitochondrial DNA differentiation (Irwin et al. 2009). Previous genomic work reported that yellowhammers and pine buntings are sister species (Alström et al. 2008; Irwin et al. 2009), but these claims have been questioned by those suggesting that past mitochondrial introgression may have eliminated genetic differences at mitochondrial markers and produced the incorrect appearance of a close species relationship between taxa that are actually not sister species in terms of nuclear DNA (Rubtsov & Opaev, 2012). Instead, based on cluster analyses of song and morphological characteristics, Rubtsov & Opaev (2012) hypothesized that the yellowhammer and pine bunting could be more closely related to the ciril and white-capped bunting respectively. Nevertheless, using hundreds of thousands of genome-wide markers, I found the same sister relationship between yellowhammers and pine buntings using genome-wide nuclear DNA that was described in previous research (Alström et al. 2008; Irwin et al. 2009). Similarly, I also recovered phylogenetic relationships between the yellowhammer, pine bunting, ciril bunting, white-capped bunting and other members of *Emberizidae* that are consistent with prior study (Alström et al. 2008).

The close relationship between yellowhammers and pine buntings as well as their large and dynamic hybrid zone seen in western and central Siberia (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) could put into question the genetic distinctiveness of these taxa. Gene flow within and out of the hybrid zone has the potential to homogenize the nuclear genome across these species' ranges. Yet, my analysis revealed that allopatric yellowhammers and pine buntings are separable into tight, well differentiated genetic clusters. This finding provides some

support for the classification of yellowhammers and pine buntings outside of the contact zone as separate taxa. A full genomic analysis of genetic differentiation between individuals within the sympatric zone is needed to further test this idea and will be addressed in Chapter 3.

Though genetically distinct, the average  $F_{ST}$  between allopatric yellowhammers and pine buntings was found to be very low when compared to other avian sister pairs (yellowhammer and pine bunting: 0.0232; pied flycatcher and collared flycatcher: 0.357 - Ellegren et al. 2012; Myrtle warbler and Audubon's warbler: 0.077-0.106 - Irwin et al. 2018). This low average suggests low overall nuclear divergence between taxa and contradicts the moderate  $F_{ST}$  averages reported from an AFLP analysis of the same populations: 0.078 based on allele frequencies and 0.140 based on band frequencies (Irwin et al. 2009). However, further investigation found that patterns of relative differentiation were highly heterogeneous across the nuclear genome with numerous peaks in  $F_{ST}$  on various chromosomes and several SNPs with  $F_{ST}$  values very close to one. It is possible that the previous AFLP analysis captured a disproportionate number of loci within these differentiation peaks thus inflating  $F_{ST}$  estimates. The comparison between these two studies suggests that caution is needed when interpreting genome-wide averages based on a relatively small number of loci. When analyzing datasets that include only a small portion of the genome—such as those obtained from AFLP procedures—these singular values can be biased towards a limited number of loci. More importantly,  $F_{ST}$  averages are unable to capture the variable nature of genetic landscapes especially when patterns of differentiation are heterogeneous.

Peaks in  $F_{ST}$  between allopatric yellowhammers and pine buntings appeared within a largely undifferentiated background on many of the larger autosomes and most significantly on

the Z chromosome. This genomic pattern is consistent with “islands of differentiation” that have been described in numerous other studies comparing closely related taxa (Harr, 2006; Nadeau et al. 2012; Irwin et al. 2018). As mentioned earlier, explanations for “islands of differentiation” tend to associate these regions with different forms of selection. The “divergence-with-gene-flow” model attributes islands of differentiation to divergent selection that causes resistance to gene flow (reviewed in Wu, 2001) while the “recurrent-selection” (Cruickshank & Hahn, 2014) and “sweep-before-differentiation” (Irwin et al. 2018) models invoke repeated selective events at these regions first in an ancestral population and then in its daughter populations. To distinguish divergence-with-gene-flow from the other hypotheses, I compared patterns of relative differentiation to patterns of absolute nucleotide differentiation between groups. In the former scenario, peaks in  $F_{ST}$  should be associated with similar peaks in  $\pi_B$  while in the latter scenario, peaks in  $F_{ST}$  should be associated with declines in  $\pi_B$  as well as in  $\pi_W$ .

In the yellowhammer and pine bunting system, I noted high congruence between  $\pi_B$  and  $\pi_W$  as windowed values of these two statistics followed a 1:1 association line. Such equivalence between  $\pi_B$  and  $\pi_W$  suggests unheeded gene flow between these taxa or, alternatively, a recent split from a common ancestor. However, a subset of windows possessing high  $F_{ST}$  and that were within “islands of differentiation” deviated from this 1:1 association suggesting that they were somewhat resistant to gene flow. It is unlikely these deviations were caused by genetic drift or an extended period of geographic separation between groups as both situations would result in the majority of windows deviating slightly from the 1:1 line rather than a small subset of window deviating significantly. Instead, this pattern suggests that some form of selection is acting to lower  $\pi_W$  relative to  $\pi_B$  in “islands of differentiation” as predicted by all the models discussed

above. However, the observation that peaks in  $F_{ST}$  tended to have low values of  $\pi_B$  provides greater support for a “recurrent-selection” or “sweep-before-differentiation” model of high  $F_{ST}$  regions rather than a “divergence-with-gene-flow” model. This finding adds to the growing body of literature connecting high differentiation with repeated bouts of selection (Campo et al. 2013; Veale & Russello, 2017; Irwin et al. 2018; Hejase et al. 2020.) Unfortunately, current theory is unable to definitely distinguish between the “recurrent-selection” and “sweep-before-differentiation” models. However, an approach that compares the coalescence times of  $F_{ST}$  peaks driven by recurrent selection to the coalescence times of regions across the rest of the genome may help make this distinction. Here, the “recurrent-selection” model predicts that  $F_{ST}$  peaks will coalesce within the same relative time period as the rest of the genome while the “sweep-before-differentiation” model predicts that  $F_{ST}$  peaks will have much more recent coalescence times.

Support for one type of model should not be taken as opposition for the other. It is likely that divergent selection and repeated bouts of selection are both important in the yellowhammer and pine bunting system and work in combination to create and maintain “islands of differentiation”. For example, Irwin et al. (2018) propose a scenario that combines the “sweep-before-differentiation” and “divergence-with-gene-flow” model where areas of past adaptive selective sweeps—global and local—between populations become important to reproductive isolation. As such, when taxa hybridize, either because gene flow between populations never ceased completely or because populations meet in an area of secondary contact, divergent selection would reinforce and potentially expand areas of high  $F_{ST}$ . Such circumstances would create the same genomic patterns observed in the yellowhammer and pine bunting system and warrant additional investigation given the extensive and dynamic nature of their hybrid zone.

Of those identified between allopatric yellowhammers and pine buntings, I saw the widest and most highly differentiated “island of differentiation” on the Z chromosome. Greater peaks in differentiation on sex chromosomes compared to autosomes is a common observation when comparing closely related species (Borge et al. 2005; Ruegg et al. 2014; Sackton et al. 2014) and is consistent with higher fixation rates on the Z chromosome due to the “faster Z/X effect” (reviewed in Meisel & Connallon, 2013; reviewed in Irwin, 2018). However, the significant width of the high  $F_{ST}$  region on this chromosome as well as large the number of SNPs that have  $F_{ST}$  values close to one suggests there could be other factors involved in this pattern.

One possible explanation for this wide, highly differentiated region could be that it corresponds to an area of low recombination leading to linked selection of nearby loci. In this situation, low recombination would create blocks of connected loci that are continuously inherited together. Strong divergent selection that increases the frequency of different versions of one SNP would similarly increase the frequency of all the loci that are linked to those versions such that different linked blocks become fixed in separate populations (reviewed in Cutter & Payseur, 2013). This would create a wide area of differentiation between taxa rather than a peak in differentiation at a single locus.

Areas of low recombination can be the product of many structural aspects of the nuclear genome including intrachromosomal inversion polymorphisms (reviewed in Smukowski & Noor, 2011). Different polymorphisms of an inversion experience little to no successful recombination often creating large blocks of linked loci (reviewed in Kirkpatrick, 2010). As mentioned above, divergent selection acting on any part of an inversion would select for the entire linked block and drag different polymorphisms of the inversion to fixation thus creating a wide area of

differentiation between taxa. The possibility of an inversion existing on chromosome Z in the yellowhammer and pine bunting system is investigated more fully in Chapter 3. Briefly, I did find some evidence of an intrachromosomal inversion polymorphism segregating within this system as well as some association between this inversion and plumage variation. Therefore, it is possible that divergent sexual selection on different plumage phenotypes associated with different inversion polymorphism is responsible for the wide “island of differentiation” seen between yellowhammers and pine buntings on the Z chromosome.

Contrary to the islands of differentiation identified between allopatric yellowhammers and pine buntings, the mtDNA introgression proposed between these groups is expected to homogenize their nuclear genomes by driving co-introgression of mitonuclear genes (Beck et al. 2015; Sloan et al. 2017; Morales et al. 2018). Natural selection would select for such co-introgression to ensure mitonuclear compatibility and proper mitochondrial functioning (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019). Consistent with this theory, my two-tailed binomial tests comparing introgression windows with protein-coding gene locations provided some evidence of preferential introgression at mitonuclear genes in allopatric yellowhammers. No comparable signal of preferential mitonuclear gene introgression was found in pine buntings suggesting that introgression occurred in the direction of pine buntings into yellowhammers. Currently, it is not known whether hypothesized mtDNA introgression occurred in a similar direction, but this information could provide further support for mitonuclear co-introgression in this system.

A significant portion of the mitonuclear genes identified as having introgressed in allopatric yellowhammer and pine bunting populations code for structural components of the

mitochondrial electron transport chain (ETC). The ETC is broken into five protein complexes which, through a series of enzymatic reactions within the mitochondria, perform oxidative phosphorylation to produce ATP necessary for organism survival (reviewed in Ernster & Schatz, 1981). Four of the five ETC complexes are made up of structural subunits encoded by both the nuclear and mitochondrial genome (reviewed in Hill, 2019). Correct fit between subunits is essential to the enzymatic function of each complex and to the flow of electrons and protons across the ETC. To put this in perspective, changing a single amino acid in one subunit can significantly disrupt its ability to even interact with other subunits within a complex (Gershoni et al. 2014). Considering the tight, heavily regulated interactions within the ETC complexes it is unsurprising that mitonuclear genes coding for ETC structural subunits would preferably introgress during mitonuclear co-introgression. If something as small as a single amino acid change has the potential to hinder ETC complex function, then the introduction of several potentially divergent and incompatible mtDNA-encoded structural subunits into a species' ETC via mtDNA introgression would likely have dire consequences on fitness. Such a situation would strongly select for the introgression of compatible nuclear encoded subunits that can restore function of the electron transport chain.

Another group of mitonuclear genes that appeared to preferentially introgress within the yellowhammer and pine bunting system were those encoding subunits of the mitochondrial ribosome or mitoribosome. Unlike the protein-protein interactions occurring within ETC complexes, mitonuclear interactions in the mitoribosome are between nucDNA encoded protein subunits and mtDNA encoded RNA (reviewed in Hill et al. 2019). Protein subunits associate closely with rRNA as both these components are necessary for the formation of a functioning

ribosome (Greber & Ban, 2016). However, nucDNA encoded subunits also interact with mtDNA encoded mRNA and tRNA during synthesis of mitochondrial proteins. Currently, research is limited on the extent and importance of interactions between mitoribosomal subunits and mitochondrial RNA. However, the fact that interactions between these factors are so extensive and are necessary for the synthesis of the mitochondrial proteins needed for structures such as ETC complexes, suggests close co-evolution between mtDNA and nuclear genes encoding mitoribosomal subunits that could select strongly for co-introgression.

The field of mitonuclear ecology is in its infancy and researchers are still learning what models and protocols best examine interactions between the mitochondrial and nuclear genomes. Mitonuclear co-introgression is particularly difficult to study as it removes genetic differences between taxa rather than creates easily identifiable differentiation peaks. Keeping these challenges in mind, I recognize the limitations of my study when identifying mitonuclear gene introgression. Reduced representation sequencing only captures a small portion of the nuclear genome scattered across a large number of chromosomes. As such, my genetic analyses were unable to detect significant signals of introgression that appeared over more narrow genomic regions making my identification of introgression windows somewhat coarse and inexact. This coarseness impedes my ability to detect mitonuclear co-introgression and may have contributed to variation in the statistical significance of the test for yellowhammer gene introgression depending on whether gene densities were taken into account. Further work examining mitonuclear co-introgression in this and other systems would benefit from a whole genome sequencing procedure that allows researchers to isolate introgression signals to narrow regions of the genome and directly investigate mitonuclear gene locations.

The yellowhammer and the pine bunting are sister species that are divergent in appearance, song and ecology (Panov et al. 2003; Rubtsov & Tarasov, 2017), but indistinguishable in mtDNA likely as a result of mtDNA introgression (Irwin et al. 2009). Such introgression has the potential to drive co-introgression of mitonuclear genes and homogenize the nuclear genomes of separate taxa. Yet, in allopatry, I found that yellowhammers and pine buntings were easily separated into distinct genetic clusters based on nuclear variation contained within “islands of differentiation”. The negative correlation between  $F_{ST}$  and  $\pi_B$  seen in my dataset suggests that these islands are the result of repeated selective events in ancestral and daughter populations as described in the “recurrent-selection” (Cruickshank & Hahn, 2014) or “sweep-before-differentiation” (Irwin et al. 2018) models. However, as predicted, I also found some evidence of mitonuclear gene introgression in the direction of pine buntings into yellowhammers that is consistent with mitonuclear co-introgression. This occurred preferentially in mitonuclear genes that code for structural components of both the ETC and the mitoribosome likely due to the potentially lethal consequences of incompatibilities between these genes and products of the mitochondrial genome. Mitonuclear incompatibilities are thought to represent an important post-zygotic reproductive barrier between taxa (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019) meaning mitonuclear co-introgression has the potential to weaken species boundaries. Support for such breakdown can be seen in the extensive and dynamic hybrid zone between yellowhammers and pine buntings (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Additional research is needed to investigate this possibility especially within the context of the “islands of differentiation” seen between allopatric

populations and to clarify the role of mitonuclear co-introgression in the process of speciation more generally.

**Table 2.1.** Geographical location, sampling size and sampling break-down for each of the sites included in this thesis. Sampling locations may include multiple sites that appeared too close together to be shown in detail in Figure 2.1. Full details for the sites included in each sampling locations can be found in Supplementary Table 2.1. The sampling location numbers that appear in the “Sampling Location” column correspond to those that appear in red in Figure 2.1. The “Sample Size” columns describes the total number of samples collected from a particular site. Columns “Allopatric E. cit” – “Allopatric E. leuc” describe the demographic breakdown of samples within each sampling location. “E. cit.” represents *Emberiza citrinella* or yellowhammers and “E. leuc.” represents *Emberiza leucocephalos* or pine buntings.

Sampling Location	Latitude (°N)	Longitude (°E)	Sample Size	Allopatric E. cit.	Near Sympatric E. cit.	Sympatric E. cit.	Hybrid	Sympatric E. leuc.	Near Sympatric E. leuc.	Allopatric E. leuc
1	57.99	12.49	1	1	0	0	0	0	0	0
2	59.81	17.05	1	1	0	0	0	0	0	0
3	51.71	18.61	1	1	0	0	0	0	0	0
4	55.28	20.97	5	5	0	0	0	0	0	0
5	65.86	21.48	2	2	0	0	0	0	0	0
6	51.38	35.84	4	3	0	0	1	0	0	0
7	55.97	38.50	18	18	0	0	0	0	0	0
8	61.45	38.67	13	12	0	0	1	0	0	0
9	43.54	40.47	1	1	0	0	0	0	0	0
10	65.85	44.24	1	1	0	0	0	0	0	0
11	58.33	44.76	1	1	0	0	0	0	0	0
12	51.20	57.27	7	7	0	0	0	0	0	0

Sampling Location	Latitude (°N)	Longitude (°E)	Sample Size	Allopatric E. cit.	Near Sympatric E. cit.	Sympatric E. cit.	Hybrid	Sympatric E. leuc.	Near Sympatric E. leuc.	Allopatric E. leuc
13	56.39	58.62	5	0	5	0	0	0	0	0
14	55.03	65.19	19	0	10	0	8	0	1	0
15	53.39	78.65	5	0	0	0	2	3	0	0
16	53.42	83.89	25	0	0	15	6	4	0	0
17	51.04	86.68	120	0	0	35	52	33	0	0
18	54.93	86.82	4	0	0	3	1	0	0	0
19	50.02	89.23	2	0	0	0	0	2	0	0
20	50.42	90.93	3	0	0	0	0	0	3	0
21	55.33	93.65	1	0	0	0	1	0	0	0
22	50.18	95.09	5	0	0	0	0	0	5	0
23	57.28	97.18	1	0	0	1	0	0	0	0
24	52.6	104.48	19	0	0	8	1	10	0	0
25	56.41	105.46	5	0	0	5	0	0	0	0
26	52.33	106.81	10	0	0	0	1	0	9	0
27	49.64	110.17	2	0	0	0	0	0	0	2
28	50.66	115.09	17	0	0	0	0	0	0	17
29	51.12	118.56	15	0	0	0	0	0	0	15

Sampling Location	Latitude (°N)	Longitude (°E)	Sample Size	Allopatric E. cit.	Near Sympatric E. cit.	Sympatric E. cit.	Hybrid	Sympatric E. leuc.	Near Sympatric E. leuc.	Allopatric E. leuc
30	50.56	143.08	8	0	0	0	0	0	0	8
		<b>Total</b>	<b>321</b>	<b>53</b>	<b>15</b>	<b>67</b>	<b>74</b>	<b>52</b>	<b>18</b>	<b>42</b>

**Table 2.2.** Summary statistics calculated while conducting mitonuclear co-introgression analysis. A total of 7187 2000bp windows were considered when determining introgression windows. A total of 134 mitonuclear genes were investigated for signals of co-introgression. “\*\*\*” indicates a significant p-value.

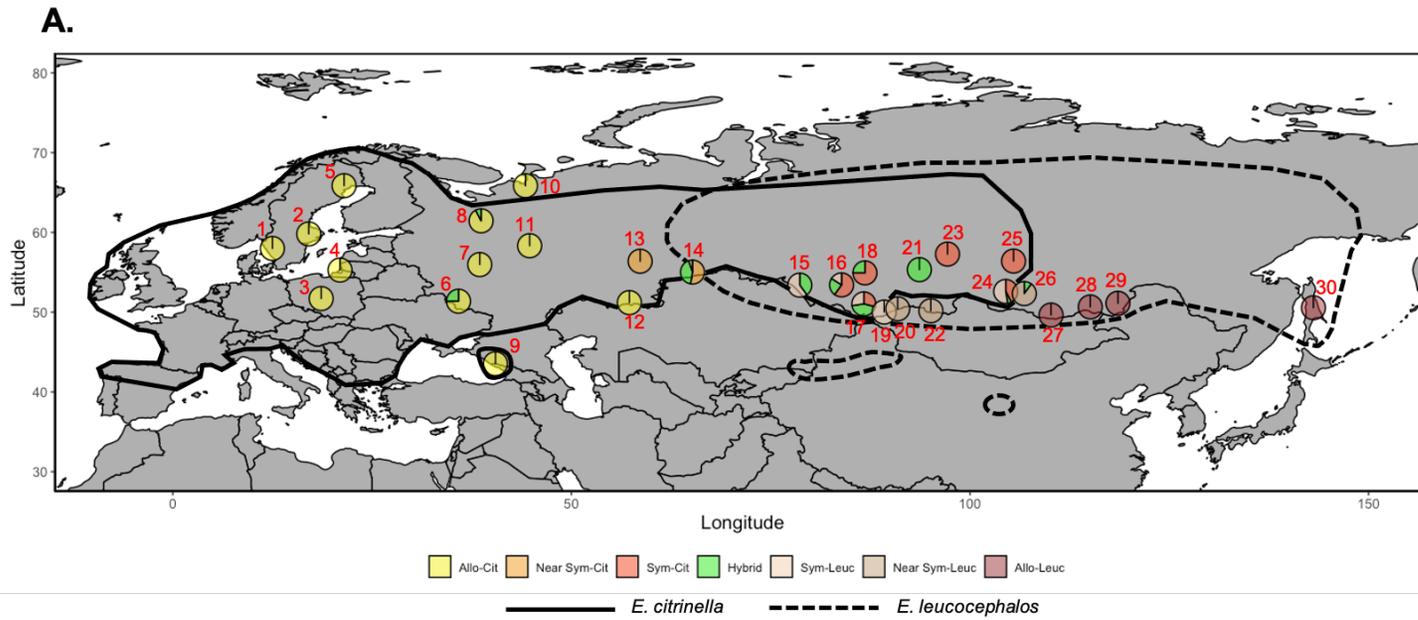
Species	# of introgression windows identified	% of mitonuclear genes appearing in introgression windows	Binomial test p-value	Fisher’s Exact test p-value
Yellowhammer	244	6.7	0.04952**	0.1311
Pine bunting	222	3.0	1	1

**Table 2.3.** Identities, chromosomal locations, window rankings and functions of mitonuclear genes that appeared within 244 yellowhammer introgression windows. Window rankings span values of 1 to 244 and are based on the average Tajima’s D value of each introgression window. The introgression window with the lowest average Tajima’s D value was given the ranking of 1 and the introgression window with the highest Tajima’s D value was given a ranking of 244. In the “Mitonuclear Gene Function” column, ETC stands for “Electron Transport Chain”. Mitonuclear gene names are written as they appear in Hill (2019).

Mitonuclear Gene	Chromosome where mitonuclear gene is found	Introgression Window Ranking	Mitonuclear Gene Function
APOPT1	5	215	Assembly factor/ancillary protein for ETC complex IV
COX5A	10	81	Structural subunit of ETC complex IV
COX17	1	48	Assembly factor/ancillary protein for ETC complex IV
MRPL1	4	214	Mitochondrial large ribosomal subunit protein
MRPL27	18	209	Mitochondrial large ribosomal subunit protein
MRPL32	2	142	Mitochondrial large ribosomal subunit protein
NDUFC1	4	86	Structural subunit of ETC complex I
mtSSB	1A	108	Single stranded DNA-binding protein
UQCR11	28	51	Structural subunit of ETC complex III

**Table 2.4.** Identities, chromosomal locations, window rankings and functions of mitonuclear genes that appeared within 222 pine bunting introgression windows. Window rankings span values of 1 to 222 and are based on the average Tajima’s D value of each introgression window. The introgression window with the lowest average Tajima’s D value was given the ranking of 1 and the introgression window with the highest Tajima’s D value was given a ranking of 222. In the “Mitonuclear Gene Function” column, ETC stands for “Electron Transport Chain”. Mitonuclear gene names are written as they appear in Hill (2019).

Mitonuclear Gene	Chromosome where mitonuclear gene is found	Introgression Window Ranking	Mitonuclear Gene Function
ATP5H5I	18	15	Structural subunit of ETC complex V
COX5A	10	25	Structural subunit of ETC complex IV
MRPL2	3	142	Mitochondrial large ribosomal subunit protein
NDUFB4	1	107	Structural subunit of ETC complex I



**B.**



Pure Citrinella (PC)



Almost Citrinella (SC)



Citrinella Hybrid (CH)



Yellow Hybrid (YH)



Pure Leucocephalos (PL)



Almost Leucocephalos (SL)

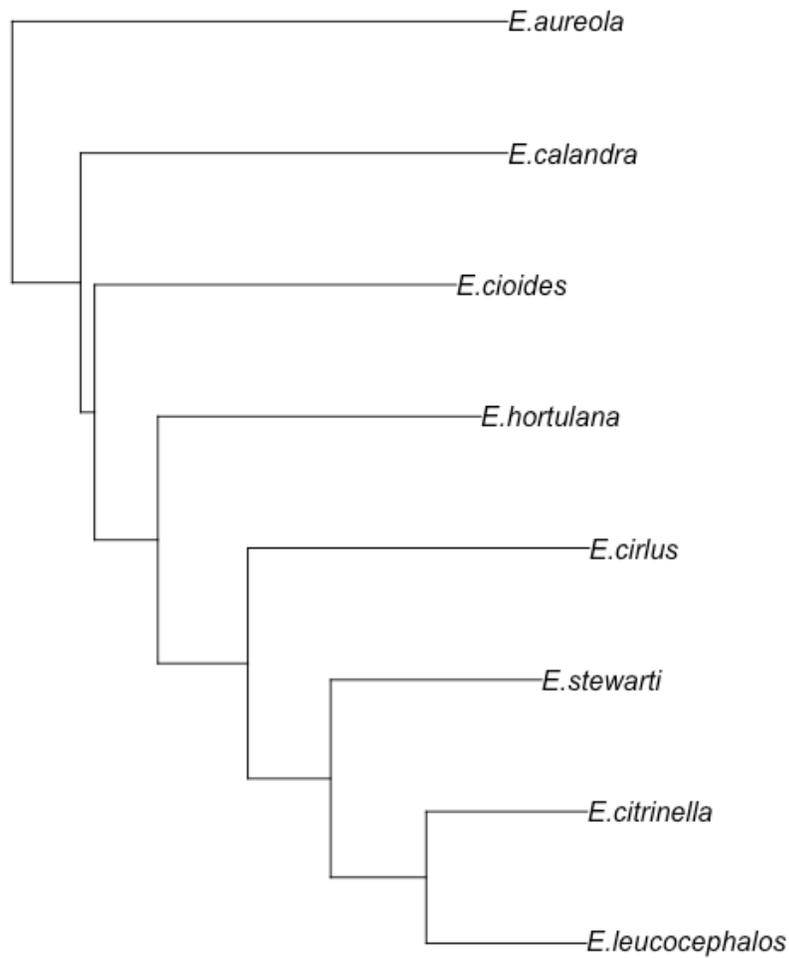


Leucocephalos Hybrid (LH)

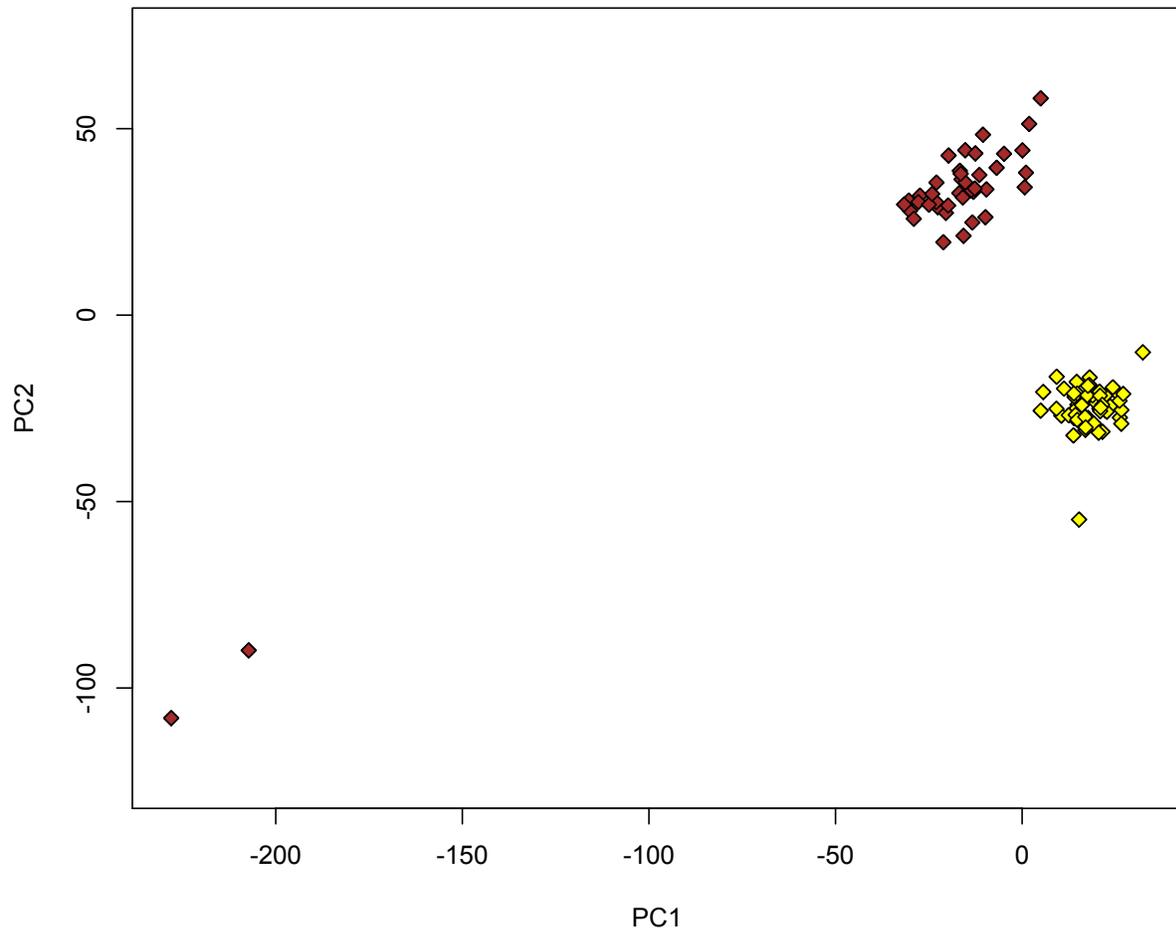


White Hybrid (WH)

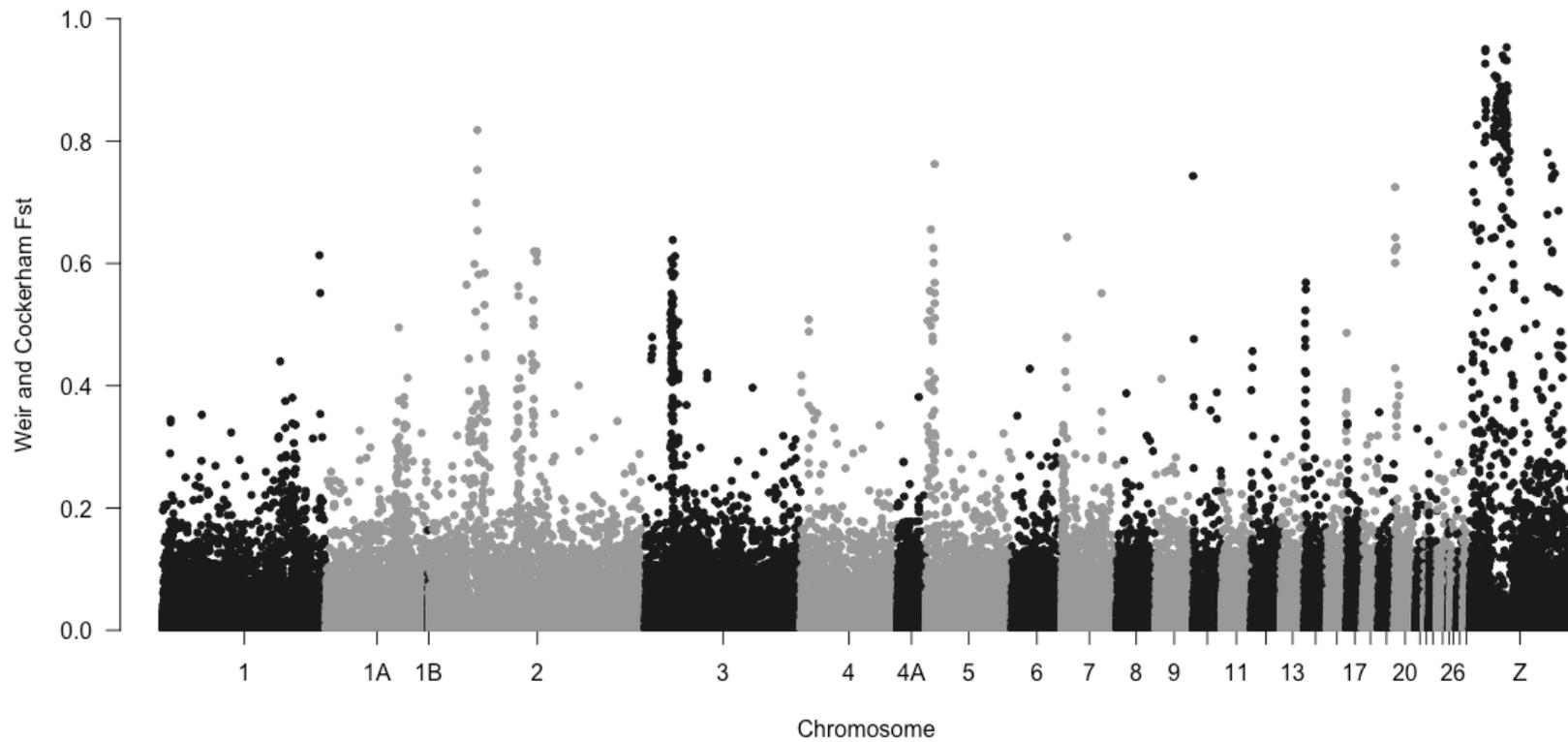
**Figure 2.1. A)** Map showing all the sampling locations included in this research. Red numbers accompanying each sampling location pie correspond to the sampling location numbers that appear in Table 1, which shows sample size and composition. Sampling locations may include multiple sites that appeared too close together to be shown in detail on the map. Full details for the sites included in each sampling locations can be found in Supplementary Table 2.1. Each sampling location pie is coloured and divided based on the proportion of each sample type that appeared within it. The sample types include: allopatric yellowhammers (Allo-Cit; yellow), near-sympatric yellowhammers (Near Sym-Cit; light orange), sympatric yellowhammers (Sym-Cit; red-orange), hybrids (Hybrid; green), sympatric pine buntings (Sym-Leuc; peach), near sympatric leucocephalos (Near Sym-Leuc; taupe) and allopatric pine buntings (Allo-Leuc; brown). The solid black line indicates the geographic range of the yellowhammer and the dashed black line indicates the geographic range of the pine bunting as described in Irwin et al. (2009). **B)** Photos depicting phenotypic variation within the yellowhammer and pine bunting system. Each photo represents one of eight phenotypic classes that individuals are divided into based on variation at three plumage traits: background colour, amount of chestnut at the brow and amount of chestnut at the throat. The photos show one example of each classes, but are unable to capture the full variation within each phenotypic group. All photos are credited to Dr. Alexander Rubtsov.



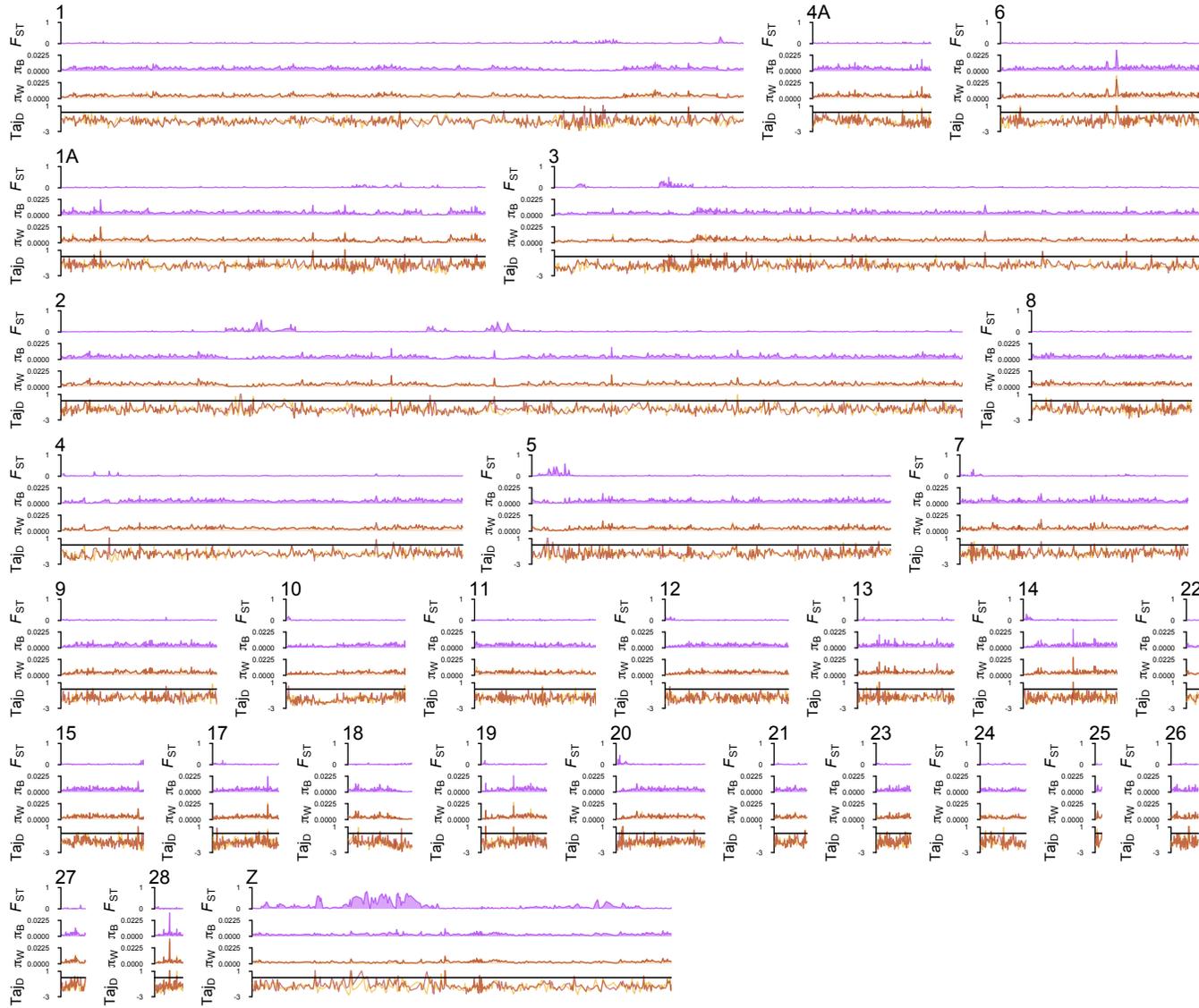
**Figure 2.2.** Neighbour-joining tree of *Emberizidae* species created using average absolute between-population nucleotide diversity ( $\pi_B$ ). Sample sizes for each species are as follows: *E. aureola* = 1, *E. calandra* = 1, *E. cioides* = 1, *E. hortulana* = 1, *E. cirrus* = 6, *E. stewarti* = 4, *E. citrinella* = 53 and *E. leucocephalos* = 42.



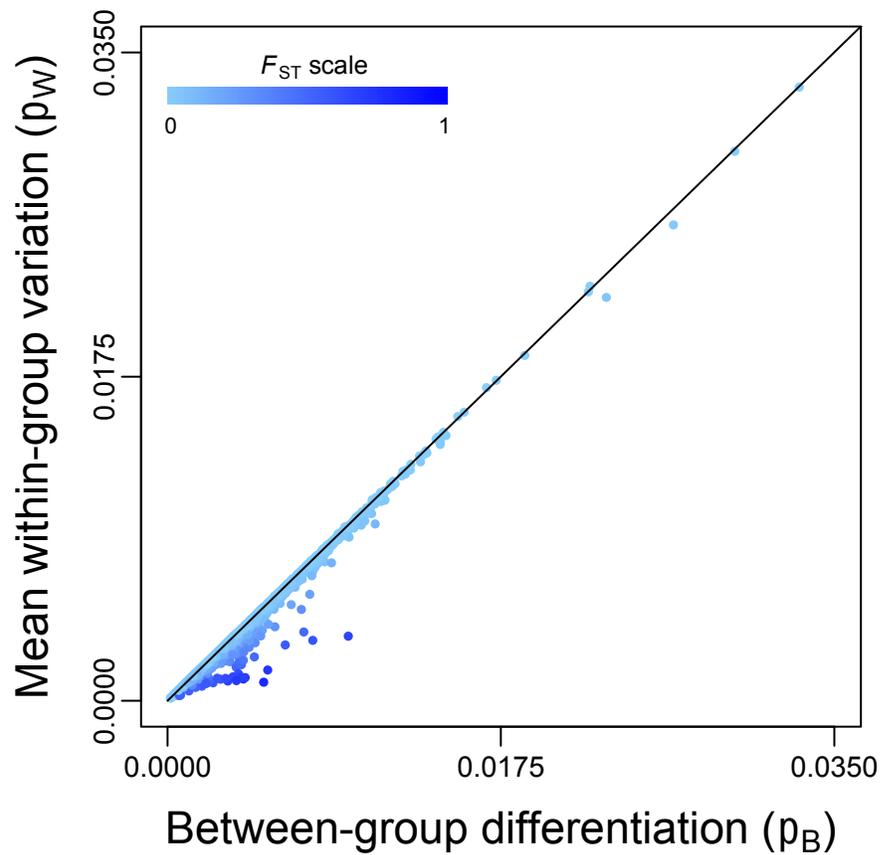
**Figure 2.3.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 53$ ) and allopatric pine buntings (brown;  $n = 42$ ). PC1 explains 3.6% of the variation among individuals and PC2 explains 2.9% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.



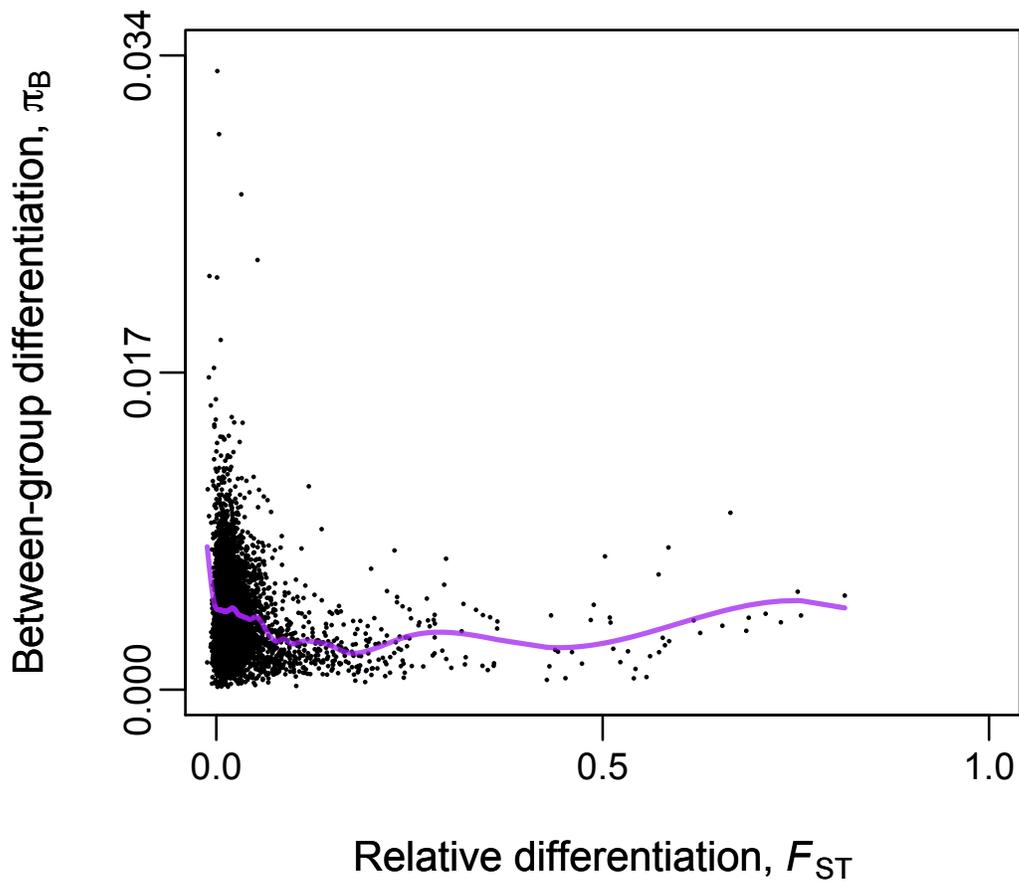
**Figure 2.4.** Relative differentiation ( $F_{ST}$ ) of 349,807 genome-wide SNPs identified between allopatric yellowhammers ( $n = 53$ ) and allopatric pine buntings ( $n = 42$ ).



**Figure 2.5.** Genome-wide patterns of genetic variation comparing allopatric yellowhammers ( $n = 53$ ) and allopatric pine buntings ( $n = 42$ ). Relative nucleotide differentiation ( $F_{ST}$ ), absolute between-population nucleotide diversity ( $\pi_B$ ), absolute within-population diversity ( $\pi_W$ ) and Tajima's D ( $Taj_D$ ) are shown as 2000 bp windowed averages across each chromosome.  $F_{ST}$  and  $\pi_B$  are shown as purple lines to indicate that values were calculated as a comparison between allopatric yellowhammers and pine buntings.  $\pi_W$  and  $Taj_D$  are shown as two separate lines (yellow = yellowhammers, brown = pine buntings) to indicate that values were calculated separately for each population.



**Figure 2.6.** The mean within-group absolute variation ( $\pi_W$ ) of allopatric yellowhammers ( $n = 53$ ) and allopatric pine buntings ( $n = 42$ ) plotted against between-group absolute differentiation ( $\pi_B$ ). Each dot represents the average value taken from a 2000 bp window of sequenced data across the nuclear genome. The black line indicates where mean within-group differentiation equals between-group differentiation. Increasing average values of  $F_{ST}$  calculated for each window are shown in darker shades of blue.



**Figure 2.7.** Association between relative differentiation ( $F_{ST}$ ) and between-group differentiation ( $\pi_B$ ) of allopatric yellowhammers ( $n=53$ ) and allopatric pine buntings ( $n = 42$ ). Each black dot represents average values calculated from a 2000bp window of sequenced data. A cubic spline fit between the variables is shown as purple line.

## **Chapter 3: A chromosomal inversion maintains divergent plumage phenotypes as two avian species merge into one**

### **3.1 Introduction**

Speciation—a key way by which new biodiversity is created—is a complex and dynamic process subject to the influence of many deterministic and stochastic forces (Coyne & Orr, 2004; Price, 2008). Due to its multifaceted nature and long evolutionary timescale, speciation has proved challenging to study particularly when focus is placed on taxa that have already achieved complete reproductive isolation and full species status. Although such comparisons allow researchers to identify the core differences that characterize groups and to postulate about the key traits that drove them apart, work in these systems ultimately focuses on the endpoint of speciation rather than the complete process. To understand the full breadth of speciation, it is necessary to investigate not only its products, but also its intermediate stages. Hybrid zones—often described as natural laboratories of evolution (Hewitt, 1988)—are representative of such stages. In these areas where taxa have yet to attain complete reproductive isolation, researchers can directly assess how divergence in allopatry influences interbreeding between sympatric groups (Barton & Hewitt, 1985; 1989; Hewitt, 1988; Gompert et al. 2017). In this way, researchers are able to identify the important reproductive barriers that limit gene flow during the initial stages of divergence and to investigate how the accumulation of such barriers may lead to cessation of interbreeding and the completion of the speciation process.

A reproductive barrier is described as any trait that restricts two populations' ability to interbreed and/or produce fit offspring thus restricting gene flow between groups (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). Though often cumulative in their effects, these barriers are highly variable in nature and can be represented by a diverse suite of morphological, behavioural, physiological and genetic characteristics. In hopes of organizing such a diversity of traits into comprehensive categories, evolutionary scientists have devised many different ways of classifying reproductive barriers. However, it is often most intuitive to think of them as acting either before the formation of a hybrid zygote (prezygotic) or following the formation of a hybrid zygote (postzygotic).

Prezygotic reproductive barriers function by preventing an interbreeding event between members of divergent taxa or by preventing fertilization of an egg cell by an individual from a divergent taxon (Mayr, 1942). Examples of prezygotic barriers can include divergence in breeding periods such that different populations display breeding behaviour at asynchronous times (e.g. Moore et al. 2005; Danley et al. 2007) or divergence in the morphology of reproductive structures such that copulation between members of separate taxa is not possible (e.g. Sota & Kubota, 1988; Wogjciech & Simmons, 2013). Yet, perhaps the most well-known and well-studied class of prezygotic barriers are those related to mate attraction and intersexual selection. Mate attraction often involves one sex—usually the males—showcasing a particular trait such as a colouration pattern, a vocal call or a behavioural display to the opposite sex in order to secure a breeding partner and produce offspring (West-Eberhard, 1983; Grant & Grant, 1997; Edwards et al. 2005). Evolution and elaboration of these display traits is driven by intersexual selection by the “choosing” sex who use these signals to differentiate between

conspecific and heterospecific individuals and to discern the most fit mate with which to breed. Divergence of intersexually selected traits and associated trait preference can generate strong prezygotic barriers by increasing discrimination against heterospecific individuals during mate choice thus preventing interbreeding.

Of the intersexually selected traits that have been identified as potential prezygotic barriers, specific focus has been placed on colouration differences between taxa. Research in various systems has provided evidence that divergent colour patterning acts as or reinforces reproductive barriers between differentiated, sympatric populations (e.g. Saetre et al. 1997; Lukhtanov et al. 2005; Seehausen et al. 2008). For example, Uy et al. (2009) investigated species recognition between different populations of chestnut flycatcher using male territorial response as a proxy for female choice. Their results showing strong discrimination based on plumage patterns suggest that plumage traits are significant mediators of interbreeding and gene flow between diverging populations (Uy et al. 2009).

Recently, research on colouration and its role in reproductive isolation has focused on determining the genetic underpinnings of these traits in order to better understand how specific genomic regions may diverge during speciation. This work has been particularly successful in identifying genes that regulate melanin-based traits such as melanocortin-1 receptor (MC1R) (Theron et al. 2001; Rosenblum et al. 2004; Gross et al. 2009) and agouti-signaling protein (ASIP) (Kingsley et al. 2009; Cerdà-Reverter et al. 2005; Haupaix et al. 2018). Carotenoid-based traits have proved more challenging to study as this class of pigments is obtained from the environment and deposited on the integument rather than produced endogenously like melanin (Hubbard et al. 2010; Mason & Bowie, 2020). Nevertheless, certain genes have been identified

that could be important for carotenoid deposition including beta-carotene oxygenase 2 (BCO2) (Toews et al. 2016b; Andrade et al. 2019) and the genes coding for “scavenger receptors” (Brelsford et al. 2017; Toomey et al. 2017). Further work is needed to continue investigating the genetic pathways controlling colouration traits and to more concretely connect differentiation in these pathways to prezygotic barriers between taxa.

Unlike prezygotic reproductive barriers, postzygotic barriers hinder gene flow between diverging populations by acting not on the interbreeding parents, but on the hybrid zygote or offspring (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). These barriers manifest as inviability, sterility or lower fitness of hybrids such that they are unable to produce offspring or produce fewer offspring than individuals with pure parentage. Interestingly, many of the traits capable of acting as prezygotic barriers within a system can also act as postzygotic barriers if hybrid phenotypes are different or intermediary between pure phenotypes (Bridle et al. 2006; Irwin, 2020). For example, in the case of colouration traits, a hybrid might possess colouration patterns that are intermediate between its parents. If this intermediate hybrid phenotype is not recognized by pure or, even, other hybrid individuals such that the individual is not able to attract a mate and produce offspring, colouration could act a postzygotic barrier.

Often, the cause of postzygotic barriers are genetic incompatibilities that evolve between differentiated groups. In this situation, genetic differentiation creates a unique suite of co-evolved alleles within each population (Bateson, 1909; Dobzhansky, 1937; Muller, 1942). Interbreeding between these populations mixes together co-evolved parental alleles creating hybrids with unique genetic combinations that have not occurred before in nature. Because these new collections of alleles did not evolve within a similar genetic background, it is likely that the

alleles of one population will function poorly or be incompatible with those from a divergent population causing negative epistatic interactions that impact hybrid fitness. Direct classification of the loci responsible for genetic incompatibilities has proved difficult to accomplish in non-model species, but research has identified important patterns related to their evolution and their association with reproductive isolation. One commonly noted trend concerning this relationship is that hybrid deficiency is often linked to strong incompatibilities on the sex chromosomes (Sætre et al. 2003; Presgraves, 2008).

The importance of sex chromosomes during the evolution of postzygotic barriers has been recognized for decades and is observed in its most extreme sense in Haldane's rule. Haldane's rule states that, in a situation where hybridization between divergent taxa produces sterile or inviable offspring, the heterogametic sex is more likely to incur these fitness losses than the homogametic sex (Haldane, 1922; reviewed in Coyne & Orr, 2004). This pattern applies to both XY sex determination systems where males are the heterogametic sex and ZW sex determination systems where the females are the heterogametic sex. From Haldane's rule, it is assumed that many of the genetic incompatibilities responsible for hybrid inviability and sterility reside on sex chromosomes, a pattern actually demonstrated in *Drosophila* (Presgraves, 2008). Explanations for this phenomenon are varied, but the hypothesis that is applicable to the widest number of systems and types of hybrid fitness loss concerns dominance interactions on sex chromosomes. Following population divergence, genetic incompatibilities that are caused by recessive alleles on the X or Z chromosome are expected to appear immediately within the heterogametic sex because these individuals do not possess another copy of the chromosome with compatible alleles that will mask negative interactions (Muller, 1940; 1942; reviewed in

Coyne & Orr, 2004). As a result, genetic incompatibilities are expected to arise at a higher rate during population divergence and to contribute more strongly to postzygotic barriers when they are caused by loci on sex chromosomes rather than by loci on autosomes. These dominance effects are supported directly by extensive lab experiments (reviewed in Coyne & Orr, 2004) and indirectly by an increasing number of studies reporting that sex chromosomes possess consistently high genetic differentiation between taxa (Borge et al. 2005; Harr, 2006; Sackton et al. 2014) and contribute strongly to speciation (Sæther et al. 2007; Presgraves, 2008).

Understanding the nature of reproductive barriers and how they mediate gene flow between taxa is a crucial part of speciation research. With the advent of new sequencing technology, it is becoming increasingly possible to identify the loci responsible for barriers and to follow these regions over the course of population divergence. Yet, in addition to isolating their genetic origins, researchers have also described important genomic characteristics that affect the evolution and maintenance of reproductive barriers. Of particular interest are chromosomal inversion polymorphisms that produce areas of low recombination between differentiated populations.

If separate populations become fixed for different forms of a chromosomal inversion, the inversion could act as a postzygotic barrier in its own right by impairing meiosis within a hybrid offspring (reviewed in Rieseberg, 2001; reviewed in Wellenreuther & Bernatchez, 2018). However, fixation of inversions that cause such hybrid deficiencies requires a very stringent set of circumstances (Walsh, 1982; Lande, 1985) and is seen rarely in animal species (Coyne et al. 1993; Bardham and Sharma. 2000). Instead, chromosomal inversions are thought to impact the evolution of reproductive isolation by suppressing recombination within hybrids (Noor et al.

2001; Rieseberg, 2001). Recombination suppression creates genomic blocks that are resistant to homogenization during interbreeding thus allowing genetic differences important to reproductive barriers to accumulate between populations even when they are connected by gene flow. In this capacity, chromosomal inversions have the potential to impact the evolution of both prezygotic and postzygotic barriers (reviewed in Hoffman & Rieseberg, 2008). For example, in a colouration-based prezygotic barrier, an inversion can maintain an association between a locus that controls male colour patterning and another locus that controls female preference for a particular colour pattern. In this way, two populations can diverge in colouration traits while concurrently diverging in their preference for these traits such that discrimination based on visual cues evolves efficiently between populations and limits interbreeding (Trickett & Butlin, 1994). In the case of a postzygotic barrier, recombination suppression within inversions allows for the maintenance of minor incompatibilities between populations that would normally be removed by natural selection during hybridization. Over time, incompatibilities continue to accumulate within inversions contributing to substantial hybrid deficiencies and reproductive isolation between taxa (Navarro & Barton, 2003). Additional research is very much needed to continue investigating the relationship between reproductive barriers and chromosomal inversions and to assess the impact of inversions during divergence and speciation more generally.

Hybrid zones between taxa that display strong patterns of islands of genomic differentiation on the Z chromosome may be particularly suitable to studying the role of inversions and sex chromosomes in speciation. These characteristics are seen (chapter 2) when comparing the yellowhammer (*Emberiza citrinella*) and the pine bunting (*Emberiza leucocephalos*), a sister pair of Palearctic songbirds that are thought to have diverged in allopatry

during the Pleistocene glaciations (Irwin et al. 2009). Currently occupying opposite sides of Eurasia, these avian species meet and hybridize extensively within a large and potentially expanding contact zone in central and western Siberia (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Yellowhammers and pine buntings are moderately divergent in both their song and their ecology, but show large differences in their plumage patterns with yellowhammers possessing yellow body plumage and minimal facial markings and pine buntings possessing white body plumage and chestnut facial markings (Panov et al. 2003; Rubtsov & Tarasov, 2017). Hybrids display highly variable, intermediate plumage patterns and can be separated into several different phenotypic classes. “Citrinella hybrids” and “leucocephalos hybrids” show moderate variation in all divergent plumage traits but can be likened to pure phenotypes whereas “white hybrids” and “yellow hybrids” possess discordant phenotypes with the body colour of one species and the facial markings of the other. Despite the aforementioned differences, previous research in the yellowhammer and pine bunting system has found that these species differ negligibly in their mitochondrial DNA and show genomic patterns consistent with adaptive introgression of mitochondrial DNA from one species into the other (Irwin et al. 2009). A genomic comparison of allopatric populations demonstrated that genetic differentiation across the nuclear genome is also limited, but that the existing differentiation separates heterospecific individuals into distinct genetic clusters (Chapter 2). Much of this genetic differentiation is contained within a large “island of differentiation” located on the Z chromosome and several smaller islands interspersed across the genome (Chapter 2). The conflicting pictures created by the substantial plumage differences between species and genetic distinctness in allopatry and by the strong pattern of mitochondrial introgression and widespread hybridization in sympatry poses

the question as to how strong reproductive barriers are between yellowhammers and pine buntings. Genomic analysis of pure and hybrid individuals within the contact zone would shed light on this inquiry and clarify the relationship between these taxa who currently possess species designations.

In this study, I employed reduced representation sequencing to examine genomic variation across the yellowhammer and pine bunting system and to assess whether patterns of genetic divergence in allopatry are maintained within the contact zone and manifest as reproductive barriers between taxa. First, I characterized general patterns of genetic differentiation and genetic ancestry in individuals that appeared phenotypically as pure or hybrid individuals near and within the sympatric zone and contrasted these patterns to those seen within the allopatric zone. Separation of individuals around the sympatric zone into distinct genetic clusters based on species designation would support the existence of reproductive barriers between taxa that are inhibiting gene flow and preventing species merging. Ancestry patterns showing a strong bias towards early generation hybrids and limited backcrossing could indicate the presence of post-zygotic barriers due to reduced hybrid fitness when compared with pure offspring. Second, I investigated whether the large, highly differentiated region on chromosome Z that was identified between allopatric populations (Chapter 2) was retained within the sympatric zone or homogenized by gene flow. Maintenance of this large “island of differentiation” could suggest that it houses loci important to reproductive barriers between taxa or that it is a region of particularly low recombination such as would be expected within a chromosomal inversion. Finally, I examined the genetic underpinnings of plumage differences between yellowhammers and pine buntings using a genome-wide-association-study and related

putative plumage genes to patterns of genetic differentiation between species inside and outside of the contact zone. Plumage traits are the main way by which yellowhammers and pine buntings are classified and may represent an important reproductive barrier between them. Identification of such a barrier would allow for better prediction of the fate of these two taxa who show potential for both continued diversification and species merging.

## **3.2 Methods**

### **3.2.1 Sampling**

The sampling methods for this study are described fully in the Methods section of Chapter 2. Briefly, a total of 335 individuals were included: 135 phenotypic yellowhammers, 112 phenotypic pine buntings, 74 phenotypic hybrids and 14 other members of Emberizidae (one yellow-breasted bunting [*Emberiza aureola*], one corn bunting [*Emberiza calandra*], one meadow bunting [*Emberiza cioides*], one ortolan bunting [*Emberiza hortulana*], four white capped buntings [*Emberiza stewarti*] and six ciril buntings [*Emberiza cirilus*]) acting as outgroups (Figure 2.1A; Table 2.1; Supplementary Table 2.1). Of these samples, 166 are analyzed for the first time in the present study. An additional 34 Emberizidae individuals were sampled, but sequencing did not produce high quality DNA reads. As a result, data from these samples were not included in the analysis.

Where possible, yellowhammer, pine bunting and hybrid males were phenotypically scored in terms of the background plumage colour, the amount of chestnut at the brow (vs. yellow or white) and the amount of chestnut at the throat (vs. yellow or white) based on the

scoring protocol presented in Panov et al. (2003) and Rubtsov and Tarasov (2017). Background colour ranged from bright yellow to pale yellow to pure white in individuals and was assessed for areas of the head and body that did not show brown or black streaking. Males were given a score from 0-7 for each of the three indicated traits with scores of 0 being consistent with the phenotype of a pure yellowhammer and scores of 7 being consistent with the phenotype of a pure pine bunting. Males were also sorted into one of eight classes based on their phenotypic scores—pure citrinella (PC), almost citrinella (SC), citrinella hybrid (CH), yellow hybrid (YH), white hybrid (WH), leucocephalos hybrid (LH), almost leucocephalos (SL) and pure leucocephalos (PL) (Figure 2.1B). Unless otherwise indicated, PC and SC individuals as well as SL and PL individuals were grouped together in analyses as phenotypic yellowhammers and phenotypic pine buntings respectively (Rubtsov and Tarasov, 2017). The remaining classes were categorized as hybrids. Yellowhammer, pine bunting and hybrid individuals were classified by geographic location as being one of “allopatric”, “near-sympatric” or “sympatric”.

### **3.2.2 DNA sequencing and identification of single nucleotide polymorphisms**

A full description of the DNA sequencing and genomic data processing methodology can be found in the Methods section of Chapter 2. Scripts for GBS read filtering were based on those of Irwin et al. (2016; 2018). It should be noted that all statistical analyses in this chapter were conducted on the genome-wide “variant site” VCF file that is comprised of SNP information for all sampled individuals. The average coverage of variable sites among the 335 Emberizidae individuals within my dataset was 16.59 while the average coverage of variable sites among the 321 yellowhammer, pine bunting and hybrid individuals analyzed in this chapter was 16.11.

### **3.2.3 Statistical analysis**

The following statistical analyses were conducted on genomic information obtained from yellowhammers, pine buntings and hybrids in all of the allopatric, near-sympatric and sympatric geographic classes. A focused genomic analysis on allopatric yellowhammers and allopatric pine buntings as well as a comparison of these groups to other members of Emberizidae can be found in Chapter 2. All of the following R (R Core Team, 2014) analyses were completed using version 3.6.2.

#### **3.2.3.1 Variant site analyses**

A total of 882,227 variant sites were identified in my sample of 335 Emberizidae individuals with 374,780 of these sites showing variation among allopatric yellowhammers and pine buntings. I analyzed variable loci using a variety of methods to investigate the underlying population structure of the sympatric zone and its surrounding areas.

First, I used the program Admixture (Alexander et al. 2009) version 1.3.0 to identify the number of ancestral populations in the yellowhammer and pine bunting system and to assign ancestry proportions to each sample. Before running Admixture, a unique “variant site” VCF file was created that included variant sites among all yellowhammers, pine buntings and hybrids samples in the study. This file was trimmed for linkage disequilibrium using Plink version 1.9 (Chang et al. 2015), creating a file of 417,164 SNPs for analysis. Within Admixture, I ran six maximum likelihood models with “K” values ranging from 1 to 6 clusters. A run was terminated when the difference between the log-likelihood values of two consecutive iterations dropped to less than  $1 \times 10^{-10}$ . Out of the six models, cross-validation error was lowest for a “K” value of

1, but was similarly low for a “K” value of 2. Based on the phenotypic, behavioural and ecological differences seen between yellowhammers and pine buntings as well as the clear genetic differentiation identified between allopatric populations on a principal components analysis (Figure 2.3), I plotted the ancestry proportions produced from the “K=2” model in R.

Next, I loaded the untrimmed genome-wide “variant site” VCF file into R that included all Emberizidae individuals. Versions of the scripts presented in Irwin et al. (2018) were used to filter SNPs and to calculate the sample size, allele frequency and Weir and Cockerham’s  $F_{ST}$  (Weir and Cockerham, 1984) for each of the 374,780 SNPs that passed quality thresholds and that were variable between allopatric populations. Allopatric yellowhammers (n=53) and allopatric pine buntings (n=42) were compared for the  $F_{ST}$  calculations. Based on information from these variable SNPs, I conducted a principal components analysis (PCA) including all yellowhammers, pine buntings and hybrids using the `pcaMethods` package (Stacklies et al. 2007) with the “`svdImpute`” command to impute missing genomic data. Furthermore, to investigate how differentiation varied between chromosomes, I also conducted chromosome-specific PCAs by filtering my genome-wide variant sites for each chromosome of interest.

Based on the unexpected population structure identified in the PCA of chromosome Z comparing yellowhammers, pine buntings and hybrids, I extracted a subset of individuals from my dataset to be graphed separately in a “Genotype-by-Individual” plot. The code for this plot is included as part of the analysis by Irwin et al. (2018) though the associated figure did not appear in the publication. In creating this figure, variant sites on chromosome Z were filtered such that SNPs with  $F_{ST}$  values lower than 0.7 were excluded. By subsetting the dataset in this way, I was able to focus on the highly differentiated loci likely responsible for the clustering patterns seen in

the PCA described above. The allele frequencies of these high  $F_{ST}$  SNPs were plotted as a series of boxes for each individual with a fully coloured box indicating homozygosity at a site and a split box indicating heterozygosity at a site. I coloured alleles for each SNP based on which allele was most common in the allopatric yellowhammers (dark purple), with the other allele being coloured light purple.

Finally, to investigate demographic and breeding trends within the yellowhammer and pine bunting system, I compared the ancestry and heterozygosity of individuals in a triangle plot. To begin with, I filtered out variant sites with  $F_{ST}$  values less than 0.6 to create a dataset of 146 SNPs. The threshold of 0.6 created a dataset that included a moderate number of SNPs that together allow easy differentiation between the taxa. To account for linkage between variant sites, I also created an additional dataset of 10 putative unlinked SNPs with  $F_{ST}$  values greater than 0.5. Here, the threshold was lowered to 0.5 to bolster the number of SNPs included within my dataset while ensuring each SNP was able to differentiate between taxa reasonably well. I applied the HlEst command from the HlEst package (Fitzpatrick, 2012) to each of these datasets to obtain maximum likelihood estimates of the ancestry index and interclass heterozygosity of each yellowhammer, pine bunting or hybrid individual. A triangle plot of these values was then produced in R.

### **3.2.3.2 Plumage trait admixture mapping**

I conducted admixture mapping on three plumage traits of interest in the yellowhammer and pine bunting system—the background plumage colour (bright yellow to pure white), the amount of chestnut at the brow (vs. yellow or white) and the amount of chestnut at the throat (vs.

yellow or white). This was completed using the program GEMMA (Zhou and Stephens, 2012) which applies a genome-wide efficient mixed model association algorithm to correlate variation in the genome to variation in phenotypic traits. GEMMA requires that all genomic data be imputed or “complete” before it is run through admixture mapping. To meet this criterion, I used the program “Bimbam” (Scheet and Stephens, 2006) to perform imputation on my genome-wide “variant sites” VCF file. The parameters that I used when imputing were: 10 EM, 40 steps, 7 clusters. It should be noted that, during imputation without a panel, Bimbam only considers genetic linkage disequilibrium information among the cohort of samples provided. It also does not take into account the different phenotypes of each sample when filling in missing genetic data. Using this imputed dataset, I created a centred relatedness matrix in GEMMA to be used in later admixture mapping. I then used GEMMA to run three univariate linear mixed models (ULMM)—one for each phenotypic trait—using a likelihood ratio test to assess association between genomic and phenotypic information at each SNP. Based on GEMMA’s internal filtering parameters, 220,220 SNPs were included in the background plumage ULMM, 220,124 SNPs were included in the brow ULMM and 220,307 SNPs were included in the throat ULMM. It should be noted that only male yellowhammer, pine bunting and hybrid individuals with phenotypic scores at all traits ( $n = 230$ ) were included in GEMMA analyses. Probability values from the likelihood ratio tests performed in each ULMM were visualized on Manhattan plots created using the package “qqman” (Turner, 2018).

### **3.2.3.3 Investigation of plumage trait dominance patterns**

Following admixture mapping, I investigated patterns of dominance for each of the SNPs significantly associated with a plumage trait of interest. In particular, I focused my attention on

the SNPs located within a putative chromosomal inversion located on chromosome Z that was significantly associated with facial plumage variation and weakly associated with body plumage variation within the yellowhammer and pine bunting system. To conduct this analysis, I created contingency tables that compared the phenotypic scores of each trait against the genotypes of significantly associated SNPs for all individuals that possessed this information. These tables were then visualized as balloon plots using the `ggpubr` package (Kassambara, 2017) and examined for patterns that suggest one ancestral phenotype is dominant over the other.

### **3.3 Results**

As described in the Results section of Chapter 2, I identified 882,227 variant sites in my genotyping-by-sequencing (GBS) dataset of 335 Emberizidae individuals. Of these sites, 374,780 were variable among allopatric populations of yellowhammers and pine buntings and were retained for statistical analysis in this chapter.

#### **3.3.1 Population structure inside and outside the hybrid zone**

To better understand the genetic relationships between different geographic populations of yellowhammers, pine buntings and their hybrids, I conducted both a principal components analysis (PCA) and an Admixture analysis of variant SNPs. In Chapter 2, I reported that allopatric populations of yellowhammers and pine buntings appeared as discrete clusters on a PCA (Figure 2.3). The genetic distinctiveness of these populations was supported by ancestry proportions produced using Admixture as all allopatric individuals possessed near-pure yellowhammer or pine bunting ancestry (Figure 3.1B).

For the most part, near-sympatric populations retained the genetic distinctness that was seen in allopatric populations. Near sympatric individuals clustered very closely with their allopatric counterparts in a PCA (Figure 3.1A; Supplementary Figure 3.1) and showed ancestry proportions near 100% (Figure 3.1B). One near-sympatric pine bunting possessed an admixed genotype with about 40% ancestry from yellowhammer populations. This individual appeared as an intermediate between allopatric clusters of yellowhammers and pine buntings in a PCA (Figure 3.1A; Supplementary Figure 3.1).

In the sympatric zone, the genetic distinctiveness seen between yellowhammer and pine bunting populations partially broke down. The majority of sympatric yellowhammers clustered closely with allopatric and near-sympatric yellowhammers in a PCA and possessed near-pure yellowhammer ancestry (Figure 3.1; Supplementary Figure 3.2). However, nine individuals experienced some admixture from pine bunting populations and had between 1-18% pine bunting ancestry (Figure 3.1B). Admixture further estimated that a single phenotypic yellowhammer from the sympatric zone possessed near-pure pine bunting ancestry. This latter bird clustered more closely with pine bunting populations in a PCA than with its putative species (Figure 3.1A). Sympatric pine buntings showed even greater variation in their genotypes. Roughly half of the sympatric pine buntings within my dataset grouped closely with near-sympatric and allopatric pine buntings along PC1 and possessed near-pure pine bunting ancestry (Figure 3.1; Supplementary Figure 3.2). The other half appeared as intermediates between allopatric populations on a PCA and possessed anywhere between 1-78% yellowhammer ancestry (Figure 3.1). Two further sympatric pine buntings clustered closely with sympatric

yellowhammers along PC1 and, contrary to their appearances, possessed more than 90% yellowhammer ancestry.

Similar to sympatric pine buntings, hybrids were also extremely varied in their genotypes. A large portion of my hybrid samples appeared as intermediates between divergent allopatric clusters in a PCA (Figure 3.1A). Nevertheless, many phenotypic hybrids possessed greater yellowhammer ancestry and tended to group more closely with yellowhammer than pine bunting populations (Figure 3.1)—a pattern that has been reported in previous research on this system (Irwin et al. 2009). A total of 33 hybrid individuals were estimated as having near-complete yellowhammer ancestry despite their intermediate plumage patterns while only two showed near-complete pine bunting ancestry (Figure 3.1B). The remaining hybrids possessed highly variable ancestry proportions with yellowhammer ancestry ranging from 3-99%. Very few individuals showed ancestry proportions close to 50-50 as might be expected in an F1 hybrid.

By splitting the hybrids into their respective phenotypic classes—citrinella hybrids, yellow hybrids, white hybrids and leucocephalos hybrids—I could appreciate additional patterns in their ancestry proportions (Supplementary Figure 3.3). Citrinella hybrids possessed consistently high yellowhammer ancestry—91.6% on average—when compared to the other three phenotypic classes; however, Admixture estimated that one citrinella hybrid possessed near-complete pine bunting ancestry. Both yellow hybrids and white hybrids tended to show comparatively higher levels of pine bunting ancestry than citrinella hybrids, but they were still skewed towards yellowhammer ancestry overall. Average yellowhammer ancestry among yellow hybrids was 74.5% while, among white hybrids, it was slightly higher at 78.2%. Only in leucocephalos hybrids did I see the average yellowhammer ancestry drop below 50%, but it was

still relatively high at 44.2%. All but one leucocephalos hybrid possessed at least 14% pine bunting ancestry and a single individual possessed near complete pine bunting ancestry. Lastly, in both hybrid females and hybrid individuals with unknown phenotypes, I saw patterns that were consistent with citrinella, yellow and white hybrids; individual ancestry was highly skewed towards yellowhammers and several individuals possessed near-complete yellowhammer ancestry.

Throughout these analyses, I found consistent discordance between genotype and phenotype as many individuals possessed the plumage phenotype of one species with the near-complete genotype of the other (Figure 3.1) or possessed an intermediate plumage phenotype with a parental genotype (Supplementary Figure 3.3). An extreme example of this discordance was seen in a pair of outliers in my PCA (Figure 3.1A). These two phenotypically divergent individuals—one sympatric pine bunting and one citrinella hybrid—clustered very closely together and both possessed near complete pine bunting ancestry (Figure 3.1B) potentially suggesting a very close familial relationship. Altogether, these patterns could imply that plumage traits within the yellowhammer and pine bunting system are controlled by a few loci of major effect rather than many loci of minor effect. Additionally, it should be noted that the presence of the aforementioned outlier pair and three additional singular outliers—two sympatric pine buntings and one leucocephalos hybrid—are not suspected to bias my PCA results as the removal of all five outliers did not compromise any of the patterns along PC1 and merely created more scatter along PC2 (Supplementary Figure 3.4).

### 3.3.2 Differentiation on the Z chromosome

In Chapter 2, I assessed patterns of genetic differentiation between allopatric populations of yellowhammers and pine buntings and identified a large island of differentiation on the Z chromosome that contained several SNPs with  $F_{ST}$  values close to one. To better understand patterns of variation in this region and infer their possible causes, I conducted a PCA of 11,147 variable SNPs located on the Z chromosome. Interestingly, I found that this analysis produced six discrete clusters—three clusters along PC1 that each separated into two further clusters along PC2 (Figure 3.2A).

Individuals were coarsely grouped along PC1 based on their species designations; most allopatric and near sympatric individuals formed clusters on either side of the axis while sympatric and hybrid individuals formed a cluster in between (Figure 3.2A). However, there was a fair amount of noise in this pattern as many sympatric or hybrid individuals also grouped together with allopatric and near sympatric individuals at the ends of the axis or, alternatively, allopatric and near-sympatric individuals appeared within the middle cluster. These discordances were likely the result of widespread genetic admixture between populations as was identified in my previous PCA and Admixture analyses (Figure 3.1). Along PC2 individuals were putatively separated based on sex with all female individuals appearing in the upper three clusters and the majority of male individuals appearing in the lower three clusters (Figure 3.2A). Two allopatric pine bunting males appeared in female cluster “1” despite their sex designations. These discrepancies may stem from low coverage of the Z chromosome in these individuals or incorrect sexing during sample collection. Individuals of unknown sex also fell into the male and female clusters, but confirming their sexual identities goes beyond the scope of this study.

To further investigate how high  $F_{ST}$  SNPs may be shaping clustering patterns in my Z chromosome PCA, I plotted the allele frequencies of SNPs with  $F_{ST}$  values equal to or greater than 0.7 for a subset of samples (Figure 3.2B-C). For females, I found that individuals in cluster “1” on the Z chromosome PCA possessed mostly pine bunting alleles at high  $F_{ST}$  SNPs while individuals in clusters “2” and “3” possessed mostly yellowhammer alleles (Figure 3.2B). Nevertheless, clusters “2” and “3” could be separated based on two small regions on either end of chromosome Z. Individuals in cluster “2” possessed pine bunting alleles at these end regions similar to individuals in cluster “1” whereas individuals in cluster “3” possessed yellowhammer alleles in these regions. This pattern is likely responsible for the intermediate nature of cluster “2” on the Z chromosome PCA (Figure 3.2A).

For male individuals, I saw a more complex pattern of inheritance at high  $F_{ST}$  SNPs (Figure 3.2C). Individuals in cluster “4” on the Z chromosome PCA possessed mostly pine bunting alleles at high  $F_{ST}$  SNPs while individuals in cluster “6” possessed mostly yellowhammer alleles. Individuals in cluster “5” were heterozygous across a large block of high  $F_{ST}$  SNPs that was bordered on either side by SNPs that were mostly homozygous for the yellowhammer allele. This large region of heterozygosity could indicate an area of low recombination (Nachman, 2002). In particular, this pattern could suggest the presence of an inversion polymorphism (reviewed in Kirkpatrick, 2010) along the Z chromosome that is segregating within yellowhammer and pine bunting system.

### 3.3.3 Demographic patterns across the hybrid zone

In the above analyses, I identified widespread admixture across the yellowhammer and pine bunting sympatric zone (Figure 3.1-3.2). To better understand the genetic relationships among individuals in the hybrid zone, I compared the ancestry of each individual against its interclass heterozygosity using triangle plots and information from high  $F_{ST}$  SNPs identified when comparing allopatric populations (Figure 3.3). When all SNPs with  $F_{ST}$  values greater than 0.6 were included in my analysis, I found that most individuals grouped in the lower corners of the triangle indicating that the great majority of their genome resembles that of one or the other parental species (Figure 3.3A). A further group of largely hybrid and sympatric individuals appeared towards the upper part of the triangle with additional allopatric and sympatric pine buntings trailing down towards one corner of the triangle.

In examining these results more closely, I found that most allopatric and near-sympatric yellowhammers and pine buntings appeared in opposing corners though pine buntings were much more variable in their positions. A large portion of phenotypic hybrids also appeared in the yellowhammer corner of the triangle while only a single hybrid individual appeared in the pine bunting corner suggesting that hybrids preferentially backcross towards yellowhammers over pine buntings. This pattern was also seen in sympatric pine buntings suggesting that many of these individuals are also highly backcrossed to yellowhammers despite their divergent appearances. Regarding the hybrids and sympatric individuals that appeared towards the top and centre of the triangle, it is likely that these birds represent earlier generation hybrids though the slight skew towards the left side of the triangle suggests some backcrossing towards pine buntings. For the allopatric and near sympatric pine buntings that also appear towards the upper

part of the triangle, it is possible that these individuals are also earlier generation hybrids or backcrosses.

With the discovery of such unusual relationships between phenotype and genotype for allopatric and near-sympatric pine buntings (Figure 3.3A), I examined the identities of the high  $F_{ST}$  SNPs included in my analysis and found that the majority were located close together on the Z chromosome. Such patterns would be expected if there is an inversion polymorphism on chromosome Z that is segregating within the system as I proposed in a previous analysis. This putative inversion has the potential to produce much of the structure I see in the triangle plot by creating close linkage between high  $F_{ST}$  SNPs. To account for such linkage, I ran the same analysis with a subset of 10 SNPs each located on a separate chromosome and with a  $F_{ST}$  score greater than or equal to 0.5 (Figure 3.3B). In the plot produced from this analysis, individuals were spread fairly evenly throughout the triangle, but the majority of individuals—from all species and geographic classes—were concentrated at the bottom. This result suggests that most individuals within the yellowhammer and pine bunting system are late-generation hybrids and that hybridization has been widespread between taxa for an extended period of time. As seen previously, allopatric and near sympatric populations were skewed towards opposing corners of the triangle, but formed much looser clusters which is still at odds with what I saw in my previous analyses (Figure 3.1). Hybrids and sympatric individuals were spread across the lower two thirds of the triangle in positions attributed to late generation hybrids or backcrosses with hybrids showing a slightly greater tendency towards backcrossing with yellowhammers. Very few individuals appeared towards the top of the triangle except for two sympatric yellowhammers which were shifted towards pure yellowhammers. It must be noted that all

demographic patterns identified in this analysis should be considered cautiously as only a limited number of SNPs were included and they likely did not capture much of the structure within the yellowhammer and pine bunting system.

### **3.3.4 Genetic underpinnings of plumage traits**

Across the yellowhammer and pine bunting system, individuals show great variation in three plumage traits: the main body colour, the amount of chestnut at the brow and the amount of chestnut at the throat. To determine what areas of the genome might be controlling this phenotypic variation, I conducted admixture mapping for each of the three traits using GEMMA. Following an internal filtering step, GEMMA tested 220,220 associations between variable SNPs and variation in body colour, 220,124 associations between variable SNPs and variation in the amount of chestnut at the brow and 220,307 associations between variable SNPs and variation in the amount of chestnut at the throat (Figure 3.4).

Beginning with body colour, I found that variation in the amount of yellow and white body plumage among bird individuals was associated with genetic variation on chromosome 20 and chromosome Z (Figure 3.4A). One SNP on chromosome 20 and four SNPs on chromosome Z were significantly associated with phenotypic variation following a Bonferroni correction. The SNP that was most significantly associated with body colour was on chromosome Z, but was not found within any particular annotated gene (Table 3.1). Nevertheless, one of the remaining three significant SNPs on chromosome Z was found within an annotated gene—*ABCA1*—and the significant SNP on chromosome 20 was found within the gene, *CHD6*. Possible associations between phenotypic and genetic variation were further seen on chromosomes 2, 7 and 14, but

none of the SNPs examined on these chromosomes passed the high significance threshold (Figure 3.4A). A number of additional SNPs on chromosomes 20 and, particularly, on chromosome Z showed a high estimated association with background colour though did not pass the significance threshold. On chromosome Z, these less strongly associated SNPs were separable into two major peaks.

When considering the amount of chestnut plumage at the brow in parental and hybrid individuals, I found a strong association between this trait and genetic variation on the Z chromosome (Figure 3.4B). Thirty-four variable SNPs split between two regions of chromosome Z were significantly associated with variation in brow plumage. The SNP most significantly associated with plumage patterns was not found within an annotated gene (Table 3.2). However, a SNP with only a slightly less significant association was found within the gene, GRAMD3, along with another, more weakly associated SNP. Three additional significant SNPs were found within the FER gene and two significant SNPs were found within the gene PJA2. Further annotated genes that contained a single SNP significantly associated with variation in brow plumage included: EFNA5, APC, MCC, CEP120 and ECPAS. A number of SNPs on chromosome 1 showed possible associations with brow plumage variation, but did not pass the significance threshold (Figure 3.4B). Additionally, as was seen for body colour, a large number of SNPs separated into two peaks on the Z chromosome were highly but insignificantly associated with brow plumage.

Comparable to brow plumage patterns, I saw a strong association between genetic variation on the Z chromosome and throat plumage patterns in the yellowhammer and pine bunting system (Figure 3.4C). Eight SNPs on chromosome Z were significantly associated with

this plumage trait. The most significantly associated SNP was not found within an annotated gene, but the second most associated SNP was found within the gene GRAMD3 (Table 3.3). None of the other significant SNPs were found within annotated genes. As was seen for brow plumage patterns, throat plumage patterns were weakly associated with a subset of SNPs on chromosome 1 and highly associated with two peaks of SNPs on the Z chromosome (Figure 3.4C). Nevertheless, none of these SNPs passed the significance threshold.

Interestingly, I found similar associations between phenotypic variation and genetic variation on the Z chromosome across all three plumage traits, but particularly when looking at eyebrow and throat plumage patterns (Table 3.1-3.3). Two SNPs were significantly associated with all three plumage traits, one SNP was associated with both body colour and brow plumage traits and six SNPs was associated with both brow and throat plumage traits. As well, the two peaks of insignificantly but highly associated SNPs seen on the Z chromosome were largely conserved across all three plumage traits with the greatest overlap seen between association patterns of eyebrow plumage and of throat plumage.

To investigate potential dominance interactions between different plumage phenotypes and their related genotypes, I created contingency tables and balloon plots that compared the phenotypic scores of all three plumage traits against any significantly associated SNPs. Potential dominance interactions were not detected for significant SNPs located on chromosome 20, so I focused my analysis on chromosome Z. Interestingly, SNPs related to plumage variation on chromosome Z possessed variable dominance patterns depending on whether they were located inside or outside of the putative chromosomal inversion. The dominance patterns seen across chromosome Z can be summarized by the four SNPs highlighted in Figure 3.5A. Two of these

SNPs occur within the putative inversion and the remaining two occur on either side of the inversion. All four SNPs are highly differentiated between allopatric populations of yellowhammers and pine buntings.

The SNP Z.4835388—shown in green—is located within a region to the left of the putative chromosomal inversion (Figure 3.5A). This SNP was significantly associated with variation in background plumage colour, but the genomic region in which it is located was also weakly associated with variation in facial plumage traits (Table 3.1; Figure 3.6). Comparisons of phenotypic scores against Z.4835388 genotypes did not suggest dominance patterns for this SNP at any of the traits of interest as heterozygotes did not show a strong bias towards one ancestral plumage phenotype and both homozygotes possessed variable phenotypes (Figure 3.5B).

The SNP Z.18131016—shown in magenta—is located near the beginning of the putative inversion on chromosome Z (Figure 3.5A) and was significantly associated with brow plumage variation within the yellowhammer and pine bunting system (Table 3.2; Figure 3.6B). The genomic region surrounding this SNP was also moderately associated with variation in background colour and variation in throat plumage among individuals (Figure 3.6A,C). For Z.18131016, I saw patterns consistent with dominance interactions as the heterozygous and one homozygous genotype showed a strong bias towards phenotypic scores of “7”—the pure pine bunting phenotype—for all plumage traits (Figure 3.5B). The remaining homozygous genotype was associated with more variable phenotypic scores across the three traits including those belonging to phenotypic yellowhammer and hybrid individuals. Altogether, these results suggest that the putative “pine bunting” allele at this SNP is dominant over the putative “yellowhammer” allele.

The SNP Z.26812248—shown in orange—is located near the end of the putative inversion located on chromosome Z (Figure 3.5A) and was significantly associated with variation in brow and throat plumage between yellowhammers, pine buntings and their hybrids (Table 3.2-3.3; Figure 3.6B-C). The genomic region surrounding this SNP was moderately associated with variation in background plumage colour (Figure 3.6A). As was observed for SNP Z.18131016, comparisons of phenotypic scores and Z.26812248 genotypes showed patterns consistent with a putative “pine bunting” allele being dominant over a putative “yellowhammer” allele for all three plumage traits considered (Figure 3.5B). As observed earlier, the heterozygous and one homozygous genotype were highly associated with pure pine bunting phenotypic scores while the other homozygous genotype was associated with variable phenotypic scores.

Finally, the SNP Z.59571043—shown in blue—is located within a region to the right of the putative inversion on chromosome Z (Figure 3.5A). This SNP and its surrounding genomic region were significantly associated with all three plumage traits (Table 3.1-3.3; Figure 3.6). Nevertheless, despite these associations, heterozygous and homozygous genotypes did not show any bias towards a particular ancestral phenotype at any of the plumage traits of interest suggesting an absence of dominance interactions at this SNP (Figure 3.5B).

I noted one additional genomic pattern while investigating dominance interactions associated with plumage variation between yellowhammers, pine buntings and their hybrids. Three SNPs—Z.23774941, Z.23774945 and Z.23774949—that are located within a 10bp region of the putative chromosome Z inversion and that were significantly associated with brow and throat variation (Table 3.2-3.3) possessed only a heterozygous genotype and a single homozygous genotype in the yellowhammer and pine bunting system. These observations are

consistent with a duplication of a genomic region or gene within the chromosomal inversion. Unfortunately, none of the three SNPs were associated with a particular annotated gene such that I am unable to relate this duplication to patterns of plumage variation or more general patterns of variation within this system.

### **3.4 Discussion**

The yellowhammer and the pine bunting are currently considered separate species—this classification attributable to their highly divergent plumage patterns and also reflected in their moderately divergent song and ecology (Panov et al. 2003; Rubtsov & Tarasov, 2017). In Chapter 2, I reported that allopatric populations of yellowhammers and pine buntings are genetically distinct from one another and that this distinction is driven mainly by a large “island of differentiation” on the Z chromosome. In this chapter, characterization of genomic differentiation across the system supported both of the aforementioned conclusions and also suggested that the genetic distinctness of these taxa extends into the near-sympatric zone. Nevertheless, widespread hybridization between yellowhammers and pine buntings in sympatry (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) as well as evidence of mitochondrial (Irwin et al. 2009) and mitonuclear gene introgression (Chapter 2) implies that there is only limited reproductive isolation between groups. Following this reasoning, analysis of sympatric individuals and hybrids revealed moderate admixture between yellowhammers and pine buntings.

From my genomic analysis of the yellowhammer and pine bunting sympatric zone, I can make several inferences about the state of reproductive barriers between these taxa. First, prezygotic barriers appear to be weak. This idea is supported by the prevalence of admixture between putatively pure sympatric populations and by the large amount of hybridization documented between taxa over decades of observation (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Prezygotic barriers related to assortative mating and species recognition may be particularly weak in this system as mapping of mating territories of sympatric individuals determined that males of both species are equally territorial towards conspecifics and heterospecifics (Rubtsov & Tarasov, 2017). These results suggest that males view individuals both within and outside of their taxon as rivals for resources. Because male territorial response is often used as a proxy for female mate choice in birds (e.g. Balakrishnan & Sorenson, 2006; Uy et al. 2009; Cruz-Yepez et al. 2020), I can argue that females may be non-discriminatory or only weakly discriminatory when choosing a mate if members of both taxa are present. As a result, females' broad acceptance of a diversity of male appearances as potential mates may drive widespread admixture and hybridization between taxa across the sympatric zone.

Like prezygotic barriers, postzygotic barriers also appear to be limited between yellowhammers and pine buntings. Within the sympatric zone, hybridization is extensive with hybrid proportions reaching almost 50% in several surveyed locations (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). Furthermore, genomic analysis of hybrids within my dataset revealed that very few individuals possessed the ancestry proportions expected in F1 hybrids. This lack of early generation hybrids was echoed in triangle plots that represented most

hybrids and, indeed, most individuals across the system as late generation hybrids or backcrosses. The presence of so many late-generation hybrids and backcrosses suggests that hybrids are not only viable as has been found in previous research (Lohrl, 1967 cited in Panov et al. 2003), but also fertile and that natural selection is not acting strongly against them. In other words, it is possible that hybrids are similar in fitness or just as fit as offspring with pure parentage.

Although the triangle plots in this study must be interpreted cautiously due to a limited number of markers and linkage between highly differentiated SNPs, further evidence for the absence of selection against hybrids can be seen in the continued expansion of the sympatric zone and near-immediate occurrence of hybridization within newly sympatric regions that has continued since it was first observed (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). In past research, scientists have suggested that the yellowhammer and pine bunting sympatric zone has expanded 1000 km east and 350 km west in 100 and 25 years respectively and that it will likely continue to expand in the future. This continued expansion is indirectly supported by the appearance of an individual in my dataset within the near-sympatric zones that possessed admixture from its sister taxa. Hybridization within newly sympatric regions is described as being high—with phenotypic hybrid proportions reaching as high as 50% initially (Rubtsov & Tarasov, 2017). Additional research is greatly needed to monitor the development of the yellowhammer and pine bunting sympatric area and to compare the fitness of hybrid and pure offspring in order to directly assess the state of postzygotic barriers between taxa.

Despite the apparent lack of strong prezygotic and postzygotic barriers between yellowhammers and pine buntings within their sympatric zone, it is possible that specific areas of the genome may house alleles that represent weak reproductive barriers not discernable when analyzing genome-wide patterns. In chapter 2, I identified a large “island of differentiation” between allopatric populations on the Z chromosome that may promote reproductive isolation between taxa and that may better separate individuals along putative species divisions. My analysis revealed six genetic clusters within the yellowhammer and pine bunting system driven by loci on this sex chromosome. Separation of the genetic clusters based on biological sex was likely the result of genetic reads from the W chromosome within female samples. Because I used a reference genome that did not include the W chromosome, it is possible that W chromosome reads that were homologous to regions of the Z chromosome mapped to this latter chromosome. This could have created a signal of increased heterozygosity within females that would distinguish them from males. The separation of individuals within each sex into three distinct genetic clusters was attributable to a strong association of genotypes along a broad region of the Z chromosome. Specifically, it appears that the intermediate genetic cluster among males was caused by a large block of heterozygosity on the Z chromosome. This heterozygous block was uniform among intermediate males suggesting tight linkage between markers in this region. Altogether, these results are consistent with an area of low recombination between genetically divergent genomic blocks (Nachman, 2002) potentially as a result of an inversion polymorphism (reviewed in Kirkpatrick, 2010) that is segregating within the yellowhammer and pine bunting system. Similar patterns were not observed among females because, as the heterogametic sex, they only possess one Z chromosome. As such, females cannot be heterozygous for regions

along this chromosome. Instead, recombination outside of the putative inversion appears to have produced intermediate females who possess the yellowhammer version of the inversion in a pine bunting genetic background. Further, double recombination within the inversion may have also contributed to the intermediacy of some females.

Chromosomal inversions preserve linkage between alleles that would otherwise be separated by recombination allowing for the accumulation of genetic differentiation among individuals (reviewed in Kirkpatrick, 2010). In a situation where separate populations become fixed for different versions of a chromosomal inversion, divergent mutations that contribute to either prezygotic or postzygotic barriers may amass within inversions and act together to cause substantial reproductive isolation between groups leading to speciation (Noor et al. 2001; Rieseberg, 2001). In the yellowhammer and pine bunting system, different versions of the putative chromosome Z inversion appear to have evolved in separate taxa—likely in allopatry prior to secondary contact—as suggested by the observation that segregation of this inversion somewhat corresponds to species divisions in allopatric and near-sympatric zones. However, as was seen in genome-wide analyses, there is little evidence of reproductive barriers related to this putative inversion between yellowhammers and pine buntings. Divergent forms of the inversion have flowed relatively uninhibited across species boundaries without large fitness effects. This was evident from the presence of many sympatric pine buntings who were homozygous for the “yellowhammer form” of the inversion and additional individuals that were heterozygous for the inversion including a number of allopatric pine buntings. From these results I propose that, if a putative inversion did arise between yellowhammers and pine buntings prior to secondary contact, the inversion does not appreciably contribute towards reproductive barriers between taxa

in the present perhaps because different forms of the inversion did not accumulate sufficient genetic differentiation (Noor et al. 2001; Rieseberg, 2001). As a result, the inversion moves freely among individuals acting more as a polymorphism of the system rather than a characteristic by which to define the two species.

Putatively pure yellowhammers and pine buntings are highly divergent in plumage traits and these characteristics have been used as chief indicators of species identity for decades (Panov et al. 2003; Rubtsov & Tarasov, 2017). My results showing some discordance between the phenotype and genotype of individuals suggest that plumage may not be a clear indicator of species or ancestry. However, these patterns also indicate that plumage characteristics may be controlled by a small number of genes of large effect and this fact—combined with the great variation in phenotype seen among hybrids—provided me with the opportunity to investigate the loci responsible for colouration patterns in this system. Indeed, I found strong associations between three plumage traits and genetic variation among yellowhammers, pine buntings and their hybrids. Interestingly, all three of the analyzed colouration traits—background colour, amount of chestnut at the brow and amount of chestnut at the throat—were strongly correlated to loci on the Z chromosome with the latter two characteristics being specifically correlated with loci within the alleged chromosomal inversion. These associations between brow and throat patterning and the chromosomal inversion imply that this region plays an important part in mediating facial plumage patterns within the yellowhammer and pine bunting system and highlights the striking congruence between the genetic loci mediating these traits. Such a connection was identified previously by Panov et al. (2003) who found a moderate correlation between brow and throat colouration in yellow and white hybrids within the sympatric zone. The

background colouration trait showed a similar, but insignificant association with the putative chromosomal inversion and an additional significant association with a region on chromosome 20 that is unique to this characteristic.

There are several ways to interpret the observed association between colouration traits and the putative chromosomal inversion on the Z chromosome in the yellowhammer and pine bunting system. First, if I consider some of the ideas proposed in chromosomal speciation models that incorporate chromosomal inversions (Noor et al. 2001; Rieseberg, 2001), it is possible that an inversion polymorphism evolved between allopatric yellowhammers and pine buntings and became associated with plumage genes. Differentiation within the inversion created divergent plumage phenotypes between the taxa and recombination suppression along the inversion (reviewed in Kirkpatrick, 2010) preserved this differentiation. As a result, ancestral yellowhammer and pine bunting plumage phenotypes were and are somewhat maintained within the sympatric zone despite pervasive gene flow. Such a proposal must be considered with a certain amount of caution however, as tight linkage between loci within the putative inversion has the potential to impact the accuracy of admixture mapping procedures.

Strong associations were seen between variation in the three plumage traits that differentiate yellowhammers and pine buntings and loci on the Z chromosome. In all three cases, the SNP that was most significantly correlated with each trait was not found within or near an annotated gene known to be related to changes in plumage phenotypes. However, additional SNPs that were significantly associated with plumage variation were located in genes connected to colouration patterning that could mediate phenotypes within this system.

Beginning with variation in background plumage colour which is yellow in phenotypic yellowhammers, white in phenotypic pine buntings and intermediate in phenotypic hybrids, I found a highly significant association with this trait and the gene ABCA1 (ATP Binding Cassette Subfamily A Member 1) on the Z chromosome. The ABCA1 gene is involved in the translocation of phospholipids across membranes as well as the formation of high-density lipoproteins (HDLs; reviewed in Oram & Vaughan, 2000) that are used to transport specific carotenoids around the body (Clevence & Bieri, 1993). A mutation in ABCA1 among the Wisconsin hypoalpha mutant (WHAM) chicken is responsible for low levels of HDLs and, as a result, low levels of carotenoids in chicken tissues (Connor et al. 2007). This mutation further produces a phenotype of a white beak and skin versus a yellow beak and skin as well as colourless plasma (Attie et al. 2002). Although ABCA1 has not been implicated in changes in plumage colouration, these results showing how gene mutations have a direct effect on carotenoid transport and cause a shift in colour equivalent to what is seen in my system suggests that it is a strong candidate for regulation of plumage patterning between yellowhammers and pine buntings. In addition to ABCA1, variation in background plumage colour was also significantly associated with a gene on chromosome 20—CHD6 (Chromodomain Helicase Binding Protein 6). The CHD6 gene is part of a complex that regulates chromatin remodeling during gene expression (Manning & Yusufzai, 2017). This gene has not been connected to the regulation of either plumage patterning or colouration in the literature suggesting that it has an unknown function controlling these traits or is linked to a gene that controls these traits but does not appear in my reduced representation sequencing dataset.

I found tight congruence between the admixture mapping results for brow and throat colouration which was consistent with the moderate correlation identified between these traits by Panov et al. (2003). Only one gene was significantly associated with both brow and throat plumage: GRAMD3 (GRAM domain containing 3). Little is known about the function of GRAMD3, but it is highly associated with retinal pigment epithelial cells in humans (Strunnikova et al. 2010). More notably, GRAMD3 has also been proposed as a candidate gene for the *Id* locus (Xu et al. 2017). The *Id* locus controls dermal shank pigmentation in certain breeds of chickens which is a trait defined by the accumulation of melanin in the dermis of the leg (McGibbon, 1974). This connection between GRAMD3 and melanin colouration is consistent with the presumed melanin-based brow and throat patterning in my system indicating that this gene may be important to the regulation of these traits. Additional genes associated with variation in brow plumage included FER (FER tyrosine kinase), which is associated with melanoma in humans (Ivanova et al. 2019), EFNA5 (Ephrin A5) which is associated with piebald skin pigmentation in sheep (García-Gómez et al. 2011) and APC (APC regulator of Wnt signaling pathway) which is associated with feather development (Widelitz et al. 2000). The CEP120 (Centrosomal protein 120), PJA2 (Praj ring finger ubiquitin ligase 2), MCC (MCC regulator of Wnt signaling pathway) and ECPAS (ECM29 proteasome adapter and scaffold) genes were also associated with brow plumage, but their known functions are not linked to any of colouration, plumage patterning or feather development. Although it is possible that these additional genes play a role in brow plumage patterning, none of these genes are significantly associated with throat plumage despite reported correlations between these traits (Panov et al. 2003). All of these brow-associated genes are also found within the putative inversion

polymorphism meaning that, rather than controlling brow plumage within this system, the observed associations could be the result of linkage to a specific gene within the inversion such as GRAMD3 that is actually regulating colouration patterns

When comparing phenotypic variation to associated SNP genotypes, I found genomic patterns consistent with dominance interactions for loci within the putative chromosomal inversion located on chromosome Z. In particular, two SNPs that are positioned within the inversion (towards opposite ends) both showed patterns consistent with pine bunting alleles being dominant to yellowhammer alleles in their effects on the plumage traits. In contrast, SNPs on chromosome Z that are located outside of the inversion but that were significantly associated with plumage variation among individuals did not show genomic patterns that strongly suggest dominance interactions. Because recombination suppression produces high linkage between loci within a chromosomal inversion (reviewed in Kirkpatrick, 2010), it is not surprising that SNPs within the putative inversion show similar dominance patterns. In fact, I can broaden these results to hypothesize that the pine bunting version of the inversion is dominant to the yellowhammer version in its role mediating plumage phenotypes in this system. Nevertheless, it is unclear how SNPs outside of the inversion impact dominance interactions to affect plumage variation, but these loci are likely responsible for some of the unexplained variation seen in my dominance analysis. Further research that characterizes the genomic structure of the yellowhammer and pine bunting Z chromosomes—especially in the region of the putative inversion—and that characterizes functional differences between divergent alleles of important candidate genes would greatly facilitate my understanding of the genetic mechanisms controlling plumage variation in this system.

If the pine bunting form of the inversion is dominant over the yellowhammer form in its effect on plumage patterns, this relationship may explain the observation that sympatric yellowhammers have less overall genomic admixture than sympatric pine buntings. In this situation, an individual who is homozygous or heterozygous for the pine bunting form of the inversion would display a pine bunting phenotype in the sympatric zone despite potentially high admixture from yellowhammers at other genomic regions. As such, admixture is “masked” within sympatric pine buntings by dominance. For individuals who are homozygous for the yellowhammer form of the inversion, admixture would not be “masked” and could potentially influence the plumage phenotype. Here, individuals who possess relatively pure yellowhammer ancestry in addition to two versions of the yellowhammer form of the inversion could display a pure yellowhammer phenotype while individuals who possess some admixture could display one of the four hybrid phenotypes. Altogether, this hypothesis is consistent with the genotypic and phenotypic trends I saw across the yellowhammer and pine bunting system. However, this reasoning does not explain the overall skew towards yellowhammer genotypes observed in the sympatric zone or reports that pine bunting and leucocephalos hybrid populations are consistently replaced by yellowhammers and other phenotypic hybrids respectively over time in the sympatric zone (Panov et al. 2003; Rubtsov and Tarasov, 2017). Keeping these latter observations in mind, it is important to contemplate whether something like preferential mating towards yellowhammers or fitness disparities may also be influencing admixture patterns in addition to dominance effects.

In regard to preferential mating as an explanation for the bias towards yellowhammer ancestry in the sympatric zone and for the replacement of pine bunting and leucocephalos hybrid

individuals, one potential reason for this behaviour could be an innate preference towards yellowhammer phenotypes. Because colouration is highly divergent between taxa, this preference could be aimed towards yellowhammer plumage patterns particularly their bright yellow body colour. In the past, the link between innate colour preferences and mate choice have been explored in guppies (*Poecilia reticulata*) where individuals' predisposition towards orange objects—particularly orange food—drives female preference for males with large orange spots (Rodd et al. 2002). This idea of sensory bias is somewhat complicated in the yellowhammer and pine bunting system by evidence of discordance between phenotype and genotype such that a yellowhammer plumage pattern may not equate to pure or high yellowhammer ancestry. However, when considering overall genomic patterns, it does appear that a preference towards yellow colouration constitutes a preference towards yellowhammer alleles.

Another possible explanation for preferential mating could be that genetic yellowhammers are socially dominant to genetic pine buntings and hybrids such that yellowhammer males are able to secure larger and better-quality territories and to attract a greater number of females within the sympatric zone. Instances of social dominance influencing hybridization and backcrossing among species have been suggested in several chickadee hybrid zones (Bronson et al. 2003; Grava et al. 2012). In both Carolina (*Poecile carolinensis*) by black-capped (*Poecile atricapillus*; Bronson et al. 2003) and black-capped by mountain chickadee (*Poecile gambeli*; Grava et al. 2012) social interactions, dominance of the former over the latter species results in biased mating patterns whereby females of the latter species are observed mating with heterospecific males, but not the other way around. Similar asymmetrical mating where genetic yellowhammers are able to secure genetic pine bunting and hybrid mates, but not

the other way around could drive unidirectional gene flow and produce the skew towards yellowhammer ancestry that I observe in my genomic analysis.

Alternatively, if preferential mating is not responsible for the observed bias in genomic ancestry, it is also possible that this pattern is driven by weak fitness differences between individuals with different genotypes such that higher yellowhammer ancestry yields slightly higher fitness. If this were the case, adaptive introgression may be responsible for the excess of yellowhammer genotypes across the sympatric zone. Adaptive introgression has been proposed to explain the lack of differentiation in the mtDNA of these two species (Irwin et al. 2009) and, in chapter 2, I also found some evidence of mitonuclear co-introgression though presumably in the direction of pine buntings into yellowhammers. As stated earlier, future research that compares the fitness of hybrid and genotypically pure individuals and that better documents interactions within the sympatric zone is sorely needed to better understand the forces shaping genomic patterns in this area.

My genomic analysis of yellowhammers, pine buntings and their hybrids across Eurasia highlights the dynamic and unpredictable nature of population divergence and emphasizes how the strength of reproductive barriers between taxa mediates the delicate balance between speciation and species merging. In this system, allopatric and near-sympatric populations of yellowhammers and pine buntings form genetically discrete clusters well separated from their sister species. However, due to a lack of strong reproductive barriers between taxa, this relationship breaks down in sympatry where hybridization and genetic admixture is prevalent. Much of the genetic differentiation seen between allopatric and near-sympatric yellowhammers and pine buntings is driven by a large “island of differentiation” on the Z chromosome which

appears to house an inversion polymorphism. This inversion is associated with plumage traits that are highly divergent between taxa and that are the main way by which these species are distinguished (Panov et al. 2003; Rubtsov & Tarasov, 2017). Different forms of this inversion generally follow species designations, but are beginning to flow between taxa much like the rest of the nuclear genome. Nevertheless, the non-recombining nature of this large linkage block hints that it could continue to contribute towards plumage polymorphisms within this system for a very long time.

Taken together, the above observations create an interesting narrative for past and future interactions between yellowhammers and pine buntings. During the last series of Pleistocene glaciations, separate populations diverged genetically while in allopatric isolation. Such divergence included the evolution of a putative inversion polymorphism that became linked to loci controlling plumage patterning and colouration traits between taxa. Eventually, the accumulation of mutations within the inversion produced highly divergent plumage phenotypes between what would be classified as yellowhammers and pine buntings. Yet, genetic divergence within and outside the inversion was not great enough to create strong reproductive barriers between groups, resulting in much hybridization and gene flow between them upon secondary contact. While the mitochondrial genome and a large part of the nuclear genome have become homogenized between the species, the inversion on chromosome Z has resisted gene flow and allowed for the retention of ancestral plumage phenotypes within the sympatric zone despite the fact that the majority of sympatric individuals are now genetically admixed. Looking forward into the future, it appears likely that the area of sympatry and hybridization between yellowhammers and pine buntings will continue to expand (Panov et al. 2003; Rubtsov &

Tarasov, 2017) potentially moving towards complete merging into one species. Yet, as the different forms of the chromosomal inversion and its associated alleles also segregate across the system, it remains possible that some version the ancestral yellowhammer and pine bunting plumage patterns will be preserved as a highly divergent phenotypic polymorphism. Such an uncommon situation highlights hybridization as both a destructive and creative force in evolution. For, although hybridization could be driving the collapse of two somewhat divergent taxa, it could also be creating a single taxon with an unusually large amount of variation that would possess great evolutionary potential in the future. These contrasting views of hybridization warrant further discussion—in regard to both the yellowhammer and pine bunting system and other systems—as evolutionary researchers continue to investigate the inner workings of the speciation process.

**Table 3.1.** Identities of SNPs showing significant association with phenotypic variation in the background colour of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form  $-\log(p\text{-value})$ . Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29811. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found within an annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Brow” indicates the amount of chestnut plumage at the brow and “Throat” indicates the amount of chestnut at the throat. “NA” indicates that a particular significant SNP was unique to the background plumage trait.

Chromosome	Position	$-\log(\text{LRT } p\text{-value})$	Gene	Significant for additional traits
20	604488	17.58	CHD6	NA
Z	4835388	17.17	ABCA1	NA
Z	59570976	18.72	NA	Brow, Throat
Z	59571043	19.90*	NA	Brow, Throat
Z	59575101	16.15	NA	Brow

**Table 3.2.** Identities of SNPs that showed a significant association with phenotypic variation in the amount of chestnut plumage at the brow of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form  $-\log(p\text{-value})$ . Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29768. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found in an annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Background” indicates the colour of the background plumage and “Throat” indicates the amount of chestnut plumage at the throat. “NA” indicates that a particular significant SNP was unique to the brow plumage trait.

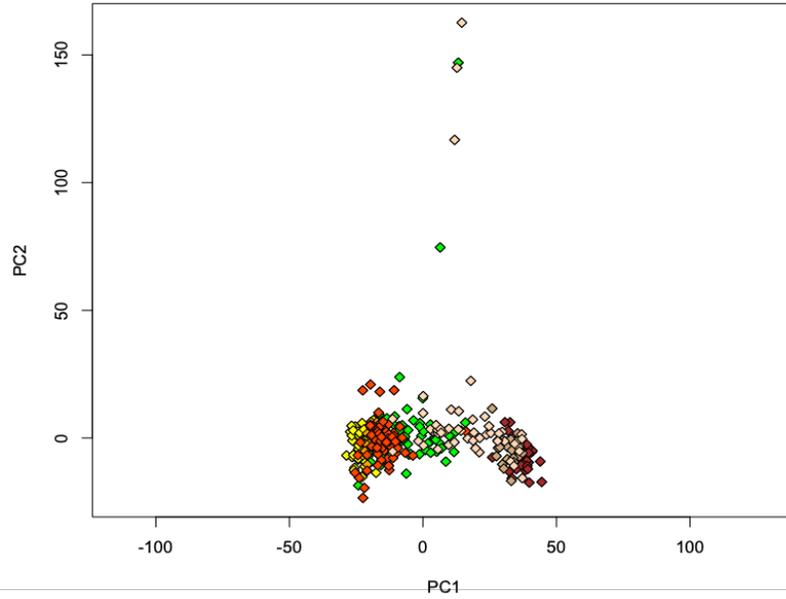
Chromosome	Position	$-\log(\text{LRT } p\text{-value})$	Gene	Significant for another phenotype?
Z	17694478	15.80	NA	NA
Z	18127032	15.75	NA	NA
Z	18127137	16.16	NA	NA
Z	18131016	17.97	NA	NA
Z	19020732	17.49	NA	NA
Z	19196302	15.47	EFNA5	NA
Z	19708803	16.08	NA	NA
Z	19768210	17.75	FER	NA
Z	19821523	17.86	FER	NA
Z	19821527	17.24	FER	NA
Z	19939726	16.58	PJA2	NA
Z	19939881	16.56	PJA2	NA
Z	21223538	15.35	APC	NA
Z	21386283	16.23	MCC	NA
Z	21843152	15.84	NA	NA
Z	23035555	15.75	CEP120	NA
Z	23774941	18.85	NA	Throat

Chromosome	Position	$-\log(\text{LRT p-value})$	Gene	Significant for another phenotype?
Z	23774945	18.85	NA	Throat
Z	23774949	18.85	NA	Throat
Z	25116054	15.58	NA	NA
Z	26023933	17.01	ECPAS	NA
Z	26402437	16.41	NA	NA
Z	26757019	22.48*	NA	Throat
Z	26760575	16.78	NA	NA
Z	26812200	16.45	GRAMD3	NA
Z	26812248	21.80	GRAMD3	Throat
Z	26869834	15.84	NA	NA
Z	27304119	15.66	NA	NA
Z	59570976	19.69	NA	Background, Throat
Z	59571035	15.99	NA	NA
Z	59571043	20.24	NA	Background, Throat
Z	59575101	16.03	NA	Background
Z	59675542	18.13	NA	NA
Z	59953541	19.32	NA	Throat

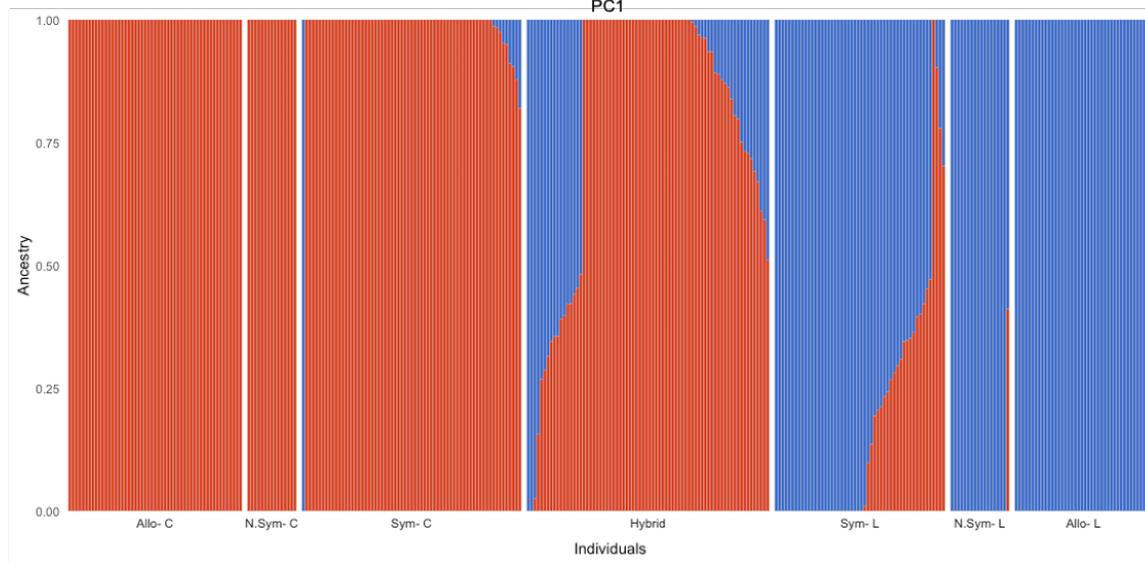
**Table 3.3.** Identities of SNPs that showed a significant association with phenotypic variation in the amount of chestnut plumage at the throat of bird individuals within the yellowhammer and pine bunting system. Locations of SNP are indicated by the “Chromosome” column which indicates the chromosomal location and the “Position” column which indicate the base pair position. P-values were calculated using a likelihood ratio test with the GEMMA program and are written in the form  $-\log(p\text{-value})$ . Larger values indicate greater significance. The Bonferroni corrected significance threshold was set at 15.29851. SNPs that occur within a gene are indicated in the “Gene” column with gene names written as they appear within the zebra finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010). “NA” indicates that a SNP was not found in an annotated gene. SNPs that are significantly associated with another plumage trait are indicated in the “Significant for additional traits” column where “Background” indicates the colour of the background plumage and “Brow” indicates the amount of chestnut plumage at the brow. “NA” indicates that a particular significant SNP was unique to the throat plumage trait.

Chromosome	Position	$-\log(\text{LRT } p\text{-value})$	Gene	Significant for another phenotype?
Z	23774941	15.64	NA	Brow
Z	23774945	15.64	NA	Brow
Z	23774949	15.64	NA	Brow
Z	26757019	17.47	NA	Brow
Z	26812248	18.91	GRAMD3	Brow
Z	59570976	17.12	NA	Background, Brow
Z	59571043	16.72	NA	Background, Brow
Z	59953541	20.04*	NA	Brow

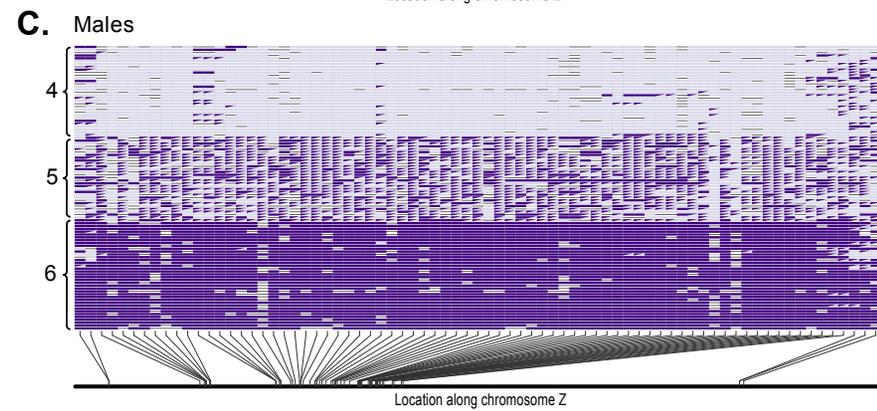
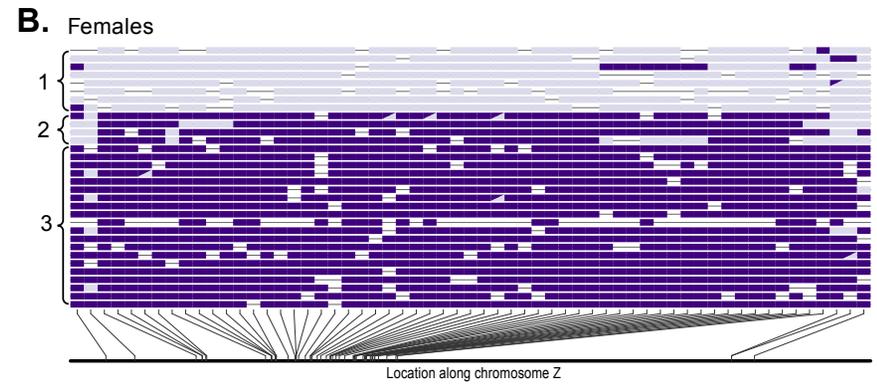
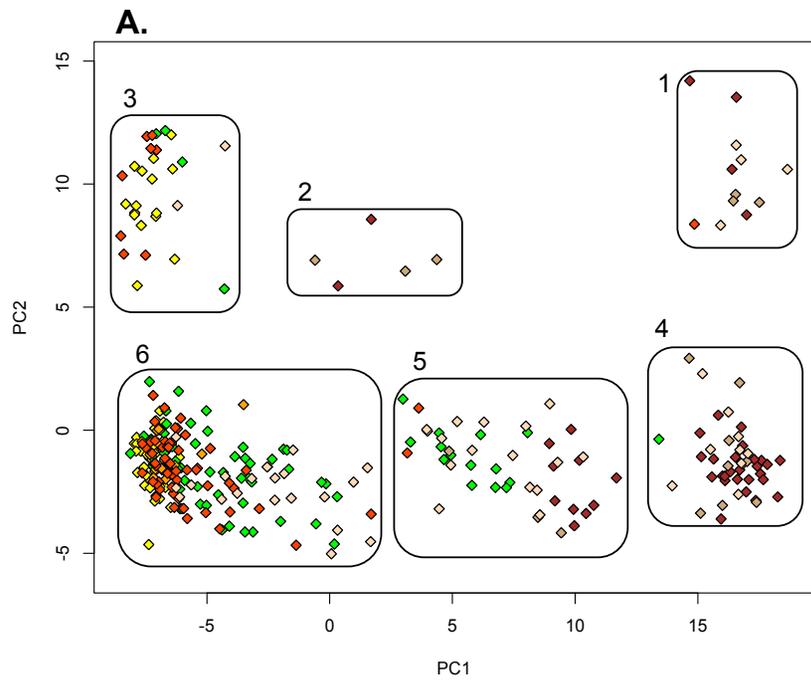
A.



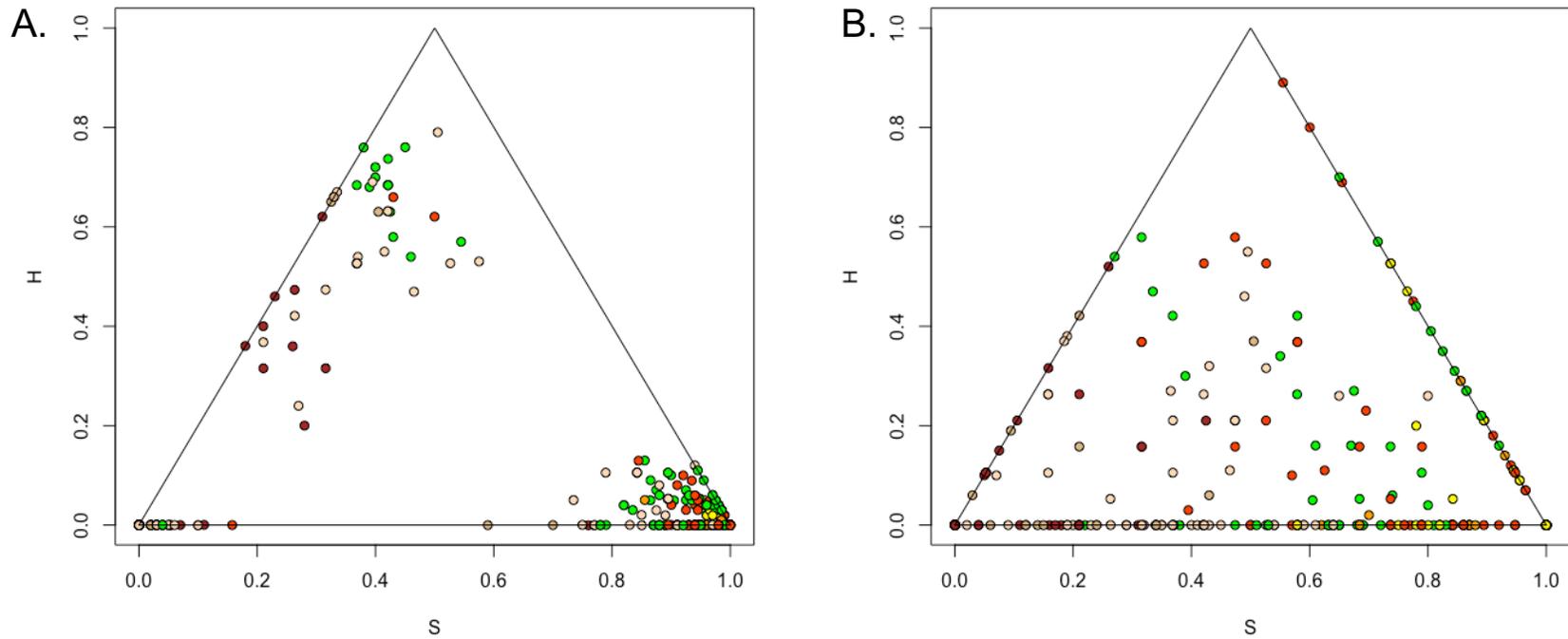
B.



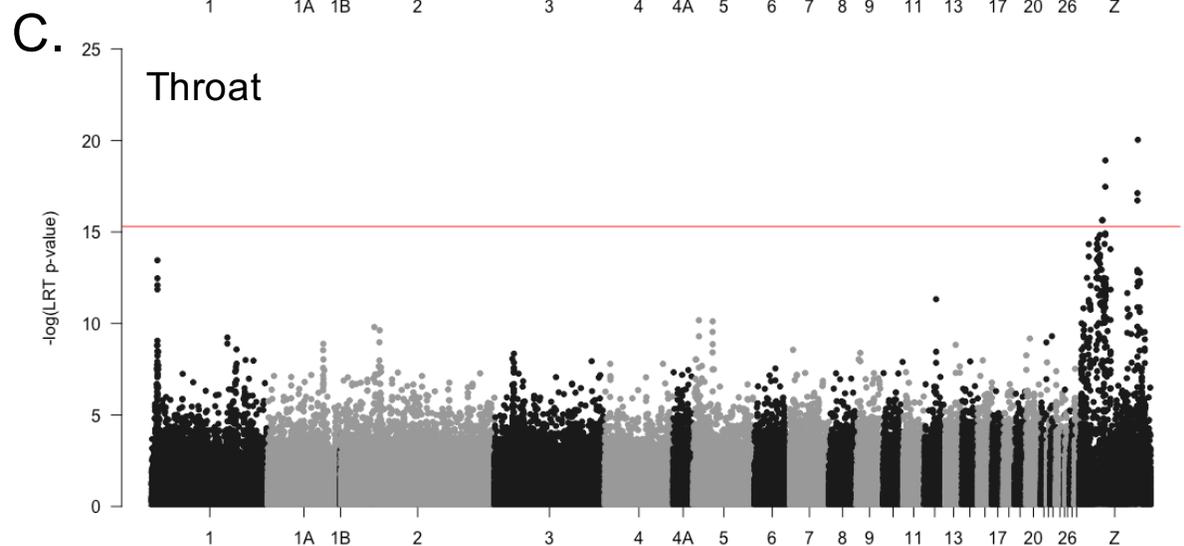
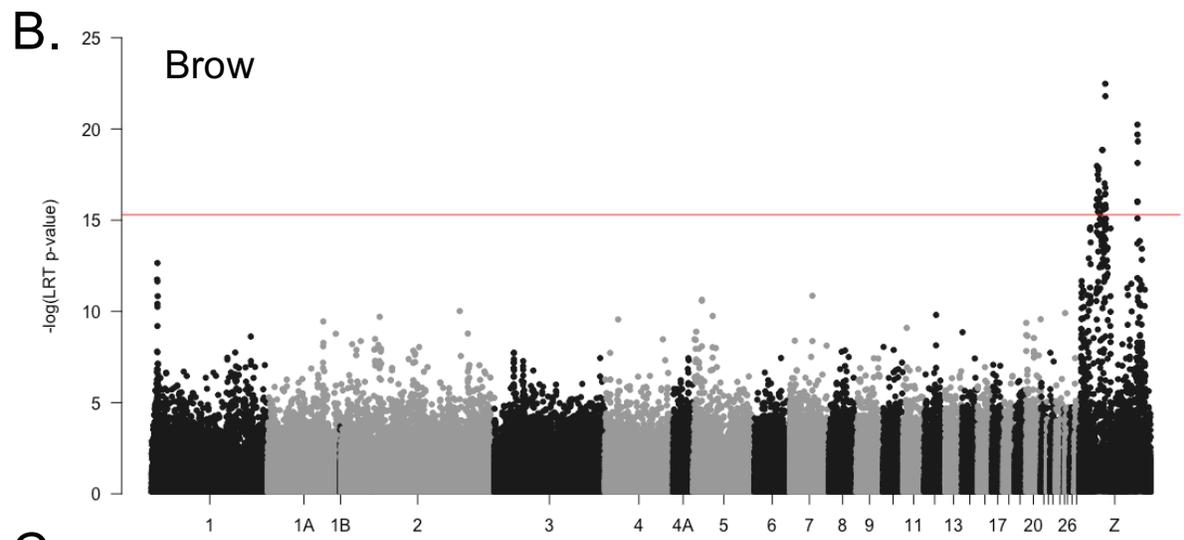
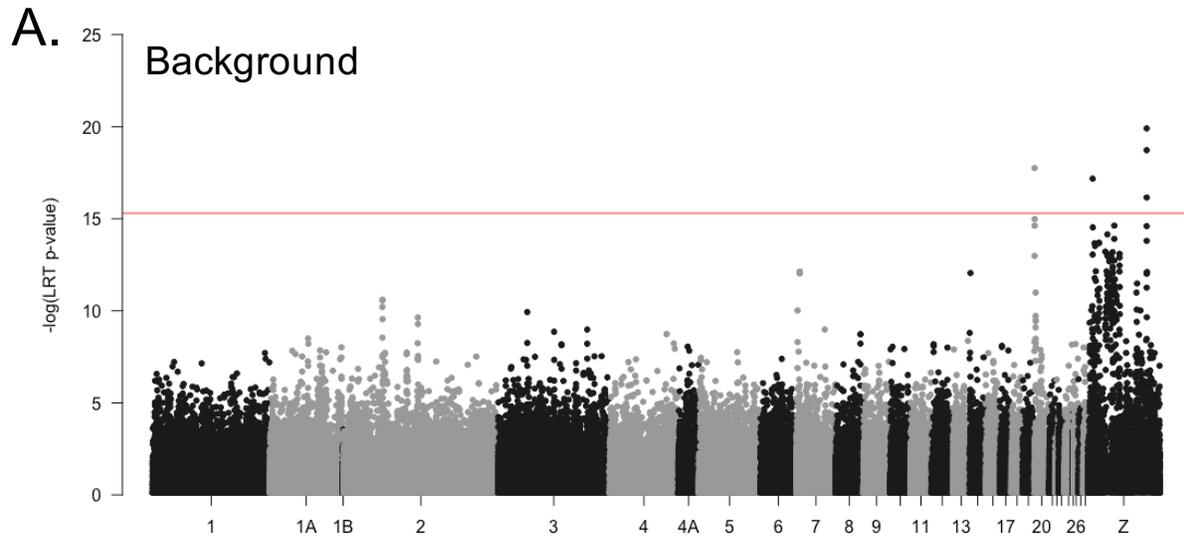
**Figure 3.1.** Genetic relationships between allopatric yellowhammers (n = 53), near-sympatric yellowhammers (n = 15), sympatric yellowhammers (n =67), allopatric pine buntings (n = 42), near-sympatric pine buntings (n = 18), sympatric pine buntings (n = 52) and hybrids (n = 74). **A)** Whole-genome principal components analysis of allopatric yellowhammers (yellow), near-sympatric yellowhammers (light orange), sympatric yellowhammers (red-orange), allopatric pine buntings (brown), near-sympatric pine buntings (taupe), sympatric pine buntings (peach) and hybrids (green). PC1 explains 1.4% of the variation among individuals and PC2 explains 0.9% of the variation among individuals. Information from 374,780 SNPs was included in this analysis. **B)** Ancestry proportions of allopatric yellowhammers (Allo- C), near-sympatric yellowhammers (N.Sym- C), sympatric yellowhammers (Sym- C), allopatric pine buntings (Allo- L), near-sympatric pine buntings (N.Sym- L), sympatric pine buntings (Sym- L) and hybrids (Hybrid) as predicted by an Admixture model with K=2. Information from 417,164 SNPs were included in this analysis.



**Figure 3.2.** Genetic differentiation across the Z chromosome among allopatric yellowhammers (n = 53), near-sympatric yellowhammers (n = 15), sympatric yellowhammers (n = 67), allopatric pine buntings (n = 42), near-sympatric pine buntings (n = 18), sympatric pine buntings (n = 52) and hybrids (n = 74). **A)** Z chromosome principal components analysis of allopatric yellowhammers (yellow), near-sympatric yellowhammers (light orange), sympatric yellowhammers (red-orange), allopatric pine buntings (brown), near-sympatric pine buntings (taupe), sympatric pine buntings (peach) and hybrids (green). PC1 explains 8.1% of the variation among individuals and PC2 explains 1.7% of the variation among individuals. Information from 11,147 SNPs was included in this analysis. Numbered boxes are used to designate each of 6 clusters within PC space. **B)** Genotype-by individuals plot of a subset of females (n = 32). **C)** Genotype-by individuals plot of a subset of males (n = 100) from the chromosome Z PCA. SNPs with  $F_{ST}$  greater or equal to 0.7 in comparisons of allopatric populations were included in this analysis. Boxes filled in with one colour indicate homozygosity at a locus and boxes split into different coloured triangles indicate heterozygosity. Light purple indicates alleles with putative pine bunting ancestry and dark purple indicates alleles with putative yellowhammer ancestry. Numbers along left side correlate to numbered clusters within the chromosome Z PCA.

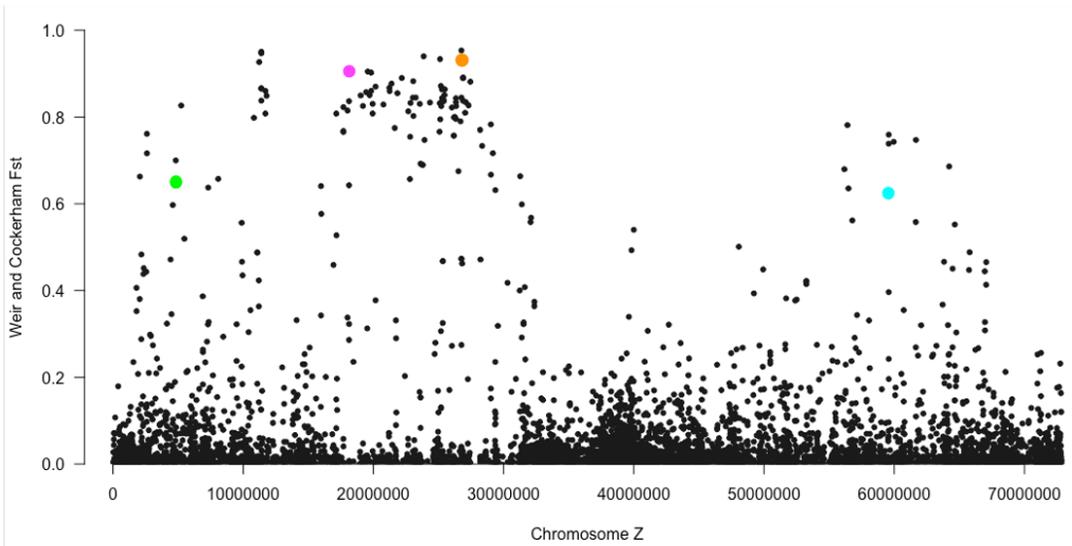


**Figure 3.3.** Triangle plots of interclass heterozygosity ( $H$ ) versus ancestry index ( $S$ ) for allopatric yellowhammers ( $n = 53$ ; yellow), near-sympatric yellowhammers ( $n = 15$ ; light orange), sympatric yellowhammers ( $n = 67$ ; red-orange), allopatric pine buntings ( $n = 42$ ; brown), near-sympatric pine buntings ( $n = 18$ ; taupe), sympatric pine buntings ( $n = 52$ ; peach) and hybrids ( $n = 74$ ; green). **A)** Analysis including 145 SNPs that possess  $F_{ST}$  values greater than or equal to 0.6 when comparing allopatric yellowhammer and allopatric pine bunting samples. **B)** Analysis including 10 unlinked SNPs possessing  $F_{ST}$  values greater than or equal to 0.5 when comparing allopatric yellowhammer and allopatric pine bunting samples.

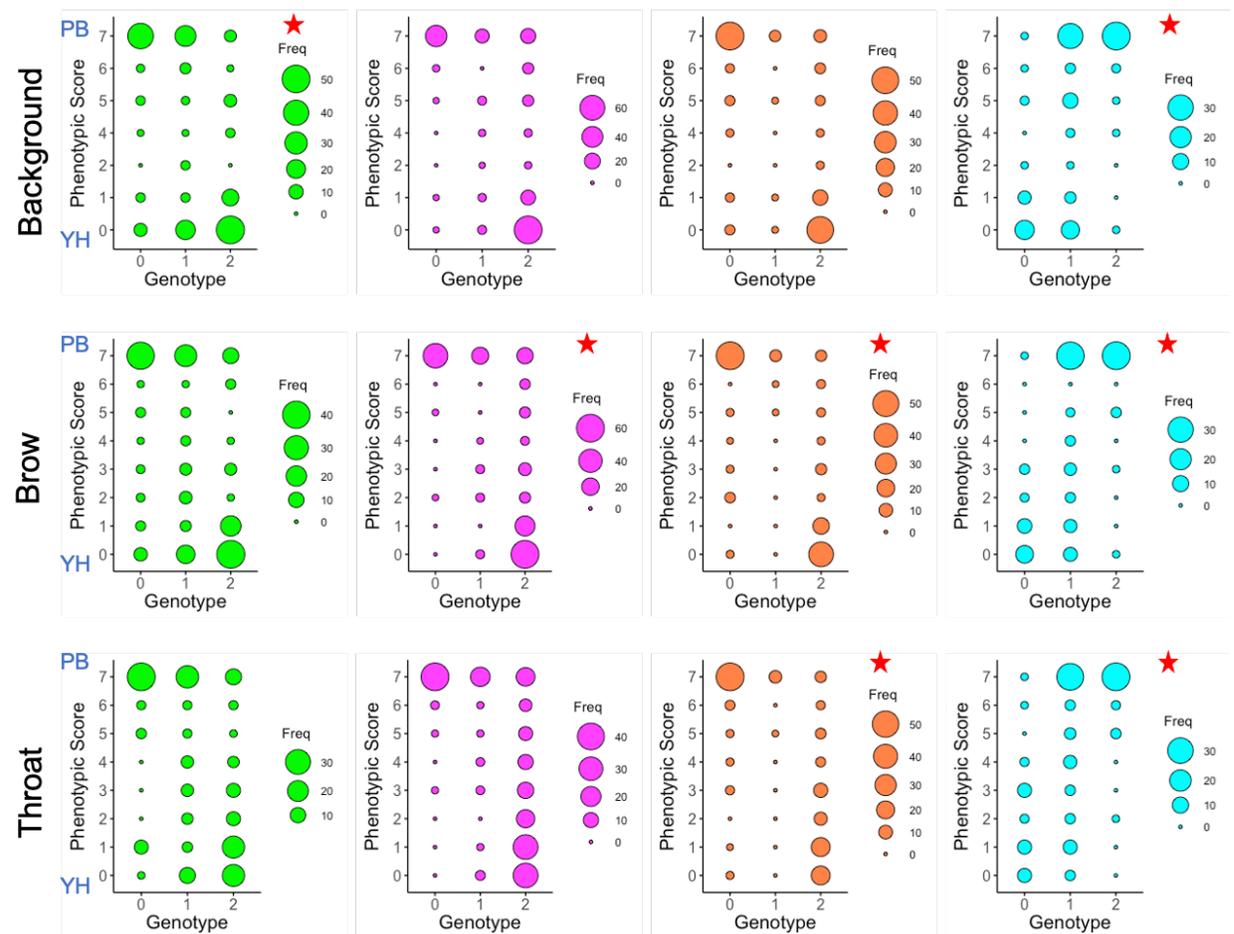


**Figure 3.4.** Associations between genome-wide SNPs and phenotypic variation in three plumage traits within the yellowhammer and pine bunting system. P-values for each SNP were determined using a likelihood ratio test calculated using the GEMMA program. Red lines indicate Bonferroni corrected significance thresholds. **A)** Associations between 220,220 genome-wide SNPs and variation in the background plumage colour of parental and hybrid individuals. **B)** Associations between 220,124 genome-wide SNPs and variation in the amount of chestnut plumage at the brow of parental and hybrid individuals. **C)** Associations between 220,307 genome-wide SNPs and variation in the amount of chestnut plumage at the throat of parental and hybrid individuals.

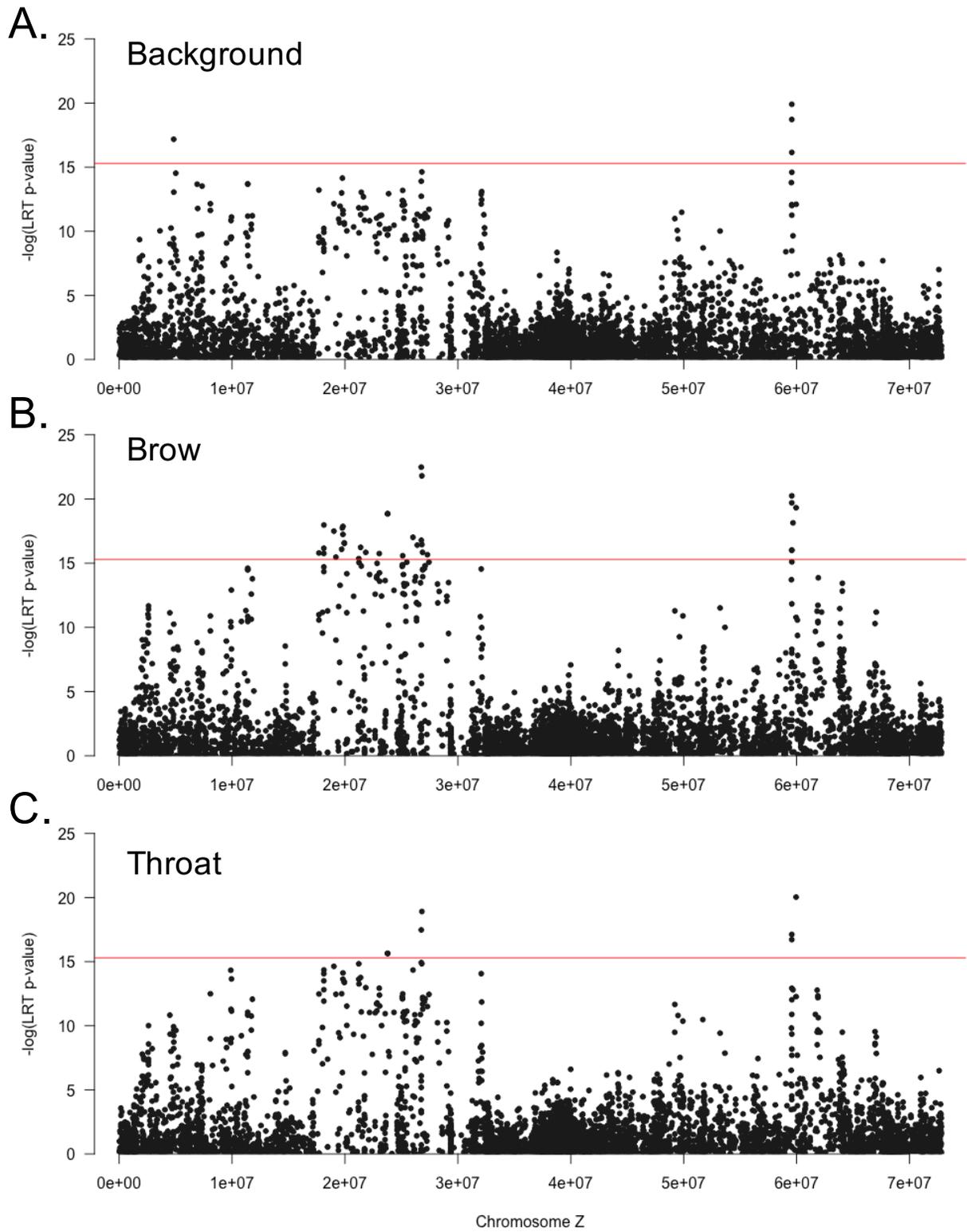
A.



B.



**Figure 3.5.** Investigation of potential dominance interactions between the alleles of SNPs that are significantly associated with plumage variation in the yellowhammer and pine bunting system as identified using GEMMA. **A)** Relative differentiation ( $F_{ST}$ ) of 11,147 SNPs on the Z chromosome calculated when comparing allopatric yellowhammers ( $n = 53$ ) and allopatric pine buntings ( $n = 42$ ). Four SNPs are highlighted in this panel: Z.4835388 (green), Z.18131016 (magenta), Z.26812248 (orange) and Z.59571043 (blue). **B)** Balloon plots illustrating the number of individuals possessing a particular phenotypic score at one three plumage traits with a particular genotype at one of the four SNPs highlighted in panel A. The plumage traits considered are: the background plumage colour (“Background”), the amount of chestnut at the brow (“Brow”) and the amount of chestnut at the throat (“Throat”). The SNP identities are indicated with the same colours as in panel A: Z.4835388 (green), Z.18131016 (magenta), Z.26812248 (orange) and Z.59571043 (blue). Dot size in each balloon plot indicates the number of individuals possessing a specific phenotypic score-genotype combination. Red stars indicate whether each of the four highlighted SNPs is significantly associated with phenotypic variation at the plumage trait of interest as determined using GEMMA. Blue labels “YH” (Yellowhammer) and “PB” (Pine Bunting) indicate phenotypic scores more commonly associated with “pure” members of each taxa.



**Figure 3.6.** Associations between SNPs located on chromosome Z and phenotypic variation in three plumage traits within the yellowhammer and pine bunting system. P-values for each SNP were determined using a likelihood ratio test calculated using the GEMMA program. Red lines indicate Bonferroni corrected significance thresholds. **A)** Associations between 6502 chromosome Z SNPs and variation in the background plumage colour of parental and hybrid individuals. **B)** Associations between 6509 chromosome Z SNPs and variation in the amount of chestnut plumage at the brow of parental and hybrid individuals. **C)** Associations between 6487 chromosome Z SNPs and variation in the amount of chestnut plumage at the throat of parental and hybrid individuals.

## Chapter 4: Conclusions

During the process of speciation, reproductive barriers accumulate between diverging populations limiting their ability to interbreed and share genetic material. Accordingly, complete reproductive isolation and full species status is achieved when interbreeding ceases entirely (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). The production of reproductive barriers is driven by genetic differentiation particularly at genes that regulate traits and behaviours related to reproduction. To answer the question of how reproductive barriers begin to develop, evolutionary scientists propose that initial differentiation occurs as a result of evolutionary forces acting on taxa during a period of allopatric isolation. When populations meet again in a zone of secondary contact, the opportunity to interbreed tests the strength of any novel reproductive barriers acting against homogenizing gene flow. Investigation of these sympatric regions provides valuable insight on the genetic loci responsible for reproductive isolation as well as on the delicate balance between population divergence and merging during speciation.

My work has demonstrated how the yellowhammer and pine bunting system provides an excellent opportunity to investigate the evolution of reproductive isolation between taxa following the trajectory of allopatric isolation and secondary contact described above. Separated during the Pleistocene glaciations, these putative species are thought to have diverged at a variety of different characteristics in isolation (Panov et al. 2003; Irwin et al. 2009; Rubtsov & Tarasov, 2017). Most significantly, yellowhammers and pine buntings differ greatly in their colouration patterns (Panov et al. 2003; Rubtsov & Tarasov, 2017)—traits that mediate reproductive interactions in many avian taxa (Price, 2008). Previous genomic work in this

system found moderate levels of nuclear genetic differentiation that are consistent with the amount of ecological, behavioural and visual divergence seen between groups (Irwin et al. 2009). Yet, in contrast to these observations, yellowhammers and pine buntings hybridize extensively within an area of secondary contact (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) and differ negligibly in their mitochondrial DNA (mtDNA; Irwin et al. 2009)—a marker commonly used to assess population divergence. This creates a conflicting picture of the state of reproductive barriers between taxa, raising the question of whether yellowhammers and pine are actually better considered members of the same species.

In this thesis, I surveyed phenotypically pure and hybrid individuals across the yellowhammer and pine bunting system and employed tens of thousands of genetic markers to assess genome-wide patterns of nuclear differentiation between taxa. Unsurprisingly, birds sampled throughout the yellowhammer and pine bunting allopatric zones separated into distinct genetic clusters along species boundaries. Nevertheless, average genome-wide differentiation was much lower than expected given the level of phenotypic divergence between taxa (Panov et al. 2003; Rubtsov & Tarasov, 2017), the amount of differentiation previously observed using a smaller number of genetic loci (Irwin et al. 2009) and the amount of differentiation noted among comparable avian sister species (Ellegren et al. 2012; Irwin et al. 2018). The disagreement between estimates of genetic divergence in previous research ( $F_{ST} = 0.078$  based on allele frequency;  $F_{ST} = 0.140$  based on AFLP band frequency; Irwin et al. 2009) and in this thesis ( $F_{ST} = 0.0232$ ) may be driven by the highly heterogenous nature of genetic differentiation between yellowhammers and pine buntings. In previous genomic work, only a small number of nuclear markers were considered when estimating genetic divergence between allopatric taxa. It is

possible that this subset of markers included a disproportionate number of SNPs within differentiation peaks such that these differentiation estimates were biased towards higher values compared to those produced from my genome-wide dataset. The low differentiation seen between yellowhammers and pine bunting may be partially driven by adaptive introgression or gene flow between taxa. In such a scenario, adaptive and neutral alleles would flow between sympatric populations as a result of hybridization and then be passed outward into allopatric populations leading to a homogenization of nuclear genomes. Adaptive introgression has been proposed previously in this system to explain the extremely low levels of differentiation seen between the mitochondrial genomes of these groups (Irwin et al. 2009).

Despite low average genetic divergence, narrow peaks in differentiation were seen between allopatric populations of yellowhammers and pine buntings across various chromosomes. In particular, there were several large differentiation peaks located on the Z chromosome. Previous evolutionary research has consistently noted higher levels of genetic differentiation over shorter time scales on sex chromosomes compared to autosomes (Thorton & Long, 2002; Borge et al. 2005; Lu & Wu, 2005; Harr, 2006; Ruegg et al. 2014; Sackton et al. 2014). These patterns are commonly attributed to differing selective pressures on sex chromosomes as part of the “faster X/Z effect” (reviewed in Meisel & Connallon, 2013; reviewed in Irwin, 2018) and to sex chromosomes’ lower effective populations sizes (Mank et al. 2010; reviewed in Irwin, 2018). Due to a variety of mechanisms, researchers postulate that sex chromosomes may play a disproportionately large role in the speciation process due to their higher rates of genetic differentiation (Sæther et al. 2007; Presgraves, 2008). In line with this idea, I found a tight association between the largest differentiation peak on the Z chromosome

and colouration differences between taxa. These colouration differences are considered one of the most likely candidates for a reproductive barrier in the yellowhammer and pine bunting system (Panov et al. 2003; Rubtsov & Tarasov, 2017).

Across the tree of life, divergent colouration phenotypes have been shown to mediate interbreeding and reproductive isolation between interacting species (e.g. Saetre et al. 1997; Lukhtanov et al. 2005; Seehausen et al. 2008; Uy et al. 2009; Seddon et al. 2013). This concept has been studied extensively in avian species due to the astounding amount of plumage diversity observed within birds. There are several ways by which plumage divergence can act as a reproductive barrier between avian taxa. Most significantly, these phenotypic differences can function as important visual signals that allow females to differentiate between con- and heterospecific males during mate choice (West-Eberhard, 1983; Edwards et al. 2005; Price, 2008). In this way, divergence of male plumage phenotypes in concert with female preferences for these phenotypes can create a strong prezygotic barrier that prevents interbreeding between speciating populations. As well, if hybrid offspring possess intermediate plumage phenotypes that females do not recognize, plumage divergence may also act as a postzygotic barrier by limiting hybrid reproductive fitness (Bridle et al. 2006; Irwin, 2020). Within the yellowhammer and pine bunting system, there is both a strong difference between the plumage patterns of heterospecific males and a variety of intermediate hybrid phenotypes (Panov et al. 2003; Rubtsov & Tarasov, 2017)—ideal conditions for plumage to act as a reproductive barrier between groups. However, unexpectedly, genomic patterns within the sympatric zone did not support the presence of such reproductive isolation.

Within their sympatric zone, yellowhammers and pine buntings are known to hybridize extensively (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) and I found genetic evidence for these interactions in a breakdown of the distinct genetic clusters observed between allopatric populations. However, levels of hybridization and gene flow within this region were also much greater than expected given the phenotypic composition of individuals within my dataset. Surprisingly, a large number of phenotypically pure individuals—particularly pine buntings—showed considerable admixture from their heterospecific counterparts. This extensive amount of hybridization within the sympatric zone suggests limited prezygotic barriers between yellowhammers and pine buntings despite their plumage differences. As well, evidence of hybrid viability (Lohrl, 1967 cited in Panov et al. 2003) and of high levels of late generation hybrids and backcrosses similarly supports negligible postzygotic barriers. Altogether, these patterns imply that reproductive isolation is extremely limited if present at all between yellowhammers and pine buntings.

Despite the apparent lack of reproductive barriers and extensive hybridization seen between yellowhammers and pine buntings, parental plumage phenotypes are maintained at relatively high levels within the sympatric zone. In fact, many phenotypically pure individuals in my dataset possessed highly admixed genomes—often comparable to the genomes of phenotypic hybrids. This poses the question as to what evolutionary force or genomic structure besides a reproductive barrier is preventing the breakdown of parental plumage patterns in the wake of gene flow. My results demonstrated a close association between plumage variation in this system and the large differentiation peak located on chromosome Z. Further genetic analysis revealed that loci within this differentiation peak were tightly linked in a pattern that is consistent with

recombination suppression. Reductions in recombination are associated with different kinds of genomic structures, but are often attributed to chromosomal inversion polymorphisms that are segregating within a system (reviewed in Smukowski & Noor, 2011).

Chromosomal inversions have recently gained traction as important drivers of speciation (e.g. Lowry & Willis, 2010; Ayala et al. 2013; Todesco et al. 2020). These conclusions are based on a chromosomal inversion's ability to suppress recombination and maintain associations between loci responsible for reproductive isolation between populations (Noor et al. 2001; Rieseberg, 2001). Stated another way, a chromosomal inversion prevents the separation of co-evolved alleles responsible for reproductive barriers during interbreeding provided that each taxon is fixed for a different version of the inversion. In the case of a barrier where a divergent plumage characteristic is used as a mate recognition signal, a chromosomal inversion can maintain an association between an allele that produces a specific phenotype and an allele that regulates female preference for that phenotype in spite of any hybridization between divergent taxa (Trickett & Butlin, 1994). Thus, a chromosomal inversion can allow for the retention of an existing reproductive barrier and can facilitate the continued divergence and speciation of taxa that have not yet achieved complete reproductive isolation.

As described earlier, I found that both pre- and postzygotic barriers are limited if they exist at all between yellowhammers and pine buntings. As such, the putative chromosomal inversion on the Z chromosome of these taxa appears to have little impact on reproductive isolation despite its close association with plumage differentiation and the fact that it roughly segregates along species boundaries. Most likely, the chromosomal inversion evolved during allopatric isolation with different forms of the inversion becoming fixed in each of the separated

populations. This inversion captured the loci responsible for plumage colouration in the system, but females may not have evolved a particular preference for the local male phenotype and, instead, simply mated with whoever was available. Thus, when phenotypically divergent yellowhammers and pine buntings met in a zone of secondary contact, females possessed little to no discrimination based on male phenotype and a reproductive barrier associated with plumage traits was unable to develop between taxa. Nevertheless, because the loci responsible for plumage differences were less susceptible to homogenization within alternative forms of an inversion, parental phenotypes were and are maintained at high levels within the sympatric zone in spite of the presumed non-discriminatory mating. Genomic patterns further suggest that the pine bunting form of the inversion and its associated plumage patterns may be dominant over the yellowhammer form of the inversion. This relationship is likely driving my observations from the sympatric zone where a large number phenotypically pure pine buntings possessed highly admixed genotypes.

Although the scenario explained above is consistent with genomic and demographic patterns within the yellowhammer and pine bunting system, there are several limitations to my conclusion that plumage variation is regulated by loci within a chromosomal inversion. First and foremost, I am unable to definitively confirm from my genetic data the presence of an inversion on chromosome Z. Patterns of heterozygosity across this chromosome imply an area of high linkage and low recombination that is associated with plumage phenotypes, but there are other ways by which recombination can be suppressed over a genomic region besides a chromosomal inversion that must be considered (reviewed in Smukowski & Noor, 2011). Second, I am also unable to discern exactly which loci and associated genes may be responsible for colouration

differences between yellowhammers and pine buntings. Due to tight linkage between markers, many of the loci within the putative inversion were significantly associated with phenotypic variation making it difficult to pinpoint the single or multiple causal SNPs. Additionally, because I used reduced representation sequencing which only captures a small portion of the nuclear genome (in this case, roughly 1.5%), it is possible that important regulatory loci within the inversion were not included in my genetic dataset for analysis. Nevertheless, despite these constraints, I identified GRAMD3 (GRAM domain containing 3) as a likely candidate gene for controlling melanin-based facial markings within this system. A SNP from this gene was highly associated with both brow and throat plumage variation—traits that are moderately correlated in the yellowhammer and pine bunting system (Panov et al. 2003)—and was also an  $F_{ST}$  outlier when comparing allopatric populations. As well, GRAMD3 has previously been connected to melanin-based pigmentation in its association with the *Id* locus in chickens (Xu et al. 2017). One further limitation to my conclusion is that the putative chromosomal inversion was not the only region of the genome correlated with plumage differences between yellowhammers and pine buntings. Additional loci outside of the inversion and on other chromosomes were significantly associated with phenotypic variation in this system. It is unclear how these loci might contribute to plumage phenotypes in conjunction with the dominance interactions proposed between different forms of the inversion. To address all these limitations, whole-genome sequencing is needed to directly characterize the area of low recombination seen on chromosome Z and to perform a functional analysis that will pin down the genetic mechanisms controlling plumage colouration between yellowhammers and pine buntings.

In the absence of barriers that permit species discrimination during mate choice, another common source of reproductive isolation is genetic incompatibility. Genetic incompatibilities develop as a result of genetic differentiation between taxa and manifest as fitness declines in hybrid offspring who possess a mixture of divergent, maladapted alleles (reviewed in Coyne & Orr, 2004; reviewed in Price, 2008). In the yellowhammer and pine bunting system, I find relatively low levels of genetic differentiation and an apparent lack of reproductive barriers suggesting that genetic incompatibilities between groups are minimal. Without such barriers, hybridization can proceed relatively unheeded between taxa allowing for homogenization in their nuclear genomes. With this in mind, I attribute patterns of low differentiation between yellowhammers and pine buntings to introgression of adaptive and neutral alleles as a result of interbreeding within the sympatric zone. Yet, this hypothesis can be expanded further in the context of general drivers of reproductive barrier loss if I consider mitonuclear interactions and potential mitonuclear co-introgression between divergent yellowhammers and pine buntings.

As part of mitonuclear research, scientists investigate how interactions between the mitochondrial and nuclear genomes regulate mitochondrial function within an organism (Hill, 2019). In most species, the mitochondrial genome has become greatly reduced meaning that effective operation of mitochondrial processes, such as oxidative phosphorylation, requires proteins encoded by various nuclear genes commonly referred to as “mitonuclear genes” (Calvo & Mootha, 2010; Lotz et al. 2014). Products of mitochondrial and mitonuclear genes function closely together within the mitochondria creating the conditions necessary for co-evolution between the mitochondrial and nuclear genomes. As a result, rapid differentiation of the mitochondrial genome between diverging populations is thought to drive similar differentiation

at mitonuclear genes (Gershoni et al. 2009; Burton & Barretto, 2012; Hill, 2019). This mitonuclear co-divergence can create strong postzygotic barriers between taxa and research has shown that mixing divergent mitochondrial and mitonuclear alleles can greatly affect mitochondrial efficiency in hybrid organisms (Kenyon & Moraes, 1997; Barrientos et al. 1998; Ellison & Burton, 2006; Burton et al. 2013). Based on all these factors and on the ubiquity of mitochondria across the tree of life, evolutionary researchers suggest that mitonuclear interactions may be an important driver of speciation among taxa (Gershoni et al. 2009; Burton & Barretto, 2012; Hill, 2019). Nevertheless, as has been noted in a large number of systems (e.g. Malmos et al. 2001; Ludwig et al. 2003; McGuire et al. 2007; Mila et al. 2011; Leavitt et al. 2017; Mastrantonio et al. 2019; Çoraman et al. 2020), mitochondrial DNA may introgress and sweep between hybridizing taxa which can drive comparable introgression of mitonuclear alleles. This co-introgression can remove any mitonuclear incompatibilities between taxa and greatly reduce the strength of postzygotic isolation between them. As such, mitonuclear interactions have the potential to act as a driver of species merging as well as speciation.

Past genomic work in the yellowhammer and pine bunting system noted negligible mtDNA differentiation between these taxa that is consistent with mtDNA introgression (Irwin et al. 2009). To address the possibility of mitonuclear co-introgression, I compared genomic signals of selection at mitonuclear genes to the rest of the genome looking for evidence of preferential selective sweeps. My results provided moderate support for some mitonuclear alleles sweeping in the direction of pine buntings into yellowhammers which could suggest past co-introgression between groups. Mitonuclear co-introgression has the potential to drive hybridization between yellowhammers and pine buntings by removing a large number of mitonuclear incompatibilities

over a relatively short period of time. Because mtDNA is inherited as a single, linked unit, introgression of mtDNA could select for co-introgression of related mitonuclear genes all at once—removing any existing mitonuclear incompatibilities together rather than one at a time. Thus, a reduction in postzygotic isolation between taxa could be both large and rapid driving a sudden influx in hybridization. In the wake of this increased rate of interbreeding, the genomes of diverged taxa would quickly homogenize removing any remaining genetic incompatibilities and preventing the production of any new ones. These conditions could lead to even higher levels of hybridization in a positive feedback loop that eventually culminates in complete population merging. Further research concerning the role of mitonuclear co-introgression in reproductive barrier loss is greatly needed to confirm whether this process is important in driving hybridization within the yellowhammer and pine bunting system and others.

In this thesis, I found some support for introgression of mitonuclear alleles from pine buntings into yellowhammers that is consistent with mitonuclear theory. However, these results must be considered with caution as there are several limitations to the methodology I employed to examine this process. First, as mentioned in my analysis of a chromosomal inversion on chromosome Z, I performed reduced representation sequencing when obtaining my genomic data meaning that I only possess information for a small portion of the nuclear genome. As such, the mitonuclear gene locations used when estimating levels of selection were broad and inexact meaning that my results could reflect genomic trends unrelated to mitonuclear co-introgression. Future work employing whole-genome sequencing would allow researchers to narrow down mitonuclear gene locations and provide definitive support for or against introgression at these regions in response to mtDNA introgression. Another limitation of this analysis is that, due to

resource constraints, I was only able to investigate signals of selection at a subset of the estimated 1500 mitonuclear genes in the nuclear genome (Calvo & Mootha, 2010; Lotz et al. 2014). I selected this subset of genes based on their protein products which all interacted directly with proteins or RNA encoded by mtDNA meaning that any signals of mitonuclear co-introgression would be particularly strong at these genomic regions. The exclusion of a large portion of the existing mitonuclear genes could imply that my results do not reflect overall trends seen across mitonuclear genes, but rather describe only specific trends at those few genes considered. Finally, my mitonuclear analysis is further limited in that it did not investigate the direction of mtDNA introgression between yellowhammers and pine buntings. My genomic data supports mitonuclear gene introgression from pine buntings into yellowhammers. However, without knowing the direction of mtDNA introgression in this system, I cannot say for certain whether mitonuclear introgression followed a similar direction—in support of mitonuclear co-introgression—or occurred in the opposite direction. Information of this kind would provide more definitive evidence for mitonuclear co-introgression between yellowhammers and pine buntings and would potentially negate some of the other restrictions to this study.

The yellowhammer and pine bunting were granted species designation based on their highly divergent appearances (Panov et al. 2003; Irwin et al. 2009; Rubtsov & Tarasov, 2017). However, the extensive hybridization seen between these taxa within their sympatric zone (Panov et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017) questions the legitimacy of distinct species designations. In this thesis, I evaluated genetic variation and admixture across the yellowhammer and pine bunting system and provided evidence that reproductive barriers are weak or even non-existent between taxa. This lack of reproductive isolation may be partially as a

result of mitonuclear co-introgression driving gene flow and genetic homogenization between taxa. However, it is also possible that strong, or even moderate, reproductive isolation never evolved between these two groups in the first place. Without strong reproductive barriers, it can be assumed that the yellowhammer and pine bunting will hybridize indiscriminately perhaps to the point of merging back into a single panmictic population. Such a future is supported by long-term surveys of the yellowhammer and pine bunting sympatric zone that suggest this region has expanded 1000 km to the east in the last one hundred years and 350 km to the west in the last twenty-five years (reviewed in Panov et al. 2003; reviewed in Rubtsov and Tarasov, 2017). However, the presence of a putative chromosomal inversion polymorphism that regulates plumage variation within this system slightly complicates the narrative of this story. Protected from recombination and genetic homogenization, it is predicted that the different versions of the inversion and their associated phenotypes could be maintained during and following population merging. As such, within the projected panmictic population, I expect to see a retention of the highly divergent, ancestral yellowhammer and pine bunting phenotypes as a freely segregating plumage polymorphism. In support of this conclusion, I saw several individuals within my sympatric dataset who possessed the ancestral phenotype of one taxon in conjunction with a highly admixed genome. Such a dramatic polymorphism may persist indefinitely if gene flow across the system remains unimpeded. However, the preservation of such genetic and phenotypic variation could also facilitate the evolution of reproductive isolation and, even, speciation if specific conditions such as another cycle of allopatric isolation were to arise in the future.

The results and conclusions discussed in this thesis are meaningful to many different aspects of speciation and evolutionary research. Beginning with the narrower implications of this

study, the classification of a potential regulator of plumage variation between yellowhammers and pine buntings is significant to understanding speciation within Emberizidae. Members of this Family are all unique in their appearances, but tend to diverge at consistent plumage traits—particularly facial markings. Keeping in mind that plumage variation is often an important reproductive barrier in avian systems (e.g. Saetre et al. 1997; Uy et al. 2009; Seddon et al. 2013), the identification of a putative chromosomal inversion highly associated with variation in facial plumage provides researchers with a strong candidate region that may be important to past and future diversification within Emberizidae. Next, my investigation of mitonuclear co-introgression highlights the importance of mitonuclear interactions in not only driving speciation, but also in countering it as well. Normally, mitonuclear interactions are discussed in the context of how mitonuclear co-divergence produces strong postzygotic barriers between taxa (Gershoni et al. 2009; Burton & Barretto, 2012; Hill, 2019). However, mitonuclear co-introgression has the potential to reduce postzygotic isolation at a particularly rapid rate and ultimately lead to population merging. This concept warrants future study especially considering the high frequency of mtDNA introgression proposed between closely related taxa (Malmos et al. 2001; Ludwig et al. 2003; McGuire et al. 2007; Mila et al. 2011; Leavitt et al. 2017; Mastrantonio et al. 2019; Çoraman et al. 2020).

Thinking more broadly, my research in the yellowhammer and pine bunting system also has important implications in understanding the origins of variation within species. My study presents an unusual trajectory for population merging where, rather than fusing into one homogenized group, I predict that highly divergent plumage phenotypes will be retained within a panmictic population of individuals. Such phenotypic variation within an interbreeding

population has only been observed within a select number of systems including the white-throated sparrows (*Zonotrichia albicollis*; Tuttle, 2003), the Gouldian finches (*Chloebia gouldiae*; Southern 1945; Brush & Seifried, 1968) and the ruffs (*Calidris pugnax*; Hogan-Warburg, 1966; Jukema & Piersma, 2006). Nevertheless, some amount of variation—whether it be phenotypic, behavioural, physiological or genetic—is common within species and is often attributed to forces such as mutation, or genetic drift (reviewed in Coyne et al., 2004; reviewed in Price, 2008). My results hint that variation within a species may also be the product of past allopatric isolation followed by population merging. In this scenario, divergence in isolation produces some variation between separated populations, but not enough to create strong reproductive barriers. When the somewhat differentiated populations meet again during secondary contact, they fuse back together into one large panmictic population due to hybridization. Yet, in spite of this fusion, the variation produced during allopatric isolation is retained to a limited extent within the panmictic population increasing the overall state of phenotypic and genetic variation within the system from what it was prior to allopatric isolation. Such a process is projected to introduce adaptive rather than random variation into a species as some divergence would have occurred adaptively in isolation. As well, population merging also has the potential to increase species variation to a greater extent and over a much shorter time scale than either mutation or drift.

Population divergence in allopatry followed by hybridization and population fusion during secondary contact greatly increases variation within a species which will, by extension, rapidly inflate the evolvability of the reformed taxon potentially allowing it to adapt to a wider range of environmental conditions. This capability can safeguard the species from future

extinction and may even pave the way for future speciation if evolutionary pressures change. The idea that hybridization may act as a creative rather than a destructive force that can drive adaptation and speciation has been explored previously in several systems including *Helianthus* sunflowers (Mitchell et al. 2019), canids (Wang et al. 2020) and *Littorina* snails (Morales et al. 2019). However, there is limited research that considers how extended hybridization leading to population merging may also act as a creative evolutionary force in the context of a single taxonomic unit rather than two. Further work investigating the effects of this process on species variation across systems and its impacts on the evolutionary trajectory of taxa is greatly needed to develop a more holistic understanding of the speciation process.

## Bibliography

- Aboim, M. A., Mavárez, J., Bernatchez, L., & Coelho, M. M. (2010). Introgressive hybridization between two Iberian endemic cyprinid fish: A comparison between two independent hybrid zones. *Journal of Evolutionary Biology*, 23(4), 817-828. doi:10.1111/j.1420-9101.2010.01953.x
- Alcaide, M., Scordato, E. S. C., Price, T. D., & Irwin, D. E. (2014). Genomic divergence in a ring species complex. *Nature*, 511(7507), 83-85. doi:10.1038/nature13285
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664. doi:10.1101/gr.094052.109
- Alström, P., Olsson, U., Lei, F., Wang, H., Gao, W., & Sundberg, P. (2008). Phylogeny and classification of the old world *Emberizini* (*Aves*, *Passeriformes*). *Molecular Phylogenetics and Evolution*, 47(3), 960-973. doi:10.1016/j.ympev.2007.12.007
- Andrade, P., Pinho, C., de Lanuza, G. P., Afonso, S., Brejcha, J., Rubin, C., Wallerman, O., Pereira, P., Sabatino, S. J., Bellati, A., Pellitteri-Rosa, D., Bosakova, Z., Bunikis, I., Carretero, M. A., Feiner, N., Marsik, P., Paupério, F., Salvi, D., Soler, L., ... Carneiro, M. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences*, 116(12), 5633-5642. doi:10.1073/pnas.1820320116
- Attie, A. D., Hamon, Y., Brooks-Wilson, A. R., Gray-Keller, M. P., MacDonald, M. L. E., Rigot, V., Tebon, A., Zhang, L-H., Mulligan, J. D., Singaraja, R. R., Bitgood, J. J., Cook, M. E., Kastelein, J. J. P., Chimini, G., & Hayden, M. R. (2002). Identification and functional analysis of a naturally occurring E89K mutation in the ABCA1 gene of the WHAM chicken. *Journal of Lipid Research*, 43(10), 1610-1617. doi:10.1194/jlr.M200223-JLR200
- Ayala, D., Guerrero, R. F., & Kirkpatrick, M. (2013). Reproductive isolation and local adaptation quantified for a chromosome inversion in a malaria mosquito. *Evolution*, 67(4), 946-958. doi:10.1111/j.1558-5646.2012.01836.x

- Balakrishnan, C. N., & Sorenson, M. D. (2006). Song discrimination suggests premating isolation among sympatric indigobird species and host races. *Behavioral Ecology*, *17*(3), 473-478. doi:10.1093/beheco/arj052
- Ballard, J. W. O., & Melvin, R. G. (2010). Linking the mitochondrial genotype to the organismal phenotype. *Molecular Ecology*, *19*(8), 1523-1539. doi:10.1111/j.1365-294X.2010.04594.x
- Bardhan, A., & Sharma, T. (2000). Meiosis and speciation: a study in a speciating *Mus terricolor* complex. *Journal of Genetics*, *79*(3), 105-111. doi:10.1007/BF02715858
- Barrientos, A., Kenyon, L., & Moraes, C. T. (1998). Human xenomitochondrial cybrids. *The Journal of Biological Chemistry*, *273*(23), 14210-14217. doi:10.1074/jbc.273.23.14210
- Barton, N. H., & Hewitt, G. M. (1985). Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, *16*(1), 113-148. doi:10.1146/annurev.es.16.110185.000553
- Barton, N. H., & Hewitt, G. M. (1989). Adaptation, speciation and hybrid zones. *Nature*, *341*(6242), 497-503. doi:10.1038/341497a0
- Bateson, W. 1909. Heredity and variation in modern lights. In A. C. Seward (Eds.) *Darwin and Modern Science* (pp. 85-101). Cambridge University Press.
- Baute, G. J., Owens, G. L., Bock, D. G., & Rieseberg, L. H. (2016). Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *American Journal of Botany*, *103*(12), 2170-2177. doi:10.3732/ajb.1600295
- Beck, E. A., Thompson, A. C., Sharbrough, J., Brud, E., & Llopart, A. (2015). Gene flow between *Drosophila yakuba* and *Drosophila santomea* in subunit V of cytochrome c oxidase: A potential case of cytonuclear cointrogression. *Evolution*, *69*(8), 1973-1986. doi:10.1111/evo.12718

- Blair, W. F. (1955). Mating call and stage of speciation in the *Microhyla olivacea-M. carolinensis* complex. *Evolution*, 9(4), 469-480. doi:10.1111/j.1558-5646.1955.tb01556.x
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120. doi:10.1093/bioinformatics/btu170
- Borge, T., Webster, M. T., Andersson, G., & Saetre, G. (2005). Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. *Genetics*, 171(4), 1861-1873. doi:10.1534/genetics.105.045120
- Brelsford, A., Toews, D. P. L., & Irwin, D. E. (2017). Admixture mapping in a hybrid zone reveals loci associated with avian feather coloration. *Proceedings of the Royal Society B: Biological Sciences*, 284(1866), 20171106. <https://doi.org/10.1098/rspb.2017.1106>
- Bridle, J. R., Saldamando, C. I., Koning, W., & Butlin, R. K. (2006). Assortative preferences and discrimination by females against hybrid male song in the grasshoppers *Chorthippus brunneus* and *Chorthippus jacobsi* (Orthoptera: Acrididae). *Journal of Evolutionary Biology*, 19(4), 1248-1256. doi:10.1111/j.1420-9101.2006.01080.x
- Bronson, C. L., Grubb, T. C., Sattler, G. D., & Braun, M. J. (2003). Mate preference: A possible causal mechanism for a moving hybrid zone. *Animal Behaviour*, 65(3), 489-500. doi:10.1006/anbe.2003.2103
- Brush, A. H., & Seifried, H. (1968). Pigmentation and feather structure in genetic variants of the Gouldian finch, *Poephila gouldiae*. *The Auk*, 85(3), 416-430. doi:10.2307/4083290
- Bryson, R. W., Adrian Nieto-Montes de Oca, Jaeger, J. R., & Riddle, B. R. (2010). Elucidation of cryptic diversity in a widespread Nearctic treefrog reveals episodes of mitochondrial gene capture as frogs diversified across a dynamic landscape. *Evolution*, 64(8), 2315-2330. doi:10.1111/j.1558-5646.2010.01014.x
- Burton, R. S., & Barreto, F. S. (2012). A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Molecular Ecology*, 21(20), 4942-4957. doi:10.1111/mec.12006

- Burton, R. S., Pereira, R. J., & Barreto, F. S. (2013). Cytonuclear genomic interactions and hybrid breakdown. *Annual Review of Ecology, Evolution, and Systematics*, 44(1), 281-302. doi:10.1146/annurev-ecolsys-110512-135758
- Calvo, S. E., & Mootha, V. K. (2010). The mitochondrial proteome and human disease. *Annual Review of Genomics and Human Genetics*, 11(1), 25-44. doi:10.1146/annurev-genom-082509-141720
- Campbell, C. R., Poelstra, J. W., & Yoder, A. D. (2018). What is speciation genomics? The roles of ecology, gene flow, and genomic architecture in the formation of species. *Biological Journal of the Linnean Society*, 124(4), 561-583. doi:10.1093/biolinnea/bly063
- Campo, D., Lehmann, K., Fjeldsted, C., Souaiaia, T., Kao, J., & Nuzhdin, S. V. (2013). Whole-genome sequencing of two north american drosophila melanogaster populations reveals genetic differentiation and positive selection. *Molecular Ecology*, 22(20), 5084-5097. doi:10.1111/mec.12468
- Cerdá-Reverter, J. M., Haitina, T., Schiöth, H. B., & Peter, R. E. (2005). Gene structure of the goldfish agouti-signaling protein: A putative role in the dorsal-ventral pigment pattern of fish. *Endocrinology*, 146(3), 1597-1610. doi:10.1210/en.2004-1346
- Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience*, 4(1) doi:10.1186/s13742-015-0047-8
- Clevidence, B. A., & Bieri, J. G. (1993). Association of carotenoids with human plasma lipoproteins. *Methods in Enzymology*, 214, 33-46.
- Connor, W. E., Duell, P. B., Kean, R., & Wang, Y. (2007). The prime role of HDL to transport lutein into the retina: Evidence from HDL-deficient WHAM chicks having a mutant ABCA1 transporter. *Investigative Ophthalmology & Visual Science*, 48(9), 4226-4231. doi:10.1167/iovs.06-1275

- Çoraman, E., Dundarova, H., Dietz, C., & Mayer, F. (2020). Patterns of mtDNA introgression suggest population replacement in Palearctic whiskered bat species. *Royal Society Open Science*, 7(6), 191805-191805. doi:10.1098/rsos.191805
- Coyne, J. A., Meyers, W., Crittenden, A. P., & Sniegowski, P. (1993). The fertility effects of pericentric inversions in *Drosophila melanogaster*. *Genetics*, 134(2), 487-496.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.
- Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, 23(13), 3133-3157. doi:10.1111/mec.12796
- Cruz-Yepez, N., González, C., & Ornelas, J. F. (2020). Vocal recognition suggests premating isolation between lineages of a lekking hummingbird. *Behavioral Ecology*, 31(4), 1046-1053. doi:10.1093/beheco/araa050
- Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying the disparity among species. *Nature Reviews. Genetics*, 14(4), 262-274. doi:10.1038/nrg3425
- 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), 2156-2158. doi:10.1093/bioinformatics/btr330
- Danley, P. D., DeCarvalho, T. N., Fergus, D. J., & Shaw, K. L. (2007). Reproductive asynchrony and the divergence of Hawaiian crickets. *Ethology*, 113(12), 1125-1132. doi:10.1111/j.1439-0310.2007.01430.x
- Dobzhansky, T. 1937. *Genetics and the Origin of Species*. Columbia University Press.
- Dobzhansky, T. (1940). Speciation as a stage in evolutionary divergence. *The American Naturalist*, 74(753), 312-321. doi:10.1086/280899

- Edwards, S. V., Kingan, S. B., Calkins, J. D., Balakrishnan, C. N., Jennings, W. B., Swanson, W. J., & Sorenson, M. D. (2005). Speciation in birds: Genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences*, *102*(Suppl 1), 6550-6557. doi:10.1073/pnas.0501846102
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., Künstner, A., Mäkinen, H., Nadachowska-Brzyska, K., Qvarnström, A., Uebbing, S., & Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, *491*(7426), 756-760. doi:10.1038/nature11584
- Ellison, C. K., & Burton, R. S. (2006). Disruption of mitochondrial function in interpopulation hybrids of *Tigriopus californicus*. *Evolution*, *60*(7), 1382-1391. doi:10.1554/06-210.1
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PloS One*, *6*(5), e19379. doi:10.1371/journal.pone.0019379
- Ernster, L., & Schatz, G. (1981). Mitochondria: A historical review. *The Journal of Cell Biology*, *91*(3), 227s-255s. doi:10.1083/jcb.91.3.227s
- Fitzpatrick, B. M. (2012). Estimating ancestry and heterozygosity of hybrids using molecular markers. *BMC Evolutionary Biology*, *12*(1), 131-131. doi:10.1186/1471-2148-12-131
- Funk, W. C., Caminer, M., & Ron, S. R. (2012). High levels of cryptic species diversity uncovered in Amazonian frogs. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1734), 1806-1814.
- García-Gómez, E., Reverter, A., Whan, V., McWilliam, S. M., Arranz, J. J., Kijas, J., & International Sheep Genomics Consortium. (2011). Using regulatory and epistatic networks to extend the findings of a genome scan: Identifying the gene drivers of pigmentation in merino sheep. *PloS One*, *6*(6), e21158-e21158. doi:10.1371/journal.pone.0021158

- Gascuel, O. (1997). BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Molecular Biology and Evolution*, *14*(7), 685-695. doi:10.1093/oxfordjournals.molbev.a025808
- Geraldes, A., Askelson, K. K., Nikelski, E., Doyle, F. I., Harrower, W. L., Winker, K., & Irwin, D. E. (2019). Population genomic analyses reveal a highly differentiated and endangered genetic cluster of northern goshawks (*Accipiter gentilis laingi*) in Haida Gwaii. *Evolutionary Applications*, *12*(4), 757-772. doi:10.1111/eva.12754
- Gershoni, M., Templeton, A. R., & Mishmar, D. (2009). Mitochondrial bioenergetics as a major motive force of speciation. *Bioessays*, *31*(6), 642-650. doi:10.1002/bies.200800139
- Gompert, Z., Mandeville, E. G., & Buerkle, C. A. (2017). Analysis of population genomic data from hybrid zones. *Annual Review of Ecology, Evolution, and Systematics*, *48*(1), 207-229. doi:10.1146/annurev-ecolsys-110316-022652
- Grant, P. R., & Grant, B. R. (1997). Genetics and the origin of bird species. *Proceedings of the National Academy of Sciences*, *94*(15), 7768-7775. doi:10.1073/pnas.94.15.7768
- Grava, A., Grava, T., Didier, R., Lait, L. A., Dosso, J., Koran, E., Burg, T. M., Otter, K. A. (2012). Interspecific dominance relationships and hybridization between black-capped and mountain chickadees. *Behavioral Ecology*, *23*(3), 566-572. doi:10.1093/beheco/arr229
- Greber, B. J., & Ban, N. (2016). Structure and function of the mitochondrial ribosome. *Annual Review of Biochemistry*, *85*(1), 103-132. doi:10.1146/annurev-biochem-060815-014343
- Gross, J. B., Borowsky, R., & Tabin, C. J. (2009). A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. *PLoS Genetics*, *5*(1), e1000326-e1000326. doi:10.1371/journal.pgen.1000326
- Haldane, J. B. S. (1922). Sex ratio and unisexual sterility in hybrid animals. *Journal of Genetics*, *12*(2), 101-109. doi:10.1007/BF02983075

- Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research*, 16(6), 730-737. doi:10.1101/gr.5045006
- Harris, R. B., Alström, P., Ödeen, A., & Leaché, A. D. (2018). Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (*Motacilla*). *Molecular Phylogenetics and Evolution*, 120, 183-195. doi:10.1016/j.ympev.2017.11.020
- Hatfield, T., & Schluter, D. (1999). Ecological speciation in sticklebacks: Environment-dependent hybrid fitness. *Evolution*, 53(3), 866-873. doi:10.1111/j.1558-5646.1999.tb05380.x
- Haupaix, N., Curantz, C., Bailleul, R., Beck, S., Robic, A., & Manceau, M. (2018). The periodic coloration in birds forms through a prepattern of somite origin. *Science*, 361(6408), 1216. doi:10.1126/science.aar4777
- Hebert, P. D. N., Cywinska, A., Ball, S. L., & deWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society. Series B: Biological Sciences*, 270(1512), 313-321. doi:10.1098/rspb.2002.2218
- Hebert, P. D., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences*, 101(41), 14812-14817. doi:10.1073/pnas.0406166101
- Hejase, H. A., Salman-Minkov, A., Campagna, L., Hubisz, M. J., Lovette, I. J., Gronau, I., & Siepel, A. (2020). Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proceedings of the National Academy of Sciences*, 117(48), 30554-30565. doi:10.1073/pnas.2015987117
- Hewitt, G. M. (1988). Hybrid zones-natural laboratories for evolutionary studies. *Trends in Ecology & Evolution*, 3(7), 158-167. doi:10.1016/0169-5347(88)90033-x
- Hill, G. E. (2019). *Mitochondrial Ecology* (First ed.). Oxford University Press.

- Hoffmann, A. A., & Rieseberg, L. H. (2008). Revisiting the impact of inversions in evolution: From population genetic markers to drivers of adaptive shifts and speciation? *Annual Review of Ecology, Evolution, and Systematics*, 39(1), 21-42. doi:10.1146/annurev.ecolsys.39.110707.173532
- Hogan-Warburg, A. J. (1966). *Social behavior of the ruff, Philomachus pugnax (L.)* (Vol. 54). Brill Archive.
- Hubbard, J. K., Uy, J. A. C., Hauber, M. E., Hoekstra, H. E., & Safran, R. J. (2010). Vertebrate pigmentation: From underlying genes to adaptive function. *Trends in Genetics*, 26(5), 231-239. doi:10.1016/j.tig.2010.02.002
- International Barcode of Life Consortium. (2021) Illuminate biodiversity to save our living planet. <https://ibol.org/>
- International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, 432(7018), 695-716. doi:10.1038/nature03154
- Irwin, D. E. (2018). Sex chromosomes and speciation in birds and other ZW systems. *Molecular Ecology*, 27(19), 3831-3851. doi:10.1111/mec.14537
- Irwin, D. E. (2020). Assortative mating in hybrid zones is remarkably ineffective in promoting speciation. *The American Naturalist*, 195(6), E150.
- Irwin, D. E., Alcaide, M., Delmore, K. E., Irwin, J. H., & Owens, G. L. (2016). Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Molecular Ecology*, 25(18), 4488-4507. doi:10.1111/mec.13792
- Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., Grossen, C., Delmore, K. E., Alcaide, M., & Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, 27(23), 4839-4855. doi:10.1111/mec.14858

- Irwin, D. E., Rubtsov, A. S., & Panov, E. N. (2009). Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biological Journal of the Linnean Society*, 98(2), 422-438. doi:10.1111/j.1095-8312.2009.01282.x
- Ivanova, I., Arulanantham, S., Barr, K., Cepeda, M., Parkins, K., Hamilton, A., Johnson, D., Penuela, S., Hess, D. A., Ronald, J. A., & Dagnino, L. (2019). Targeting FER kinase inhibits melanoma growth and metastasis. *Cancers*, 11(3), 419. doi:10.3390/cancers11030419
- Jones, K. S., & Weisrock, D. W. (2018). Genomic data reject the hypothesis of sympatric ecological speciation in a clade of *Desmognathus* salamanders. *Evolution*, 72(11), 2378-2393. doi:10.1111/evo.13606
- Jukema, J., & Piersma, T. (2006). Permanent female mimics in a lekking shorebird. *Biology Letters*, 2(2), 161-164. doi:10.1098/rsbl.2005.0416
- Kassambara, A. (2017). *ggpubr: Publication Ready Plots*. Statistical Tools for High-Throughput Data Analysis. <http://www.sthda.com/english/articles/24-ggpubr-publication-ready-plots/>
- Kenyon, L., & Moraes, C. T. (1997). Expanding the functional human mitochondrial DNA database by the establishment of primate xenomitochondrial cybrids. *Proceedings of the National Academy of Sciences*, 94(17), 9131-9135. doi:10.1073/pnas.94.17.9131
- Kerr, K. C. R., Stoeckle, M. Y., Dove, C. J., Weigt, L. A., Francis, C. M., & Hebert, P. D. N. (2007). Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7(4), 535-543. doi:10.1111/j.1471-8286.2007.01670.x
- Kingsley, E. P., Manceau, M., Wiley, C. D., & Hoekstra, H. E. (2009). Melanism in *Peromyscus* is caused by independent mutations in agouti. *PloS One*, 4(7), e6435-e6435. doi:10.1371/journal.pone.0006435

- Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology*, 8(9), e1000501. doi:10.1371/journal.pbio.1000501
- Kirschel, A. N. G., Nwankwo, E. C., Pierce, D. K., Lukhele, S. M., Moysi, M., Ogolowa, B. O., Hayes, S. C., Monadjem, A., & Brelsford, A. (2020). CYP2J19 mediates carotenoid colour introgression across a natural avian hybrid zone. *Molecular Ecology*, 29(24), 4970-4984. doi.org/10.1111/mec.15691
- Lande, R. (1985). The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity*, 54(3), 323-332. doi:10.1038/hdy.1985.43
- Leavitt, D. H., Marion, A. B., Hollingsworth, B. D., & Reeder, T. W. (2017). Multilocus phylogeny of alligator lizards (*Elgaria, anguinae*): Testing mtDNA introgression as the source of discordant molecular phylogenetic hypotheses. *Molecular Phylogenetics and Evolution*, 110, 104-121. doi:10.1016/j.ympev.2017.02.010
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760. doi:10.1093/bioinformatics/btp324
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The sequence alignment map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079. doi:10.1093/bioinformatics/btp352
- Lohrl, H. (1967). Zur Verwandtschaft von Fichtenammer (*Emberiza leucocephalos*) und Goldammer (*Emberiza citrinella*). *Vogelwelt*, 88, 148-152.
- Lotz, C., Lin, A. J., Black, C. M., Zhang, J., Lau, E., Deng, N., Wang, Y., Zong, N. C., Choi, J. H., Xu, T., Liem, D. A., Korge, P., Weiss, J. N., Hermjakob, H., Yates, J. R., III., Apweiler, R., & Ping, P. (2014). Characterization, design, and function of the mitochondrial proteome: From organs to organisms. *Journal of Proteome Research*, 13(2), 433-446. doi:10.1021/pr400539j
- Lowry, D. B., & Willis, J. H. (2010). A widespread chromosomal inversion polymorphism contributes to a major life-history transition, local adaptation, and reproductive isolation. *PLoS Biology*, 8(9), e1000500. doi:10.1371/journal.pbio.1000500

- Lu, J., & Wu, C. (2005). Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proceedings of the National Academy of Sciences*, *102*(11), 4063-4067. doi:10.1073/pnas.0500436102
- Ludwig, A., Congiu, L., Pitra, C., Fickel, J., Gessner, J., Fontana, F., Patarnello, T., & Zane, L. (2003). Nonconcordant evolutionary history of maternal and paternal lineages in Adriatic sturgeon. *Molecular Ecology*, *12*(12), 3253-3264. doi:10.1046/j.1365-294X.2003.01999.x
- Lukhtanov, V. A., Dantchenko, A. V., Kandul, N. P., Plotkin, J. B., Pierce, N. E., & Haig, D. (2005). Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies. *Nature*, *436*(7049), 385-389. doi:10.1038/nature03704
- Lynch, M., Koskella, B., & Schaack, S. (2006). Mutation pressure and the evolution of organelle genomic architecture. *Science*, *311*(5768), 1727-1730. doi:10.1126/science.1118884
- Malmos, K. B., Sullivan, B. K., & Lamb, T. (2001). Calling behavior and directional hybridization between two toads (*Bufo microscaphus* × *B. woodhousii*) in Arizona. *Evolution*, *55*(3), 626-630. doi:10.1554/0014-3820(2001)055[0626:CBADHB]2.0.CO;2
- Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic drift. *Molecular Biology and Evolution*, *27*(3), 661-670. doi:10.1093/molbev/msp282
- Manning, B. J., & Yusufzai, T. (2017). The ATP-dependent chromatin remodeling enzymes CHD6, CHD7, and CHD8 exhibit distinct nucleosome binding and remodeling activities. *The Journal of Biological Chemistry*, *292*(28), 11927-11936. doi:10.1074/jbc.M117.779470
- Mason, N. A., & Bowie, R. C. (2020). Plumage patterns: Ecological functions, evolutionary origins, and advances in quantification. *The Auk*, *137*(4), ukaa060. doi:10.1093/auk/ukaa060

- Mason, N. A., & Taylor, S. A. (2015). Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, 24(12), 3009-3025. doi:10.1111/mec.13140
- Mastrantonio, V., Urbanelli, S., & Porretta, D. (2019). Ancient hybridization and mtDNA introgression behind current paternal leakage and heteroplasmy in hybrid zones. *Scientific Reports*, 9(1), 19177-9. doi:10.1038/s41598-019-55764-w
- Mayr, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.
- Mayr, E. 1954. Change of genetic environment and evolution. In J. Huxley, A. C. Hardy, and E. B. Ford (Eds.), *Evolution as a Process* (pp. 157-180). Allen & Unwin
- Mayr, E. (1963). *Animal Species and Evolution*. Belknap Press of Harvard University Press.
- Mayr, E. (1970). *Populations, Species, and Evolution: An Abridgment of Animal Species and Evolution*. Belknap Press of Harvard University Press.
- McGibbon, W. H. (1974). A shank color mutation in Cornell random bred S.C. white leghorns. *Poultry Science*, 53(3), 1251-1253. doi:10.3382/ps.0531251
- McGuire, J. A., Linkem, C. W., Koo, M. S., Hutchison, D. W., Lappin, A. K., Orange, D. I., Lemos-Espinal, J., Riddle, B. R., & Jaeger, J. R. (2007). Mitochondrial introgression and incomplete lineage sorting through space and time: Phylogenetics of crotaphytid lizards. *Evolution*, 61(12), 2879-2897. doi:10.1111/j.1558-5646.2007.00239.x
- McKay, B. D., & Zink, R. M. (2010). The causes of mitochondrial DNA gene tree parphyly in birds. *Molecular Phylogenetics and Evolution*, 54(2), 647-650. doi:10.1016/j.ympev.2009.08.024
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytzky, A., Garimella, K., Altshuler, D., Gabriel, S., Daley, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, 20(9), 1297-1303. doi:10.1101/gr.107524.110

- Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in Genetics*, 29(9), 537-544. doi:10.1016/j.tig.2013.05.009
- Melo-Ferreira, J., Alves, P. C., Freitas, H., Ferrand, N., & Boursot, P. (2009). The genomic legacy from the extinct *Lepus timidus* to the three hare species of Iberia: Contrast between mtDNA, sex chromosomes and autosomes. *Molecular Ecology*, 18(12), 2643-2658. doi:10.1111/j.1365-294X.2009.04221.x
- Mila, B., Toews, D. P., Smith, T. B., & Wayne, R. K. (2011). A cryptic contact zone between divergent mitochondrial DNA lineages in southwestern North America supports past introgressive hybridization in the yellow-rumped warbler complex (*Aves: Dendroica coronata*). *Biological Journal of the Linnean Society*, 103(3), 696-706.
- Mitchell, N., Owens, G. L., Hovick, S. M., Rieseberg, L. H., & Whitney, K. D. (2019). Hybridization speeds adaptive evolution in an eight-year field experiment. *Scientific Reports*, 9(1), 6746-6746. doi:10.1038/s41598-019-43119-4
- Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution*, 49(4), 718-726. doi:10.2307/2410325
- Moore, I. T., Bonier, F., & Wingfield, J. C. (2005). Reproductive asynchrony and population divergence between two tropical bird populations. *Behavioral Ecology*, 16(4), 755-762. doi:10.1093/beheco/ari049
- Morales, H. E., Pavlova, A., Amos, N., Major, R., Kilian, A., Greening, C., & Sunnucks, P. (2018). Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nature Ecology & Evolution*, 2(8), 1258-1267. doi:10.1038/s41559-018-0606-3
- Morales, H. E., Faria, R., Johannesson, K., Larsson, T., Panova, M., Westram, A. M., & Butlin, R. K. (2019). Genomic architecture of parallel ecological divergence: Beyond a single environmental contrast. *Science Advances*, 5(12), eaav9963-eaav9963. doi:10.1126/sciadv.aav9963

- Muller, H. J. 1940. Bearing of the *Drosophila* work on systematics. In J. S. Huxley (Eds.) *The New Systematics* (pp. 185-268). Clarendon Press.
- Muller, H. J. 1942. Isolating mechanisms, evolution, and temperature. *Biology Symposium*, 6, 71-125.
- Nachman, M. (2002). Variation in recombination rate across the genome: evidence and implications. *Current Opinion in Genetics & Development*, 12(6), 657-663.  
doi:10.1016/s0959-437x(02)00358-1
- Nadeau, N. J., Whibley, A., Jones, R. T., Davey, J. W., Dasmahapatra, K. K., Baxter, S. W., Quail, M. A., Joron, M., Ffrench-Constant, R. H., Blaxter, M. L., Mallet, J., & Jiggins, C. D. (2012). Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1587), 343-353. doi:10.1098/rstb.2011.0198
- Navarro, A., & Barton, N. H. (2003). Accumulating postzygotic isolation genes in parapatry: A new twist on chromosomal speciation. *Evolution*, 57(3), 447-459.  
doi:10.1111/j.0014-3820.2003.tb01537.x
- Noor, M. A. F., Grams, K. L., Bertucci, L. A., & Reiland, J. (2001). Chromosomal inversions and the reproductive isolation of species. *Proceedings of the National Academy of Sciences*, 98(21), 12084-12088. doi:10.1073/pnas.221274498
- Oram, J. F., & Vaughan, A. M. (2000). ABCA1-mediated transport of cellular cholesterol and phospholipids to HDL apolipoproteins. *Current Opinion in Lipidology*, 11(3), 253-260.  
doi:10.1097/00041433-200006000-00005
- Ortiz-Barrientos, D., Engelstädter, J., & Rieseberg, L. H. (2016). Recombination rate evolution and the origin of species. *Trends in Ecology & Evolution*, 31(3), 226-236.  
doi:10.1016/j.tree.2015.12.016
- Panov, E. N., Rubtsov, A. A., & Monzиков, D. G. (2003). Hybridization between yellowhammer and pine bunting in Russia. *Dutch Birding*, 25, 17-31.

- Panov, E. N., Rubtsov, A. S., & Mordkovich, M. V. (2007). New data on the relationships between two species of buntings (*Emberiza citrinella* and *E. leucocephalos*) hybridizing in the areas of overlap of their ranges, *Zoologicheskii Zhurnal*, 86(11), 1362–1378.
- Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526-528. doi:10.1093/bioinformatics/bty633
- Presgraves, D. C. (2008). Sex chromosomes and speciation in *Drosophila*. *Trends in Genetics*, 24(7), 336-343. doi:10.1016/j.tig.2008.04.007
- Price, T. (2008). *Speciation in birds*. Roberts and Co.
- Pulido-Santacruz, P., Aleixo, A., & Weir, J. T. (2018). Morphologically cryptic Amazonian bird species pairs exhibit strong postzygotic reproductive isolation. *Proceedings of the Royal Society B: Biological Sciences*, 285(1874), 20172081. doi:10.1098/rspb.2017.2081
- R Core Team (2014). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <http://www.r-project.org/>
- Rieseberg, L. H. (2001). Chromosomal rearrangements and speciation. *Trends in Ecology & Evolution*, 16(7), 351-358. doi:10.1016/s0169-5347(01)02187-5
- Rodd, F. H., Hughes, K. A., Grether, G. F., & Baril, C. T. (2002). A possible non-sexual origin of mate preference: Are male guppies mimicking fruit? *Proceedings of the Royal Society. Series B: Biological Sciences*, 269(1490), 475-481. doi:10.1098/rspb.2001.1891
- Rosenblum, E. B., Hoekstra, H. E., & Nachman, M. W. (2004). Adaptive reptile color variation and the evolution of the Mc1r gene. *Evolution*, 58(8), 1794-1808.

- Rubtsov, A.S. 2007. Variability of songs of the yellowhammer (*Emberiza citrinella*) and pine bunting (*Emberiza leucocephala*) as an evidence of population structure and evolutionary history of the species, *Zoologicheskii Zhurnal*, 86(7), 863–876.
- Rubtsov, A. S., & Opaev, A. S. (2012). Phylogeny reconstruction of the yellowhammer (*Emberiza citrinella*) and pine bunting (*Emberiza leucocephala*) based on song and morphological characters. *Biology Bulletin*, 39(9), 715-728. doi:10.1134/S1062359012090087
- Rubtsov, A. S., & Tarasov, V. V. (2017). Relations between the yellowhammer (*Emberiza citrinella*) and the pine bunting (*Emberiza leucocephalos*) in the forested steppe of the Trans-Urals. *Biology Bulletin*, 44(9), 1059-1072. doi:10.1134/S1062359017090114
- Ruegg, K., Anderson, E. C., Boone, J., Pouls, J., & Smith, T. B. (2014). A role for migration-linked genes and genomic islands in divergence of a songbird. *Molecular Ecology*, 23(19), 4757-4769. doi:10.1111/mec.12842
- Sackton, T. B., Corbett-Detig, R. B., Nagaraju, J., Vaishna, L., Arunkumar, K. P., & Hartl, D. L. (2014). Positive selection drives faster-z evolution in silkmoths. *Evolution*, 68(8), 2331-2342. doi:10.1111/evo.12449
- Sæther, S. A., Sætre, G., Borge, T., Wiley, C., Svedin, N., Andersson, G., Veen, T., Haavie, J., Servedio, M. R., Bureš, S., Král, M., Hjernquist, M. B., Gustafsson, L., Träff, J., & Qvarnström, A. (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, 318(5847), 95-97. doi:10.1126/science.1141506
- Sætre, G., Borge, T., Lindroos, K., Haavie, J., Sheldon, B. C., Primmer, C., & Syvänen, A. (2003). Sex chromosome evolution and speciation in *Ficedula* flycatchers. *Proceedings of the Royal Society B: Biological Sciences*, 270(1510), 53-59. <https://doi.org/10.1098/rspb.2002.2204>
- Saetre, G.P., Moum, T., Bureš, S., Král, M., Adamjan, M., & Moreno, J. (1997). A sexually selected character displacement in flycatchers reinforces premating isolation. *Nature* 387(6633), 589–592. doi:10.1038/42451

- 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. *American Journal of Human Genetics*, 78(4), 629-644. doi:10.1086/502802
- Scott, G. R., Schulte, P. M., Egginton, S., Scott, A. L. M., Richards, J. G., & Milsom, W. K. (2011). Molecular evolution of cytochrome c oxidase underlies high-altitude adaptation in the bar-headed goose. *Molecular Biology and Evolution*, 28(1), 351-363. doi:10.1093/molbev/msq205
- Seddon, N., Botero, C. A., Tobias, J. A., Dunn, P. O., MacGregor, H. E., Rubenstein, D. R., Uy, J. A. C., Weir, J. T., Whittingham, L. A., & Safran, R. J. (2013). Sexual selection accelerates signal evolution during speciation in birds. *Proceedings of the Royal Society B: Biological Sciences*, 280(1766), 20131065.
- Seehausen, O., Terai, Y., Magalhaes, I. S., Carleton, K. L., Mrosso, H. D. J., Miyagi, R., van der Sluijs, I., Schneider, M. V., Maan, M. E., Tachida, H., Imai, H., & Okada, N. (2008). Speciation through sensory drive in cichlid fish. *Nature*, 455(7213), 620-626. doi:10.1038/nature07285
- Sloan, D. B., Havird, J. C., & Sharbrough, J. (2017). The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Molecular Ecology*, 26(8), 2212-2236. doi:10.1111/mec.13959
- Smukowski, C. S., & Noor, M. A. F. (2011). Recombination rate variation in closely related species. *Heredity*, 107(6), 496-508. doi:10.1038/hdy.2011.44
- Sota, T., & Kubota, K. (1998). Genital lock-and-key as a selective agent against hybridization. *Evolution*, 52(5), 1507-1513. doi:10.1111/j.1558-5646.1998.tb02033.x
- Southern, H. N. (1945). Polymorphism in *Poephila gouldiae* gould. *Journal of Genetics*, 47(1), 51-57.

- Stacklies, W., Redestig, H., Scholz, M., Walther, D., & Selbig, J. (2007). pcaMethods-a bioconductor package providing PCA methods for incomplete data. *Bioinformatics*, 23(9), 1164-1167. doi:10.1093/bioinformatics/btm069
- Strunnikova, N. V., Maminishkis, A., Barb, J. J., Wang, F., Zhi, C., Sergeev, Y., Chen, W., Edwards, A. O., Stambolian, D., Abecasis, G., Swaroop, A., Munson, P. J., & Miller, S. S. (2010). Transcriptome analysis and molecular signature of human retinal pigment epithelium. *Human Molecular Genetics*, 19(12), 2468-2486. doi:10.1093/hmg/ddq129
- Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123(3), 585-595.
- Theron, E., Hawkins, K., Bermingham, E., Ricklefs, R. E., & Mundy, N. I. (2001). The molecular basis of an avian plumage polymorphism in the wild: A melanocortin-1-receptor point mutation is perfectly associated with the melanic plumage morph of the bananaquit, *Coereba flaveola*. *Current Biology*, 11(8), 550-557.
- Thornton, K., & Long, M. (2002). Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. *Molecular Biology and Evolution*, 19(6), 918-925. doi:10.1093/oxfordjournals.molbev.a004149
- Todesco, M., Owens, G. L., Bercovich, N., Légaré, J., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muñoz, S., Nielsen, R., Donovan, L. A., ... Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, 584(7822), 602-607. doi:10.1038/s41586-020-2467-6
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907-3930. doi:10.1111/j.1365-294X.2012.05664.x
- Toews, D. P. L., Campagna, L., Taylor, S. A., Balakrishnan, C. N., Baldassarre, D. T., Deane-Coe, P. E., Harvey, M. G., Hooper, D. M., Irwin, D. E., Judy, C. D., Mason, N. A., McCormack, J. E., McCracken, K. G., Oliveros, C. H., Safran, R. J., Scordato, E. S. C., Stryjewski, K. F., Tigano, A., Uy, J. A. C., & Winger, B. M. (2016a). Genomic

- approaches to understanding population divergence and speciation in birds. *The Auk*, 133(1), 13-30. doi:10.1642/AUK-15-51.1
- Toews, D. L., Taylor, S., Vallender, R., Brelsford, A., Butcher, B., Messer, P., & Lovette, I. (2016b). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, 26(17), 2313-2318. doi:10.1016/j.cub.2016.06.034
- Toews, D. P. L., & Irwin, D. E. (2008). Cryptic speciation in a Holarctic passerine revealed by genetic and bioacoustic analyses. *Molecular Ecology*, 17(11), 2691-2705. doi:10.1111/j.1365-294x.2008.03769.x
- Tonzo, V., Papadopoulou, A., & Ortego, J. (2019). Genomic data reveal deep genetic structure but no support for current taxonomic designation in a grasshopper species complex. *Molecular Ecology*, 28(17), 3869-3886. doi:10.1111/mec.15189
- Toomey, M. B., Lopes, R. J., Araújo, P. M., Johnson, J. D., Gazda, M. A., Afonso, S., Mota, P. G., Koch, R. E., Hill, G. E., Corbo, J. C., & Carneiro, M. (2017). High-density lipoprotein receptor SCARB1 is required for carotenoid coloration in birds. *Proceedings of the National Academy of Sciences*, 114(20), 5219-5224. doi:10.1073/pnas.1700751114
- Trickett, A. J., & Butlin, R. K. (1994). Recombination suppressors and the evolution of new species. *Heredity*, 73(4), 339-345. doi:10.1038/hdy.1994.180
- Turner, S. (2018). qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. *Journal of Open Source Software*, 3(25), 731. doi:10.21105/joss.00731
- Tuttle, E. M. (2003). Alternative reproductive strategies in the white-throated sparrow: Behavioral and genetic evidence. *Behavioral Ecology*, 14(3), 425-432. doi:10.1093/beheco/14.3.425
- Uy, J. A. C., Moyle, R. G., & Filardi, C. E. (2009). plumage and song differences mediate species recognition between incipient flycatcher species of the Solomon Islands. *Evolution*, 63(1), 153-164. doi:10.1111/j.1558-5646.2008.00530.x

- Veale, A. J., & Russello, M. A. (2017). An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Scientific Reports*, 7(1), 1747-10. doi:10.1038/s41598-017-01890-2
- Voelker, G. (2002). Systematics and historical biogeography of wagtails: Dispersal versus vicariance revisited. *The Condor*, 104(4), 725-739. doi:10.1650/0010-5422(2002)104[0725:SAHBOW]2.0.CO;2
- Walsh, J. B. (1982). Rate of accumulation of reproductive isolation by chromosome rearrangements. *The American Naturalist*, 120(4), 510-532. doi:10.1086/284008
- Wang, M., Wang, S., Li, Y., Jhala, Y., Thakur, M., Otecko, N. O., Si, J., Chen, H., Shapiro, B., Nielsen, R., Zhang, Y., & Wu, D. (2020). Ancient hybridization with an unknown population facilitated high altitude adaptation of canids. *Molecular Biology and Evolution*, 37(9), 2616-2629. doi:10.1093/molbev/msaa113
- Warren, W. C., Clayton, D. F., 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(7289), 757-762. doi:10.1038/nature08819
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358-1370. doi:10.1111/j.1558-5646.1984.tb05657.x
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, 33(6), 427-440. doi:10.1016/j.tree.2018.04.002
- West-Eberhard, M. J. (1983). Sexual selection, social competition, and speciation. *The Quarterly Review of Biology*, 58(2), 155-183. doi:10.1086/413215
- Widelitz, R. B., Jiang, T., Lu, J., & Chuong, C. (2000).  $\beta$ -catenin in epithelial morphogenesis: Conversion of part of avian foot scales into feather buds with a mutated  $\beta$ -catenin. *Developmental Biology*, 219(1), 98-114. doi:10.1006/dbio.1999.9580

- Wojcieszek, J. M., & Simmons, L. W. (2013). Divergence in genital morphology may contribute to mechanical reproductive isolation in a millipede. *Ecology and Evolution*, 3(2), 334-343. doi:10.1002/ece3.466
- Wu, C. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851-865. doi:10.1046/j.1420-9101.2001.00335.x
- Xu, J., Lin, S., Gao, X., Nie, Q., Luo, Q., & Zhang, X. (2017). Mapping of id locus for dermal shank melanin in a Chinese indigenous chicken breed. *Journal of Genetics*, 96(6), 977-983. doi:10.1007/s12041-017-0862-z
- Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics*, 44(7), 821-824. doi:10.1038/ng.2310
- Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17(9), 2107-2121. doi:10.1111/j.1365-294X.2008.03

## Appendices

### Appendix A Supplementary Tables and Figures

#### A.1 Supplementary Tables

**Supplementary Table 2.1.** Detailed information on all the samples included in this study. Explanations for the abbreviations used in the “Phenotypic Class” and “Geographic Distribution” columns can be found in the methods sections of Chapter 2. In the “sex” column, “m” stands for male, “f” stands for female and “uk” stands for unknown. The “TH” column contains phenotypic scores for each individual for the throat plumage trait. The “BR” column contains phenotypic scores for each individual for the brow plumage trait. The “BG” column contains phenotypic scores for each individual for the background colour plumage trait. In the “Pheno Class” (Phenotypic Class) column, “OUT” stands for outgroup, “FML” stands for female and “UK” stands for unknown. The numbers in the “Sampling Location” column correspond to those that appear in Figure 2.1A. In all columns, a “NA” observation stands for “Not Applicable”.

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_ASR00_01	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.26	115.21	28
Emberiza_GBS1_ASR05_05	<i>E. hortulana</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA
Emberiza_GBS1_ASR05_14	<i>E. citrinella</i>	m	1	2	1	SC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS1_ASR05_17	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS1_ASR05_18	<i>E. citrinella</i>	m	2	1	1	SC	Allopatric	Collected	51.2	57.27	12

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_ASR05_33	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_35	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_36	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_37	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_43	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_45	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_47	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS1_ASR05_54	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_55	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_56	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_59	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_61	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_66	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_ASR05_68	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS1_AWH_067	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS1_AWH_076	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Burke Museum, USA	52.528	104.297	24

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_AWH_077	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS1_AWH_148	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	56.405	105.463	25
Emberiza_GBS1_AWH_169	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	56.405	105.463	25
Emberiza_GBS1_BKS_1572	<i>E. citrinella</i>	m	1	0	0	SC	Allopatric	Burke Museum, USA	54.573	39.182	7
Emberiza_GBS1_BKS_1646	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Burke Museum, USA	54.573	39.182	7
Emberiza_GBS1_BKS_1821	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Burke Museum, USA	51.412	34.547	6
Emberiza_GBS1_BKS_1841	<i>E. citrinella</i>	m	1	1	1	SC	Near Sympatric	Burke Museum, USA	56.318	59.18	13
Emberiza_GBS1_BKS_1928	<i>E. citrinella</i>	m	1	0	1	SC	Near Sympatric	Burke Museum, USA	56.456	58.065	13
Emberiza_GBS1_BKS_1945	<i>E. citrinella</i>	m	1	0	0	SC	Near Sympatric	Burke Museum, USA	56.456	58.065	13

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_BKS_1957	<i>E. citrinella</i>	m	2	1	0	SC	Near Sympatric	Burke Museum, USA	56.456	58.065	13
Emberiza_GBS1_BKS_1988	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Burke Museum, USA	55.54	39.352	7
Emberiza_GBS1_BKS_1989	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Burke Museum, USA	55.54	39.352	7
Emberiza_GBS1_BKS_2040	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Burke Museum, USA	55.54	39.352	7
Emberiza_GBS1_BKS_2041	<i>E. citrinella</i>	m	1	0	0	SC	Allopatric	Burke Museum, USA	55.54	39.352	7
Emberiza_GBS1_CDS_4883	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS1_CDS_4906	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS1_DAB_296	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS1_DAB_297	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	51.346	106.5096	26

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_ENP05_06	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS1_ENP05_08	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS1_ENP05_14	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	52.53	85.2	16
Emberiza_GBS1_ENP05_16	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS1_ENP05_20	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS1_ENP05_22	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS1_ENP05_25	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS1_ENP97_11	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS1_ENP97_12	Hybrid	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS1_ENP97_25	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	58.33	44.76	11
Emberiza_GBS1_EVN_327	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_EVN_331	<i>E. citrinella</i>	m	0	0	2	SC	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_EVN_357	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_EVN_363	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_EVN_573	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Darwin Museum, Russia	57.71	39.34	7
Emberiza_GBS1_EVN_574	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Darwin Museum, Russia	57.71	39.34	7
Emberiza_GBS1_IUK_615	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_IUK_631	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_IVF_309	Hybrid	m	3	2	4	WH	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_IVF_390	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS1_IVF_658	<i>E. cioides</i>	uk	NA	NA	NA	OUT	Outgroup	Drovetsky expedition	NA	NA	NA
Emberiza_GBS1_IVF_682	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Drovetsky expedition	50.5036	115.0029	28
Emberiza_GBS1_JML_291	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS1_M05_05	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	55.28	20.97	4

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_MSU_N247	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	50.38	95.1	22
Emberiza_GBS1_MSU_N339	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	50.4343	91.4182	20
Emberiza_GBS1_MSU_P66	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	50.13	95.09	22
Emberiza_GBS1_NVN_015	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Darwin Museum, Russia	57.71	39.34	7
Emberiza_GBS1_RCF_1807	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	50.39	91.36	20
Emberiza_GBS1_RCF_1971	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	50.04	95.08	22
Emberiza_GBS1_RCF_2235	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS1_SVD_2134	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Burke Museum, USA	43.54	40.47	9
Emberiza_GBS1_SVD_2695	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_SVD_3479	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Drovetsky expedition	49.6439	110.1652	27
Emberiza_GBS1_SVD_518	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Burke Museum, USA	52.46	104.41	24
Emberiza_GBS1_SVD_519	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.46	104.41	24
Emberiza_GBS1_SVD_531	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Burke Museum, USA	52.46	104.41	24
Emberiza_GBS1_SWM_03	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Swedish NHM	57.99	12.49	1
Emberiza_GBS1_SWM_10	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Swedish NHM	65.86	21.48	5
Emberiza_GBS1_SWM_12	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Swedish NHM	65.86	21.48	5
Emberiza_GBS1_VGR_400	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS1_VGR_440	<i>E. citrinella</i>	m	1	0	1	SC	Sympatric	Burke Museum, USA	56.405	105.463	25
Emberiza_GBS1_VGR_508	<i>E. citrinella</i>	m	0	0	1	SC	Sympatric	Burke Museum, USA	56.405	105.463	25

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS1_VM_282a	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Burke Museum, USA	50.44	143.18	30
Emberiza_GBS2_ASR05_02	<i>E. citrinella</i>	m	1	0	0	SC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS2_ASR05_15	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS2_ASR05_20	<i>E. citrinella</i>	m	2	1	0	SC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS2_ASR05_22	<i>E. citrinella</i>	m	2	1	0	SC	Allopatric	Collected	51.2	57.27	12
Emberiza_GBS2_ASR05_32	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_38	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_39	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_40	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_41	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_44	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_46	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_49	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	51.12	118.56	29
Emberiza_GBS2_ASR05_51	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS2_ASR05_52	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS2_ASR05_57	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS2_ASR05_58	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS2_ASR05_60	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS2_ASR05_67	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_ASR98_11	Hybrid	m	4	2	0	CH	Sympatric	Darwin Museum, Russia	53.32	107.12	26
Emberiza_GBS2_ASR98_17	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Darwin Museum, Russia	52.8	104.74	24
Emberiza_GBS2_ASR99_01	Hybrid	m	0	0	4	WH	Sympatric	Darwin Museum, Russia	55.33	93.65	21
Emberiza_GBS2_AWH_039	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_AWH_152	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	56.405	105.463	25
Emberiza_GBS2_BKS_1583	<i>E. citrinella</i>	m	1	1	0	SC	Allopatric	Burke Museum, USA	54.573	39.182	7
Emberiza_GBS2_BKS_1609	<i>E. citrinella</i>	m	1	0	0	SC	Allopatric	Burke Museum, USA	54.573	39.182	7
Emberiza_GBS2_BKS_1654	<i>E. citrinella</i>	m	2	1	0	SC	Allopatric	Burke Museum, USA	51.349	37.123	6
Emberiza_GBS2_BKS_1667	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Burke Museum, USA	51.349	37.123	6

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_BKS_1710	Hybrid	m	4	3	1	CH	Allopatric	Burke Museum, USA	51.349	37.123	6
Emberiza_GBS2_BKS_1859	<i>E. citrinella</i>	m	1	0	1	SC	Near Sympatric	Burke Museum, USA	56.318	59.18	13
Emberiza_GBS2_BKS_2017	<i>E. citrinella</i>	m	1	0	1	SC	Allopatric	Burke Museum, USA	55.54	39.352	7
Emberiza_GBS2_DAB_291	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS2_DAB_299	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS2_DAB_301	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS2_DAB_303	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS2_DAB_308	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Near Sympatric	Burke Museum, USA	51.346	106.5096	26
Emberiza_GBS2_ENP05_04	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS2_ENP05_09	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS2_ENP05_11	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_ENP05_12	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS2_ENP05_15	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS2_ENP05_21	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS2_ENP05_24	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS2_ENP05_26	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	53.35	83.75	16
Emberiza_GBS2_ENP97_01	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.83	77.67	15
Emberiza_GBS2_ENP97_03	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.83	77.67	15
Emberiza_GBS2_ENP97_05	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_07	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_09	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_10	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_13	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_15	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_20	Hybrid	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_ENP97_21	Hybrid	uk	NA	NA	NA	UK	Sympatric	Collected	54.85	83.11	16
Emberiza_GBS2_EVN_328	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS2_EVN_366	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_EVN_570	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Darwin Museum, Russia	57.71	39.34	7
Emberiza_GBS2_F_12483	<i>E. cirrus</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA
Emberiza_GBS2_FMNH_01	<i>E. stewarti</i>	uk	NA	NA	NA	OUT	Outgroup	Field Museum, USA	NA	NA	NA
Emberiza_GBS2_IUK_2306	<i>E. aureola</i>	uk	NA	NA	NA	OUT	Outgroup	Drovetsky expedition	NA	NA	NA
Emberiza_GBS2_IUK_2343	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Drovetsky expedition	49.6439	110.1652	27
Emberiza_GBS2_IUK_702	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS2_IUK_703	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Bell Museum, USA	61.45	38.67	8
Emberiza_GBS2_IUK_801	<i>E. citrinella</i>	f	NA	NA	NA	FML	Allopatric	Bell Museum, USA	65.85	44.24	10
Emberiza_GBS2_M05_01	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	55.28	20.97	4
Emberiza_GBS2_M05_03	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	55.28	20.97	4
Emberiza_GBS2_M05_08	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	55.28	20.97	4
Emberiza_GBS2_M05_10	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Collected	55.28	20.97	4

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_MSU_N71	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	50.13	95.09	22
Emberiza_GBS2_RCF_1949	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Burke Museum, USA	50.44	90.01	20
Emberiza_GBS2_RCF_1970	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Near Sympatric	Burke Museum, USA	50.04	95.08	22
Emberiza_GBS2_RCF_2191b	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_RCF_2192	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_RCF_2230	<i>E. leucocephalos</i>	m	6	7	7	SL	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_SVD_1978	<i>E. calandra</i>	uk	NA	NA	NA	OUT	Outgroup	Burke Museum, USA	NA	NA	NA
Emberiza_GBS2_SVD_3563	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Drovetsky expedition	50.5036	115.0029	28
Emberiza_GBS2_SVD_514	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Burke Museum, USA	52.46	104.41	24

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS2_SVD_524	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Burke Museum, USA	52.46	104.41	24
Emberiza_GBS2_SVD_533	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Burke Museum, USA	57.28	97.18	23
Emberiza_GBS2_SWM_11	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	Swedish NHM	59.81	17.05	2
Emberiza_GBS2_VGR_343	Hybrid	m	3	2	0	CH	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_VGR_350	<i>E. leucocephalos</i>	uk	NA	NA	NA	UK	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_VGR_355	<i>E. leucocephalos</i>	m	6	7	7	SL	Sympatric	Burke Museum, USA	52.528	104.297	24
Emberiza_GBS2_VM_285a	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Burke Museum, USA	50.44	143.18	30
Emberiza_GBS2_ZMUC_09	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Allopatric	ZMUC, Denmark	51.71	18.61	3
Emberiza_GBS4_F_13069	<i>E. cirrus</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA
Emberiza_GBS4_FM_347946	<i>E. stewarti</i>	uk	NA	NA	NA	OUT	Outgroup	Field Museum, USA	NA	NA	NA
Emberiza_GBS4_K_Z2756	<i>E. cirrus</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS4_KZ_2766	<i>E. cirius</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA
Emberiza_GBS4_MIM_165	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS4_R06_01	Hybrid	m	3	1	0	CH	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS4_RYA_2397	Hybrid	m	2	0	7	WH	Sympatric	Zoological museum, Russia	53.48	78.82	15
Emberiza_GBS4_RYA_2680	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS4_RYA_3178	<i>E. leucocephalos</i>	m	NA	NA	NA	UK	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS4_SVN_2335	<i>E. leucocephalos</i>	m	NA	NA	NA	UK	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS4_XD_548	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_552	Hybrid	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_603	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_610	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_612	Hybrid	m	2	0	4	WH	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS4_XD_613	Hybrid	m	4	4	6	LH	Near Sympatric	Collected	55.22	65.42	14

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS4_XD_617	Hybrid	m	5	5	7	LH	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS4_XD_619	<i>E. citrinella</i>	m	1	0	0	SC	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS4_XD_621	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	55.19	65.26	14
Emberiza_GBS4_XD_623	<i>E. citrinella</i>	m	1	0	0	SC	Near Sympatric	Collected	55.19	65.26	14
Emberiza_GBS4_XD_624	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	55.19	65.26	14
Emberiza_GBS4_XD_625	<i>E. citrinella</i>	m	1	0	0	SC	Near Sympatric	Collected	55.19	65.26	14
Emberiza_GBS4_XD_630	Hybrid	m	5	5	6	LH	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS4_XD_631	<i>E. citrinella</i>	m	1	0	0	SC	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS4_XD_632	Hybrid	m	4	3	2	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_634	<i>E. citrinella</i>	m	2	0	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_635	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_636	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_637	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_639	Hybrid	m	5	5	7	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_640	<i>E. leucocephalos</i>	m	7	7	4	SL	Sympatric	Collected	50.73	86.34	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS4_XD_645	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_647	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_649	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_650	Hybrid	m	5	5	6	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_653	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_654	<i>E. leucocephalos</i>	m	7	7	6	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_655	Hybrid	m	7	7	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_656	Hybrid	m	6	7	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_657	<i>E. citrinella</i>	m	1	0	0	SC	Allopatric	Collected	56.06	36.13	7
Emberiza_GBS4_XD_658	<i>E. citrinella</i>	m	0	0	1	SC	Allopatric	Collected	56.06	36.13	7
Emberiza_GBS4_XD_659	<i>E. citrinella</i>	m	1	1	0	SC	Allopatric	Collected	56.06	36.13	7
Emberiza_GBS4_XD_663	Hybrid	m	1	2	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_666	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_670	<i>E. citrinella</i>	m	0	0	2	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_672	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_674	Hybrid	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_684	<i>E. citrinella</i>	m	0	1	0	SC	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS4_XD_688	Hybrid	f	NA	NA	NA	FML	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS4_XD_699	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.64	87.96	17
Emberiza_GBS4_XD_700	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.64	87.96	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS4_XD_792	Hybrid	m	4	2	6	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_798	Hybrid	m	0	0	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_800	Hybrid	m	3	0	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_930	Hybrid	m	7	7	1	YH	Sympatric	Collected	50.31	87.6	17
Emberiza_GBS4_XD_934	Hybrid	m	5	5	5	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_940	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_941	Hybrid	m	3	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_946	Hybrid	m	4	4	5	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_947	Hybrid	m	6	3	1	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_948	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_949	Hybrid	m	0	0	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_950	Hybrid	m	5	3	1	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_957	Hybrid	m	4	3	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_958	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_960	Hybrid	m	7	7	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_961	Hybrid	m	3	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_962	Hybrid	m	7	6	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_967	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_968	Hybrid	m	3	2	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_971	Hybrid	m	5	4	7	LH	Sympatric	Collected	50.73	86.34	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS4_XD_972	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_976	Hybrid	m	3	2	1	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_977	Hybrid	m	3	2	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_979	Hybrid	m	7	7	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_990	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_994	Hybrid	m	5	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_995	Hybrid	m	2	3	7	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_XD_997	<i>E. citrinella</i>	m	2	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS4_ZM_1164	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Zoological museum, Russia	54.93	86.82	18
Emberiza_GBS4_ZM_1184	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Zoological museum, Russia	54.93	86.82	18
Emberiza_GBS4_ZM_1335	Hybrid	m	3	3	7	WH	Sympatric	Zoological museum, Russia	51.88	80.1	15
Emberiza_GBS5_ASR05_62_2	<i>E. leucocephalos</i>	m	7	7	7	PL	Allopatric	Collected	50.21	115.06	28
Emberiza_GBS5_EAK_344	<i>E. stewarti</i>	uk	NA	NA	NA	OUT	Outgroup	Zoological museum, Russia	NA	NA	NA
Emberiza_GBS5_ENP05_02	<i>E. citrinella</i>	uk	NA	NA	NA	UK	Sympatric	Collected	51.96	85.96	17
Emberiza_GBS5_F_13088	<i>E. cirrus</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS5_FM_347945	<i>E. stewarti</i>	uk	NA	NA	NA	OUT	Outgroup	Field Museum, USA	NA	NA	NA
Emberiza_GBS5_KZ_2762	<i>E. cirrus</i>	uk	NA	NA	NA	OUT	Outgroup	Collected	NA	NA	NA
Emberiza_GBS5_RYA_3003	<i>E. leucocephalos</i>	m	NA	NA	NA	UK	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS5_SVN_2336	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Allopatric	Zoological museum, Russia	50.68	142.97	30
Emberiza_GBS5_XD_549	Hybrid	m	7	7	1	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_550	Hybrid	m	5	3	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_602	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_605	<i>E. leucocephalos</i>	m	7	7	4	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_608	Hybrid	m	4	1	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_609	<i>E. leucocephalos</i>	m	7	7	6	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_611	Hybrid	m	3	3	0	CH	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS5_XD_615	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS5_XD_616	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	55.22	65.42	14
Emberiza_GBS5_XD_620	Hybrid	m	3	4	0	CH	Near Sympatric	Collected	55.19	65.26	14

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS5_XD_622	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	55.19	65.26	14
Emberiza_GBS5_XD_626	<i>E. citrinella</i>	m	0	0	0	PC	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS5_XD_627	Hybrid	m	6	6	0	YH	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS5_XD_628	Hybrid	m	7	7	1	YH	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS5_XD_629	<i>E. leucocephalos</i>	m	7	7	7	PL	Near Sympatric	Collected	54.67	64.88	14
Emberiza_GBS5_XD_633	Hybrid	m	6	5	0	YH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_638	Hybrid	m	1	0	6	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_642	Hybrid	m	3	2	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_643	Hybrid	m	2	1	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_646	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_648	<i>E. citrinella</i>	m	1	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_651	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_660	<i>E. citrinella</i>	m	1	1	0	SC	Allopatric	Collected	56.06	36.13	7
Emberiza_GBS5_XD_661	<i>E. citrinella</i>	m	0	0	0	PC	Allopatric	Collected	56.06	36.13	7
Emberiza_GBS5_XD_667	Hybrid	m	4	4	1	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_668	Hybrid	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_683	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Collected	53.09	83.84	16

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS5_XD_685	<i>E. citrinella</i>	m	1	1	1	SC	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS5_XD_686	Hybrid	m	5	5	7	LH	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS5_XD_687	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS5_XD_689	Hybrid	m	4	3	0	CH	Sympatric	Collected	51.57	85.56	17
Emberiza_GBS5_XD_690	Hybrid	m	1	1	5	WH	Sympatric	Collected	53.09	83.84	16
Emberiza_GBS5_XD_694	Hybrid	m	6	5	7	LH	Sympatric	Collected	50.64	87.96	17
Emberiza_GBS5_XD_698	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.64	87.96	17
Emberiza_GBS5_XD_794	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_795	<i>E. citrinella</i>	m	0	0	1	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_797	<i>E. leucocephalos</i>	m	6	6	6	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_799	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_924	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.02	89.23	19
Emberiza_GBS5_XD_925	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.02	89.23	19
Emberiza_GBS5_XD_927	<i>E. leucocephalos</i>	m	7	7	6	SL	Sympatric	Collected	50.31	87.6	17
Emberiza_GBS5_XD_928	<i>E. citrinella</i>	m	2	1	0	SC	Sympatric	Collected	50.31	87.6	17
Emberiza_GBS5_XD_931	Hybrid	m	4	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_942	<i>E. citrinella</i>	m	1	0	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_943	<i>E. citrinella</i>	m	1	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_944	Hybrid	m	3	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_952	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS5_XD_954	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_955	Hybrid	m	5	3	7	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_956	<i>E. citrinella</i>	m	2	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_963	Hybrid	m	0	0	7	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_964	<i>E. leucocephalos</i>	m	7	7	5	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_965	<i>E. citrinella</i>	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_966	<i>E. citrinella</i>	m	2	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_969	Hybrid	m	5	5	5	LH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_970	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_973	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_974	Hybrid	m	1	1	5	WH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_975	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_978	<i>E. citrinella</i>	m	1	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_980	<i>E. leucocephalos</i>	m	6	7	4	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_981	<i>E. leucocephalos</i>	f	NA	NA	NA	FML	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_982	<i>E. citrinella</i>	m	0	0	1	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_983	Hybrid	m	3	1	0	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_984	<i>E. citrinella</i>	m	2	1	0	SC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_985	<i>E. leucocephalos</i>	m	6	6	7	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_987	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17

Sample ID	Species	Sex	TH	BR	BG	Pheno Class	Geographic Distribution	Source	Latitude (°N)	Longitude (°E)	Sampling Location
Emberiza_GBS5_XD_989	<i>E. leucocephalos</i>	m	7	7	5	SL	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_992	<i>E. citrinella</i>	m	0	0	0	PC	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_993	Hybrid	m	4	3	1	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_XD_999	Hybrid	m	3	0	1	CH	Sympatric	Collected	50.73	86.34	17
Emberiza_GBS5_ZM_1162	<i>E. citrinella</i>	m	1	0	0	PC	Sympatric	Zoological museum, Russia	54.93	86.82	18
Emberiza_GBS5_ZM_1183	Hybrid	m	7	7	0	YH	Sympatric	Zoological museum, Russia	54.93	86.82	18
Emberiza_GBS5_ZM_1197	<i>E. leucocephalos</i>	m	7	7	7	PL	Sympatric	Zoological museum, Russia	53.43	83.93	16
Emberiza_GBS5_ZM_1200	Hybrid	m	2	0	6	WH	Sympatric	Zoological museum, Russia	53.43	83.93	16
Emberiza_GBS5_ZM_1220	Hybrid	m	3	3	0	CH	Sympatric	Zoological museum, Russia	53.43	83.93	16
Emberiza_GBS5_ZM_1267	<i>E. leucocephalos</i>	m	6	6	7	SL	Sympatric	Zoological museum, Russia	53.37	78.02	15
Emberiza_GBS5_ZM_1375	<i>E. citrinella</i>	m	1	0	0	PC	Sympatric	Zoological museum, Russia	53.24	83.51	16

**Supplementary Table. 2.2.** Detailed information on the genomic locations and functions of the 134 mitonuclear genes investigated in this study. In the “Function” column, “ETC” stand for electron transport chain. Mitonuclear gene names are written as they appear in Hill (2019).

Mitonuclear Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
DNUFV3	1	4107093	4115372	4111232.5	Structural subunit of ETC complex I
MRPL30	1	31234806	31237071	31235938.5	Mitochondrial large ribosomal subunit protein
NARS2	1	86711128	86758715	86734921.5	aminoacyl-tRNA synthetase
NDUFB4	1	92119879	92122800	92121339.5	Structural subunit of ETC complex I
COX17	1	92525354	92527833	92526593.5	Assembly factor/ancillary protein for ETC complex IV
TIMMDC1	1	103173445	103180367	103176906	Assembly factor/ancillary protein for ETC complex I
MRPL39	1	112594984	112609353	112602168.5	Mitochondrial large ribosomal subunit protein
ATP5J	1	112654573	112658971	112656772	Structual subunit of ETC complex V
GARS	2	4436681	4465691	4451186	aminoacyl-tRNA synthetase
MRPL32	2	32999048	33001488	33000268	Mitochondrial large ribosomal subunit protein
FARS2	2	44068036	44292764	44180400	aminoacyl-tRNA synthetase
MRPL3	2	62358147	62388654	62373400.5	Mitochondrial large ribosomal subunit protein
MRPL36	2	91135578	91135946	91135762	Mitochondrial large ribosomal subunit protein
NDUFS6	2	91137018	91144551	91140784.5	Structural subunit of ETC complex I
NDUFV2	2	104093616	104113139	104103377.5	Structural subunit of ETC complex I
MRPL15	2	116382079	116395598	116388838.5	Mitochondrial large ribosomal subunit protein
TMEM70	2	124369603	124374001	124371802	Assembly factor/ancillary protein for ETC complex V

Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
NDUFAF6	2	132905588	132922459	132914023.5	Assembly factor/ancillary protein for ETC complex I
UQCRB	2	133375437	133380089	133377763	Structural subunit of ETC complex III
COX6C	2	134936993	134941743	134939368	Structural subunit of ETC complex IV
MRPL13	2	143508441	143536509	143522475	Mitochondrial large ribosomal subunit protein
NDUFB9	2	145176891	145187833	145182362	Structural subunit of ETC complex I
NDUFAF5	3	5098403	5105436	5101919.5	Assembly factor/ancillary protein for ETC complex I
MRPL33	3	7321655	7343178	7332416.5	Mitochondrial large ribosomal subunit protein
EPRS	3	9675151	9718481	9696816	aminoacyl-tRNA synthetase
LRPPRC	3	16539177	16615242	16577209.5	Assembly factor/ancillary protein for ETC complex IV
MRPL2	3	19867530	19869125	19868327.5	Mitochondrial large ribosomal subunit protein
PET117	3	23719042	23720746	23719894	Assembly factor/ancillary protein for ETC complex IV
AARS2	3	31074744	31093395	31084069.5	aminoacyl-tRNA synthetase
MRPL14	3	31303292	31314337	31308814.5	Mitochondrial large ribosomal subunit protein
TFB2M	3	33650775	33661865	33656320	Mitochondrial Transcription Factor B2
NDUFAF7	3	34173535	34180273	34176904	Assembly factor/ancillary protein for ETC complex I
TFB1M	3	54775521	54800918	54788219.5	Mitochondrial Transcription Factor B1
MRPL18	3	57591256	57593649	57592452.5	Mitochondrial large ribosomal subunit protein
NDUFAF4	3	74973096	74976828	74974962	Assembly factor/ancillary protein for ETC complex I
COX7A	3	83154874	83158710	83156792	Structural subunit of ETC complex IV

Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
MRPL19	3	107287972	107290732	107289352	Mitochondrial large ribosomal subunit protein
COX18	4	1462451	1466733	1464592	Assembly factor/ancillary protein for ETC complex IV
MRPL1	4	2274392	2279731	2277061.5	Mitochondrial large ribosomal subunit protein
NDUFC1	4	9943216	9945937	9944576.5	Structural subunit of ETC complex I
MRPL35	4	65701473	65703901	65702687	Mitochondrial large ribosomal subunit protein
NDUFS8	5	7953211	7954951	7954081	Structural subunit of ETC complex I
DNUFV1	5	8059469	8061462	8060465.5	Structural subunit of ETC complex I
MRPL23	5	13894159	13905268	13899713.5	Mitochondrial large ribosomal subunit protein
NDUFS3	5	21325493	21329936	21327714.5	Structural subunit of ETC complex I
NDUFAF1	5	23445722	23451657	23448689.5	Assembly factor/ancillary protein for ETC complex I
COX16	5	26716109	26760850	26738479.5	Assembly factor/ancillary protein for ETC complex IV
NUBPL	5	34372043	34456539	34414291	Assembly factor/ancillary protein for ETC complex I
DNUFB1	5	45442861	45445862	45444361.5	Structural subunit of ETC complex I
APOPT1	5	51884411	51890971	51887691	Assembly factor/ancillary protein for ETC complex IV
TFAM	6	4545977	4553892	4549934.5	Mitochondrial Transcription Factor A
NDUFB8	6	16430817	16434575	16432696	Structural subunit of ETC complex I
COX15	6	21184712	21188712	21186712	Assembly factor/ancillary protein for ETC complex IV
Twinkle	6	22511790	22514648	22513219	mtDNA Helicase
NDUFA10	7	1914517	1940529	1927523	Structural subunit of ETC complex I

Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
ATP5G3	7	16650485	16653714	16652099.5	Structural subunit of ETC complex V
NDUFS1	7	20857200	20871606	20864403	Structural subunit of ETC complex I
FASTKD2	7	21078508	21088901	21083704.5	Assembly factor/ancillary protein for ETC complex IV
NDUFB3	7	22144540	22146995	22145767.5	Structural subunit of ETC complex I
DARS	7	33263917	33300247	33282082	aminoacyl-tRNA synthetase
DARS2	8	2714489	2729279	2721884	aminoacyl-tRNA synthetase
UQCRH	8	18754346	18755566	18754956	Structural subunit of ETC complex III
ATPAF1	8	19066784	19075132	19070958	Assembly factor/ancillary protein for ETC complex V
MRPL37	8	22833154	22837987	22835570.5	Mitochondrial large ribosomal subunit protein
PARS2	8	22937524	22940092	22938808	aminoacyl-tRNA synthetase
FARSB	9	8954302	8991289	8972795.5	aminoacyl-tRNA synthetase
NDUFB5	9	19985641	19991221	19988431	Structural subunit of ETC complex I
COX5A	10	1899415	1901700	1900557.5	Structural subunit of ETC complex IV
POLG	10	12982224	12992974	12987599	Subunit of DNA polymerase gamma
COX4	11	21106	24901	23003.5	Structural subunit of ETC complex IV
AARS	11	5000660	5017549	5009104.5	aminoacyl-tRNA synthetase
KARS	11	12556178	12564535	12560356.5	aminoacyl-tRNA synthetase
UQCRC1	11	14269328	14270098	14269713	Structural subunit of ETC complex III
ACAD9	12	1576425	1596683	1586554	Assembly factor/ancillary protein for ETC complex I

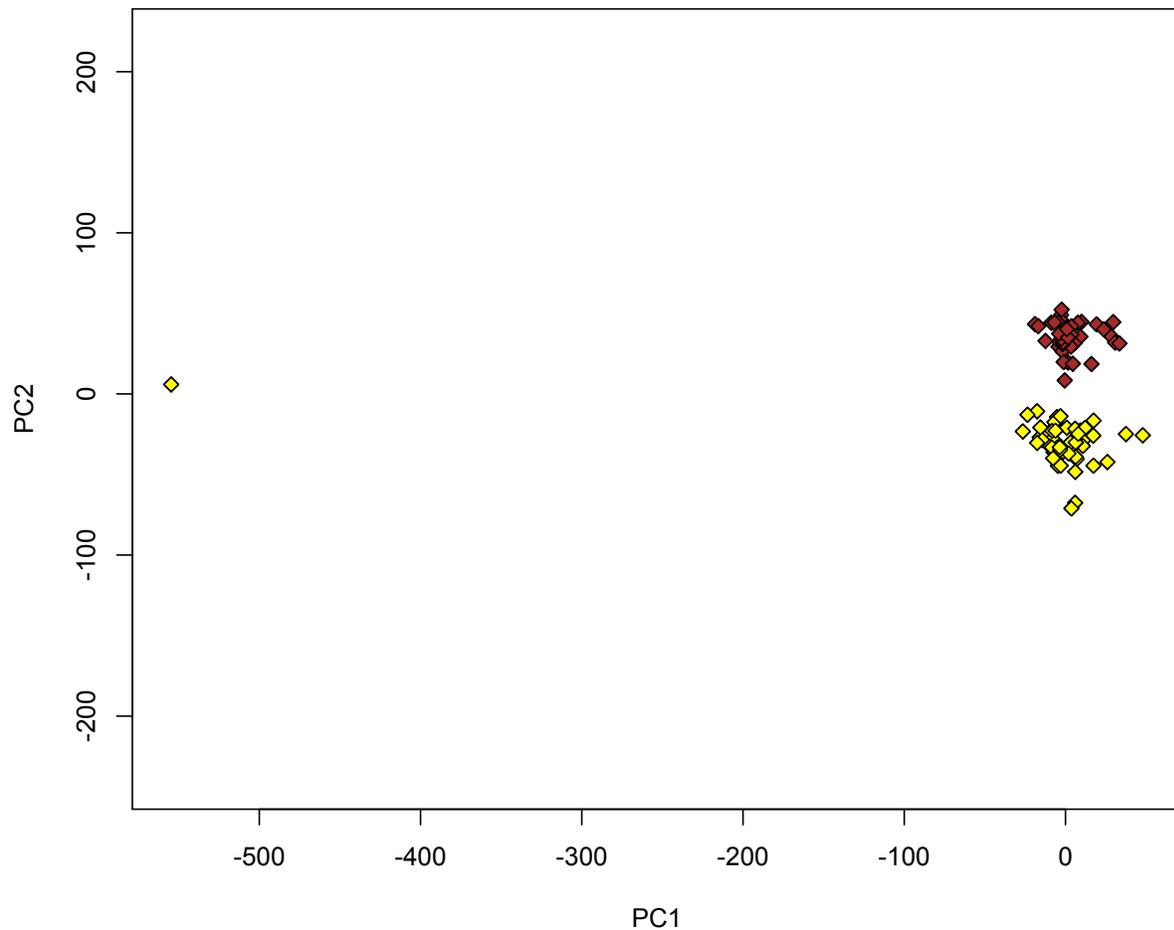
Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
UQRC1	12	9243581	9252902	9248241.5	Structural subunit of ETC complex III
NDUFAF3	12	12252246	12253315	12252780.5	Assembly factor/ancillary protein for ETC complex I
UQCRQ	13	1160234	1161216	1160725	Structural subunit of ETC complex III
MRPL22	13	5404744	5411508	5408126	Mitochondrial large ribosomal subunit protein
NDUFA2	13	15719950	15720141	15720045.5	Structural subunit of ETC complex I
HARS	13	16514578	16527881	16521229.5	aminoacyl-tRNA synthetase
NDUFB6	14	1409930	1411272	1410601	Structural subunit of ETC complex I
MRPL28	14	2948403	2980093	2964248	Mitochondrial large ribosomal subunit protein
NDUFAB1	14	8853879	8856325	8855102	Structural subunit of ETC complex I
NDUFB10	14	9479824	9481800	9480812	Structural subunit of ETC complex I
ATP5J2	14	11124929	11126597	11125763	Structural subunit of ETC complex V
COX19	14	13640400	13642700	13641550	Assembly factor/ancillary protein for ETC complex IV
UQCRC2	14	15520722	15531290	15526006	Structural subunit of ETC complex III
MRPL40	15	8155635	8160277	8157956	Mitochondrial large ribosomal subunit protein
COX6A	15	10968121	10969778	10968949.5	Structural subunit of ETC complex IV
UQCR10	15	12850007	12851340	12850673.5	Structural subunit of ETC complex III
MRPL41	17	1410903	1411829	1411366	Mitochondrial large ribosomal subunit protein
SURF1	17	7671080	7676756	7673918	Assembly factor/ancillary protein for ETC complex IV
NDUFA8	17	9921532	9923704	9922618	Structural subunit of ETC complex I

Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
MRPL12	18	1709716	1712644	1711180	Mitochondrial large ribosomal subunit protein
POLG2	18	3399365	3405672	3402518.5	Subunit of DNA polymerase gamma
COX10	18	3831022	3927582	3879302	Assembly factor/ancillary protein for ETC complex IV
SCO1	18	5465970	5471599	5468784.5	Assembly factor/ancillary protein for ETC complex IV
MRPL38	18	8209846	8218187	8214016.5	Mitochondrial large ribosomal subunit protein
ATP5H	18	8874306	8876875	8875590.5	Structural subunit of ETC complex V
MRPL27	18	9156395	9159570	9157982.5	Mitochondrial large ribosomal subunit protein
TTC19	19	7948708	7956750	7952729	Assembly factor/ancillary protein for ETC complex III
ATP5E	20	12072535	12073407	12072971	Structural subunit of ETC complex V
MRPL20	21	4190134	4191696	4190915	Mitochondrial large ribosomal subunit protein
NDUFS5	23	3948096	3948305	3948200.5	Structural subunit of ETC complex I
ATP5L	24	238825	240411	239618	Structural subunit of ETC complex V
FOXRED1	24	7573063	7578176	7575619.5	Assembly factor/ancillary protein for ETC complex I
TARS2	25	1192211	1197044	1194627.5	aminoacyl-tRNA synthetase
SARS	26	3075507	3081339	3078423	aminoacyl-tRNA synthetase
ATP5F1	26	3213285	3216396	3214840.5	Structural subunit of ETC complex V
ATP5G1	27	4567588	4568848	4568218	Structural subunit of ETC complex V
NDUFA11	28	138984	139260	139122	Structural subunit of ETC complex I
UQCR11	28	1073915	1076925	1075420	Structural subunit of ETC complex III

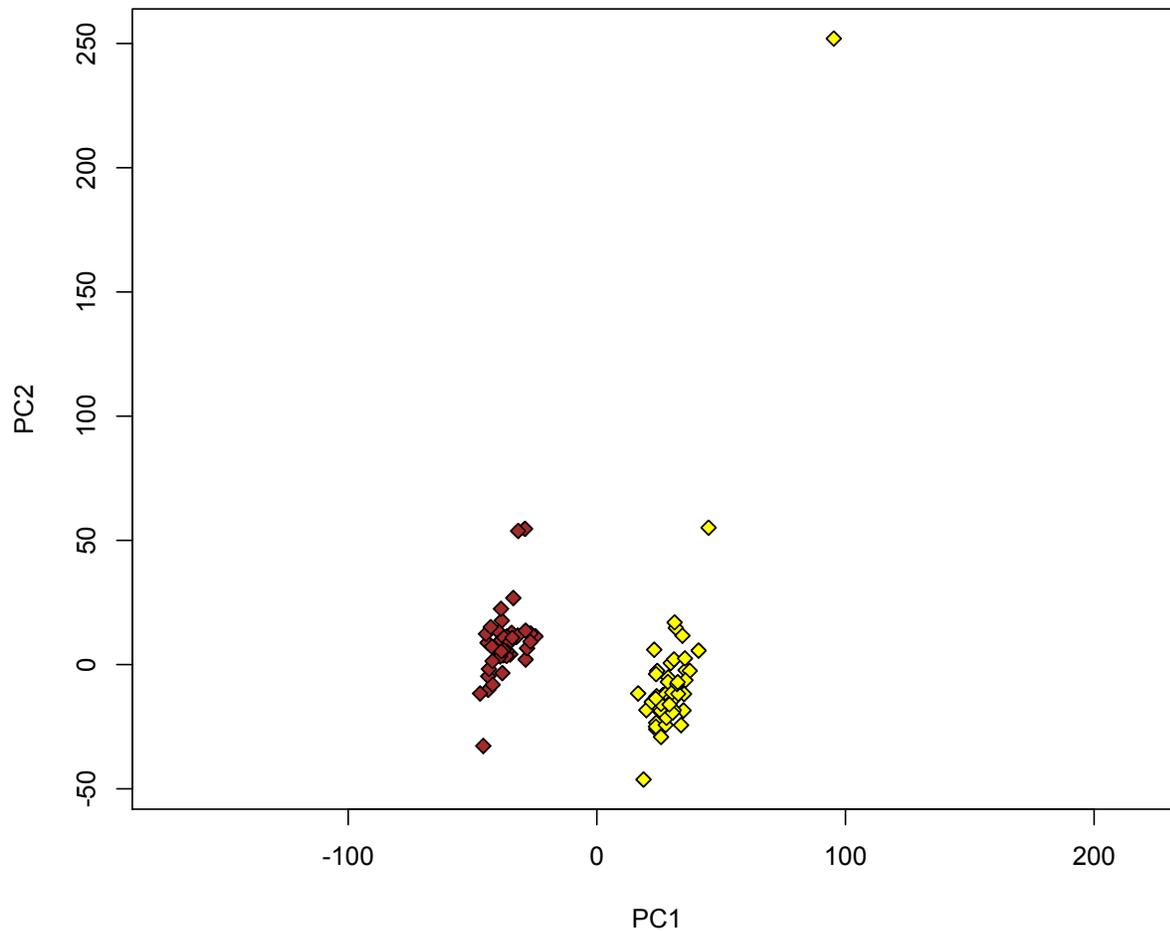
Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
POLRMT	28	2901392	2916002	2908697	Mitochondrial RNA polymerase
NDUFS7	28	4121690	4125088	4123389	Structural subunit of ETC complex I
ATP50D	28	4237453	4239005	4238229	Structural subunit of ETC complex V
ATP5C1	1A	3456418	3463670	3460044	Structural subunit of ETC complex V
NDUFA5	1A	21570348	21574483	21572415.5	Structural subunit of ETC complex I
MRPL42	1A	44154725	44161681	44158203	Mitochondrial large ribosomal subunit protein
NDUFA12	1A	44613892	44626365	44620128.5	Structural subunit of ETC complex I
NDUFA6	1A	48813487	48815331	48814409	Structural subunit of ETC complex I
mtSSB	1A	58878378	58881966	58880172	Single stranded DNA-binding protein
NDUFB2	1A	59314052	59315170	59314611	Structural subunit of ETC complex I
NDUFA9	1A	63597529	63609489	63603509	Structural subunit of ETC complex I
ATP5O	1B	248176	252126	250151	Structural subunit of ETC complex V
NDUFA1	4A	9441670	9442588	9442129	Structural subunit of ETC complex I
COX7B	4A	18172692	18174866	18173779	Structural subunit of ETC complex IV
ATP5I	Z	3439938	3442139	3441038.5	Structural subunit of ETC complex V
ATP5A1	Z	32785423	32792998	32789210.5	Structural subunit of ETC complex V
MRPL17	Z	40019570	40020454	40020012	Mitochondrial large ribosomal subunit protein
TARS	Z	40838333	40852658	40845495.5	aminoacyl-tRNA synthetase
NDUFS4	Z	46636290	46683735	46660012.5	Structural subunit of ETC complex I

Mitochondrial Gene	Chromosome	Start Position (bp)	End Position (bp)	Centre Position (bp)	Function
NDUFAF2	Z	49401620	49454084	49427852	Assembly factor/ancillary protein for ETC complex I
PTCD2	Z	66031974	66046558	66039266	Assembly factor/ancillary protein for ETC complex III
COX7C	Z	69702408	69704744	69703576	Structural subunit of ETC complex IV

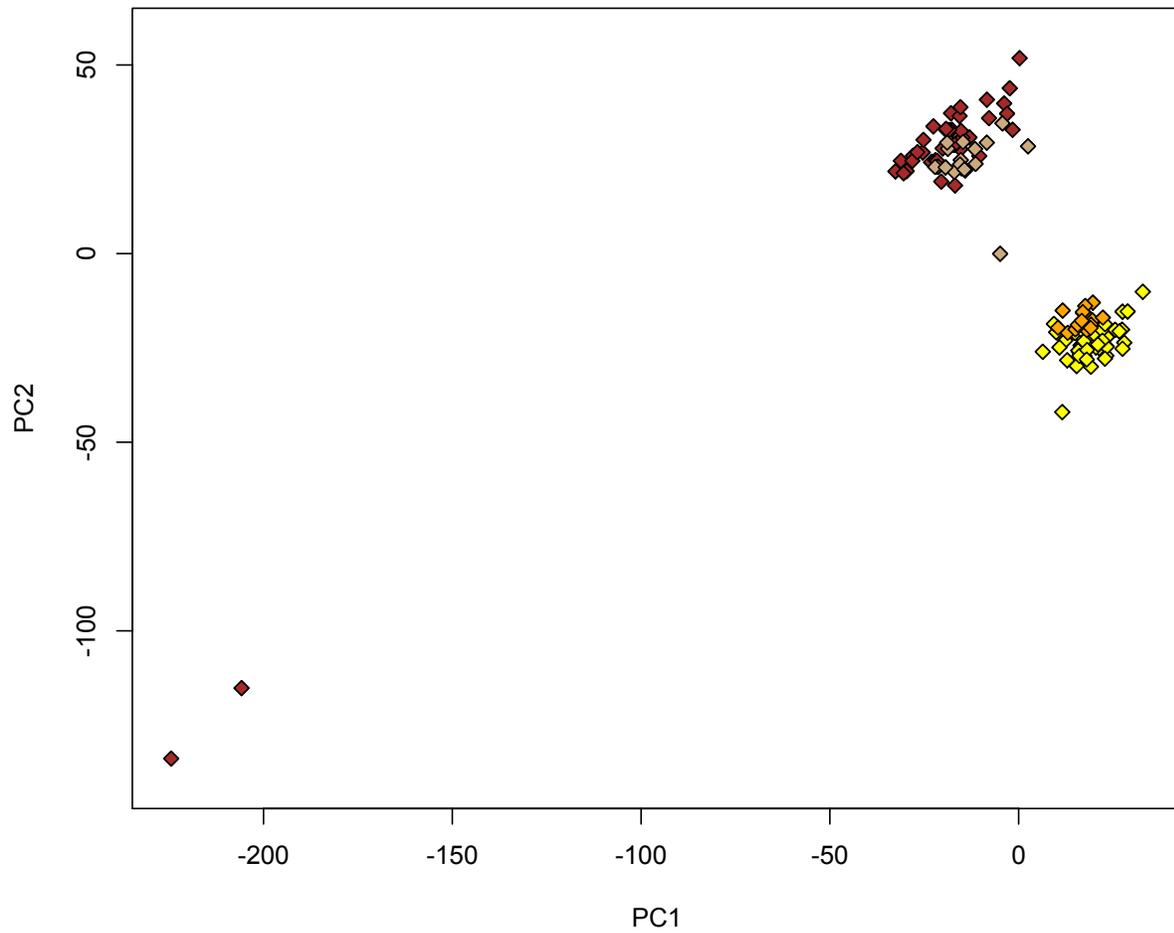
## A.2 Supplementary Figures



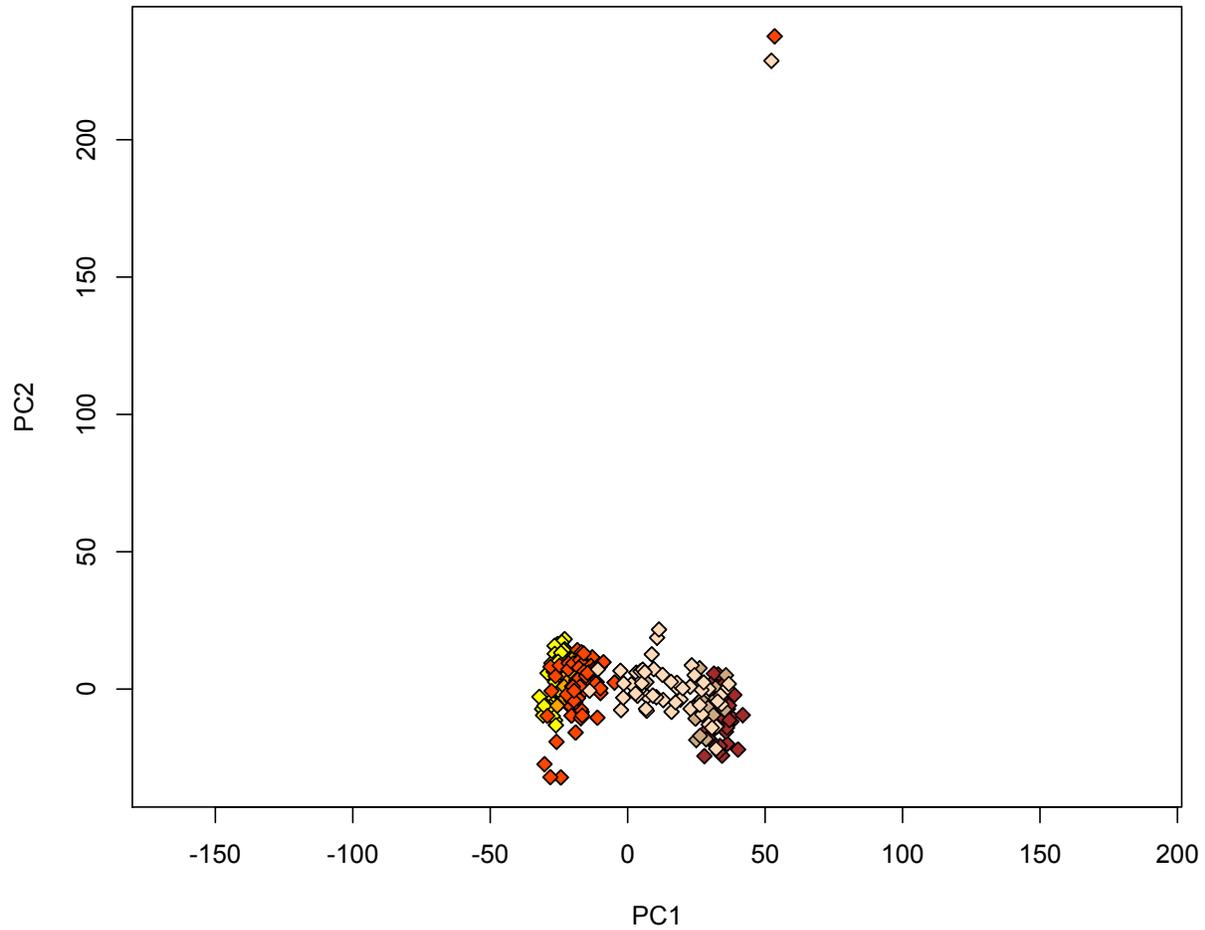
**Supplementary Figure 2.1.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 53$ ) and allopatric pine buntings (brown;  $n = 41$ ) following the removal of sample “*Emberiza\_GBS2\_ASR05\_49*”. This sample was one of a pair of outliers that appeared in the principal components analysis shown in Figure 1. PC1 explains 8.6% of the variation among individuals and PC2 explains 3.0% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.



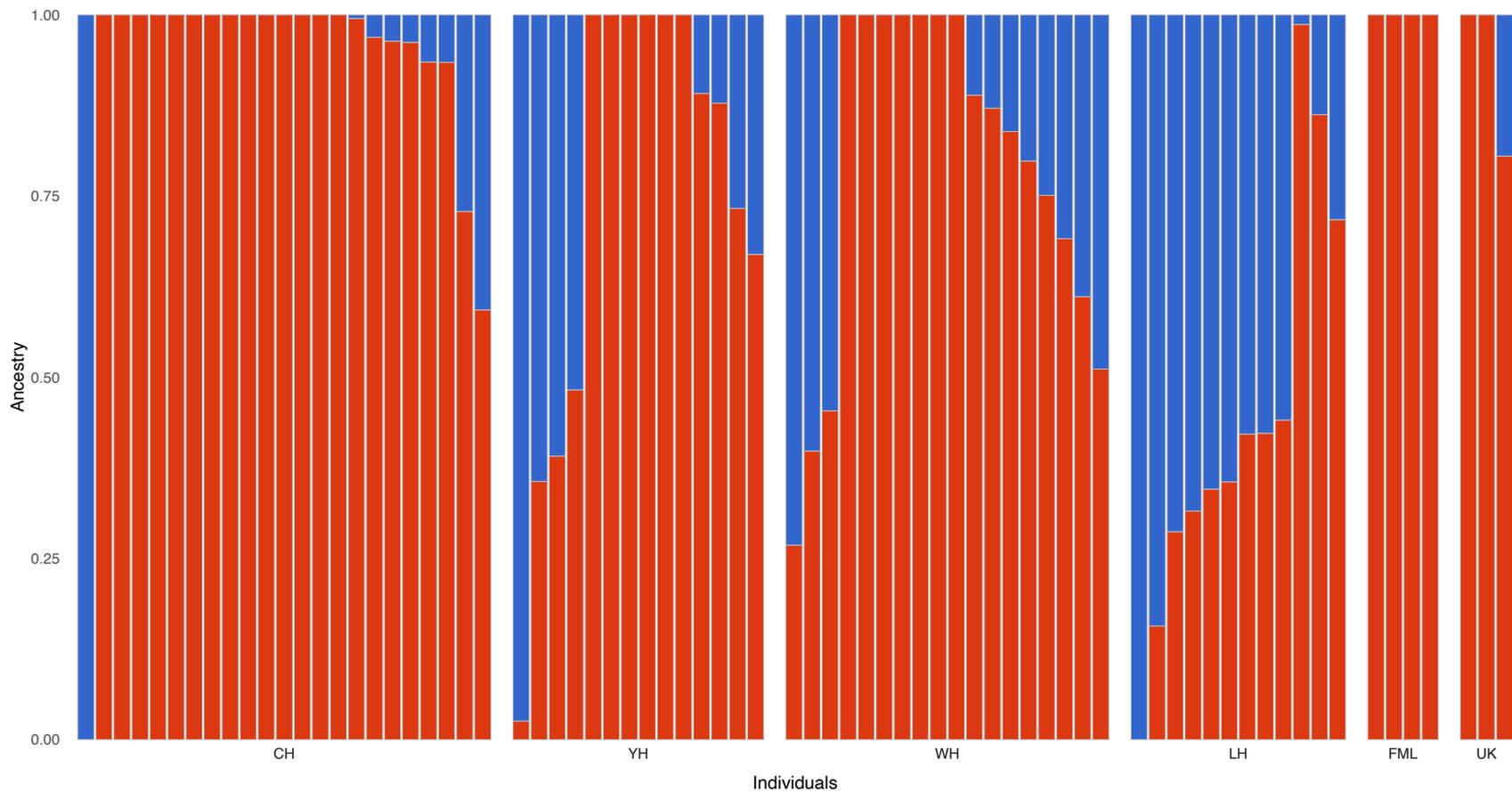
**Supplementary Figure 2.2.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 52$ ) and allopatric pine buntings (brown;  $n = 41$ ) following the removal of sample “*Emberiza\_GBS2\_ASR05\_49*” and “*Emberiza\_GBS2\_BKS\_1609*”. “*Emberiza\_GBS2\_ASR05\_49*” was one of a pair of outliers that appeared in the principal components analysis shown in Figure 1. “*Emberiza\_GBS2\_BKS\_1609*” was an outlier that appeared in the principal components analysis shown in Supplementary Figure S1. PC1 explains 3.2% of the variation among individuals and PC2 explains 2.6% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.



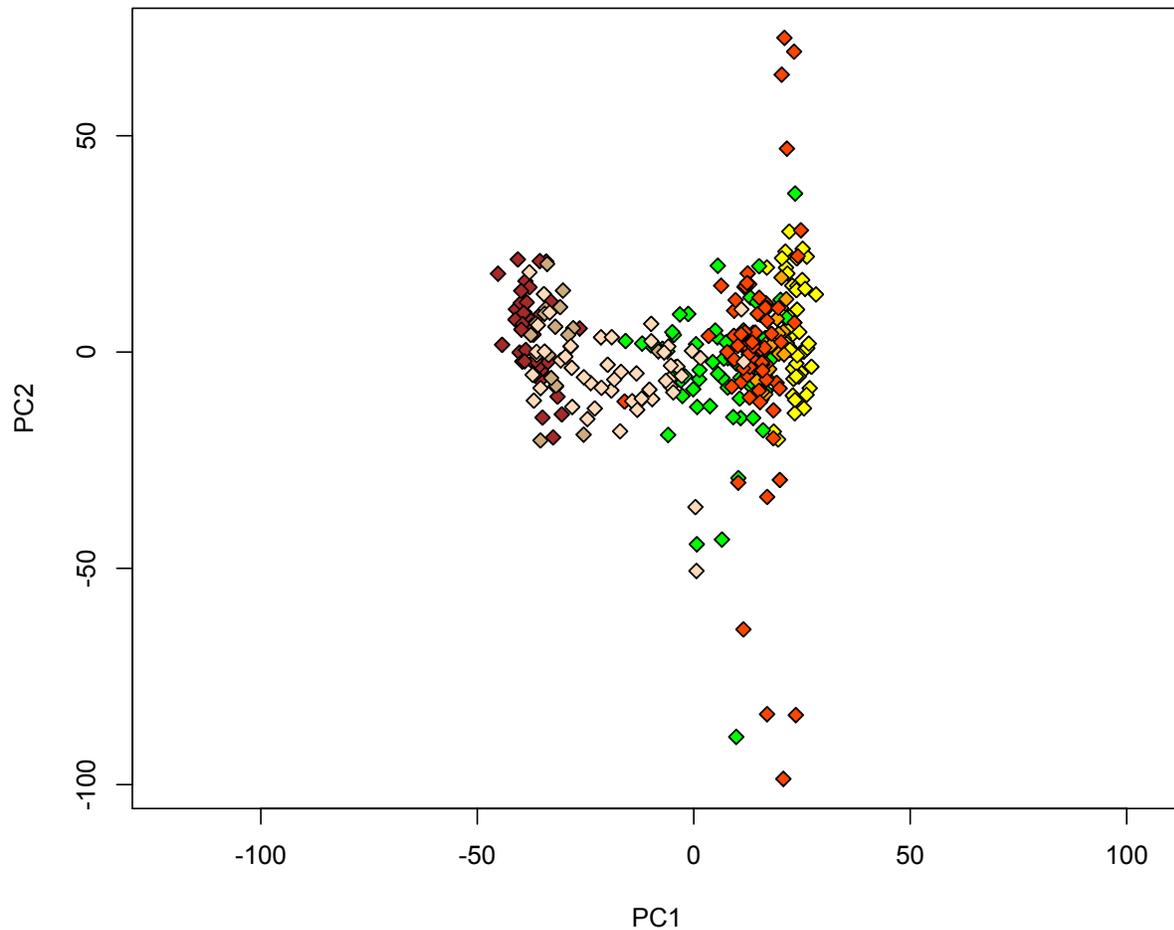
**Supplementary Figure 3.1.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 53$ ), near-sympatric yellowhammers (light orange;  $n = 15$ ), allopatric pine buntings (brown;  $n = 42$ ) and near-sympatric pine buntings (taupe;  $n = 18$ ). PC1 explains 2.9% of the variation among individuals and PC2 explains 2.5% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.



**Supplementary Figure 3.2.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 53$ ), near-sympatric yellowhammers (light orange;  $n = 15$ ), sympatric yellowhammers (red-orange;  $n = 67$ ), allopatric pine buntings (brown;  $n = 42$ ), near-sympatric pine buntings (taupe;  $n = 18$ ) and sympatric pine buntings (peach;  $n = 52$ ). PC1 explains 1.7% of the variation among individuals and PC2 explains 1.5% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.



**Supplementary Figure 3.3.** Ancestry proportions of hybrid individuals split into the phenotypic classes described in Rubtsov & Tarasov, 2017. Phenotypic classes include: citrinella hybrid (CH, n = 23), yellow hybrid (YH, n = 14), white hybrid (WH, n = 18), leucocephalos hybrid (LH, n = 12), female (FML, n = 4) and unknown class (UK, n = 3). Ancestry proportions were predicted by an Admixture model with K=2 and information from 417,164 SNPs.



**Supplementary Figure 3.4.** Whole-genome principal components analysis of allopatric yellowhammers (yellow;  $n = 53$ ), near-sympatric yellowhammers (light orange;  $n = 15$ ), sympatric yellowhammers (red-orange;  $n = 67$ ), allopatric pine buntings (brown;  $n = 42$ ), near-sympatric pine buntings (taupe;  $n = 18$ ), sympatric pine buntings (peach;  $n = 49$ ) and hybrids (green;  $n = 72$ ) following the removal of five outliers identified in Figure 1A: “*Emberiza\_GBS4\_XD\_632*”, “*Emberiza\_GBS4\_XD\_636*”, “*Emberiza\_GBS4\_XD\_639*”, “*Emberiza\_GBS4\_XD\_972*” and “*Emberiza\_GBS5\_XD\_970*”. PC1 explains 1.4% of the variation among individuals and PC2 explains 0.8% of the variation among individuals. Information from 374,780 SNPs was included in this analysis.