

ARTICLE



High heterogeneity in genomic differentiation between phenotypically divergent songbirds: a test of mitonuclear co-introgression

Ellen Nikelski ^{1,3}✉, Alexander S. Rubtsov² and Darren Irwin ¹

© The Author(s), under exclusive licence to The Genetics Society 2022

Comparisons of genomic variation among closely related species often show more differentiation in mitochondrial DNA (mtDNA) and sex chromosomes than in autosomes, a pattern expected due to the differing effective population sizes and evolutionary dynamics of these genomic components. Yet, introgression can cause species pairs to deviate dramatically from general differentiation trends. The yellowhammer (*Emberiza citrinella*) and pine bunting (*E. leucocephalos*) are hybridizing avian sister species that differ greatly in appearance and moderately in nuclear DNA, but that show no mtDNA differentiation. This discordance is best explained by adaptive mtDNA introgression—a process that can select for co-introgression at nuclear genes with mitochondrial functions (mitonuclear genes). To better understand these discordant differentiation patterns and characterize nuclear differentiation in this system, we investigated genome-wide differentiation between allopatric yellowhammers and pine buntings and compared it to what was seen previously in mtDNA. We found significant nuclear differentiation that was highly heterogeneous across the genome, with a particularly wide differentiation peak on the sex chromosome Z. We further investigated mitonuclear gene co-introgression between yellowhammers and pine buntings and found support for this process in the direction of pine buntings into yellowhammers. Genomic signals indicative of co-introgression were common in mitonuclear genes coding for subunits of the mitochondrion and electron transport chain complexes. Such introgression of mitochondrial DNA and mitonuclear genes provides a possible explanation for the patterns of high genomic heterogeneity in genomic differentiation seen among some species groups.

Heredity (2023) 130:1–13; <https://doi.org/10.1038/s41437-022-00580-8>

INTRODUCTION

Evolution in eukaryotes is shaped by changes in multiple genomic components that differ in their modes of inheritance: mitochondrial DNA (mtDNA) is usually inherited through the matrilineal line, autosomes are inherited through both parental lines and sex chromosomes are inherited differentially depending on the sex of both parent and offspring (Avice 2000). During speciation, populations of a single species diverge genetically, often in isolation, with the strength and pattern of genetic differentiation varying across the different genomic components (reviewed in Coyne and Orr 2004; reviewed in Price 2008). This variation arises due to differences in each component's rate of evolution as well as the degree to which each component contributes toward reproductive isolation and is resistant to gene flow between populations in secondary contact. Most commonly, speciating taxa will show clear differentiation in mtDNA (e.g., Hebert et al. 2004; Kerr et al. 2007), moderate differentiation in sex chromosomes (e.g., Thornton and Long 2002; Borge et al. 2005; Lu and Wu 2005; Harr 2006; Ruegg et al. 2014; Sackton et al. 2014), and comparatively modest differentiation across autosomes (Harr 2006; Nadeau et al. 2012; Irwin et al. 2018).

Mitochondrial DNA often shows strong differentiation between speciating populations (e.g., Hebert et al. 2004; Kerr et al. 2007). These patterns are partially driven by the mitochondrial genome's uniparental inheritance and haploid nature which decrease its effective population size to $\frac{1}{4}$ that of autosomal DNA under the assumptions of neutrality and an equal sex ratio (Moore 1995). Combined with a relatively high mutation rate (Lynch et al. 2006), this low effective population size leads to the rapid fixation of mutations via genetic drift and strong mtDNA differentiation between taxa. Additionally, a lack of recombination across the mitochondrial genome can further contribute to mtDNA divergence through rampant genetic hitchhiking (Hill 2020). Here, positive selection for an adaptive mtDNA mutation causes the fixation of genetic variants across the mitochondrial genome resulting in strong differentiation between speciating taxa.

Sex chromosomes—specifically the Z (Borge et al. 2005; Ruegg et al. 2014; Sackton et al. 2014) and X chromosomes (Thornton and Long 2002; Lu and Wu 2005; Harr 2006)—tend to show more moderate levels of genetic differentiation between speciating populations, but often much greater differentiation than autosomes. This trend is driven by two important characteristics of Z/X

¹Department of Zoology, and Biodiversity Research Centre, 6270 University Blvd., University of British Columbia, Vancouver, BC, Canada. ²State Darwin Museum, Moscow, Russia.

³Present address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada. Associate editor: Paul Sunnucks.

✉email: ellen.nikelski@mail.utoronto.ca

Received: 28 April 2022 Revised: 23 November 2022 Accepted: 23 November 2022

Published online: 3 December 2022

chromosomes. First, on Z/X chromosomes, beneficial recessive mutations are immediately exposed to selection in the heterogametic sex allowing these advantageous variants to fix more rapidly than they would on autosomes (Meisel and Connallon 2013; Irwin 2018). Second, because Z/X chromosomes are inherited as either one or two copies depending on the sex of the offspring, these chromosomes have a lower effective population size than autosomes (Mank et al. 2010; Irwin 2018). This lower effective population size allows for the fixation of a greater number of neutral and slightly deleterious mutations due to less effective purifying selection and a larger role of genetic drift. Working in tandem, these two characteristics drive genetic divergence of Z/X chromosomes between speciating populations in a phenomenon known as the “faster Z/X effect” (Meisel and Connallon 2013; Mank et al. 2010; Irwin 2018).

Across autosomes, genetic differentiation between speciating taxa tends to be modest compared to mtDNA and sex chromosomes and is often characterized by “islands of differentiation”—genomic regions of high relative differentiation that appear within a background of low relative differentiation (Harr 2006; Nadeau et al. 2012; Hejase et al. 2020). Explanations for these “islands” often invoke reduced gene flow between speciating taxa during secondary contact (Wu 2001) and/or repeated bouts of selection prior to and following secondary contact (Cruickshank and Hahn 2014; Irwin et al. 2018). In the former scenario, differentiation islands are hypothesized to house the loci responsible for reproductive barriers between taxa making them resistant to the homogenizing influence of gene flow (Wu 2001). In the latter scenario, differentiation islands are described as genomic areas that have experienced recurrent selection or selective sweeps (Cruickshank and Hahn 2014; Irwin et al. 2018). These events reduce genetic diversity first in the ancestral population and then in both daughter populations producing relative peaks in genetic differentiation between taxa.

An increasing number of studies have reported differentiation patterns that vary from the above description of what is usually seen between speciating populations (Irwin et al. 2009; Yannic et al. 2010; Bryson et al. 2010). In a number of taxonomically widespread cases, mtDNA shows dramatically low differentiation between taxa compared to nuclear DNA (Toews and Brelsford 2012) and observable phenotypes. One hypothesis that may explain these observations is introgression of mitochondrial haplotypes from one population into another following secondary contact and hybridization. Introgression of mtDNA can occur neutrally due to various processes such as sex-biased dispersal or hybrid zone movement (Toews and Brelsford 2012); however, the extreme degree of discordance observed between mitochondrial and nuclear differentiation in certain systems (e.g., Alves et al. 2008; Irwin et al. 2009) suggests that mtDNA introgression may be driven by selection and occur adaptively by providing a fitness advantage to individuals in the receiving population.

Introgression of foreign mtDNA could provide fitness advantages through two major avenues. First, because variation in mtDNA has been associated with variation in mitochondrial efficiency under different abiotic conditions (e.g., Ballard et al. 2007), mtDNA introgression may allow the receiving population to better adapt within a changing or novel environment (Hulsey et al. 2016; Sloan et al. 2017; Hill 2019a). For example, mtDNA introgression has been tentatively associated with thermal adaptation in rabbits (Alves et al. 2008) and cichlids (Hulsey et al. 2016). The second way that mtDNA introgression can provide a fitness advantage is by replacing a mitochondrial genome with a high mutational load (Sloan et al. 2017; Hill 2019a). As described earlier, the low effective population size (Moore 1995) and high mutation rate (Lynch et al. 2006) of mtDNA can lead to the rapid fixation of mutations, including deleterious mutations. This tendency combined with genetic hitchhiking of deleterious mutations (Hill 2020), can lead to a high mtDNA

mutational load which may decrease mitochondrial efficiency. Through introgression of a foreign mitochondrial haplotype with a lower mutational load, the receiving population may be able to regain greater mitochondrial function (Llopart et al. 2014; Hulsey et al. 2016; Sloan et al. 2017; Hill 2019a).

Adaptive mtDNA introgression presents a compelling hypothesis for the discordant differentiation patterns observed between some taxa (Irwin et al. 2009; Yannic et al. 2010; Bryson et al. 2010). This idea becomes even more interesting when we consider recent work that suggests strong coevolution between the mitochondrial and nuclear genomes (Hill 2019b). In most bilaterian animals, the mitochondrial genome consists of 37 genes. Because of this low gene content, mitochondrial processes are reliant on more than 1000 proteins encoded by “mitonuclear genes” across the nuclear genome (Calvo and Mootha 2010; Lotz et al. 2014). A few hundred of these mitonuclear proteins interact closely with products of the mitochondrial genome (i.e., RNA and proteins) or with the mitochondrial genome itself (Gershoni et al. 2009; Burton and Barreto 2012; Hill 2019b). Direct interactions between mitochondrial and mitonuclear products are necessary for mtDNA replication, transcription and translation (Diodato et al. 2014; Greber and Ban 2016) and for the formation of complexes I, III, IV and V in the electron transport chain (ETC; Hill 2019b) which performs oxidative phosphorylation in the mitochondrion (reviewed in Ernster and Schatz 1981). These associations between mitochondrial and mitonuclear products imply tight coevolution between the mitochondrial and nuclear genomes due to selection for mitochondrial efficiency: changes in one create selective pressure for compatible changes in the other (Gershoni et al. 2009; Burton and Barreto 2012; Hill 2019b).

Such coevolution suggests that hybridization and gene flow between speciating populations may be selected against due to recombination that separates coevolved mitochondrial and mitonuclear alleles exposing genetic incompatibilities (mitonuclear incompatibilities) within hybrid individuals. In this way, mitonuclear coevolution has the potential to contribute toward reproductive isolation and select against mtDNA introgression between taxa. Nevertheless, if the fitness advantages of mtDNA introgression outweigh the fitness disadvantages of mitonuclear incompatibilities, this process may still be able to occur. In that case, mtDNA introgression could select for co-introgression of coevolved mitonuclear alleles that together optimize mitochondrial function (Sloan et al. 2017; Hill 2019a). Evidence for mitonuclear co-introgression has been found in some systems (Beck et al. 2015; Morales et al. 2018; Wang et al. 2021), but the general significance of this process is still up for debate.

The yellowhammer (Passeriformes: Emberizidae: *Emberiza citrinella*) and pine bunting (*E. leucocephalos*) system (Fig. 1) is one that may capture the complex interplay between genetic differentiation and mitonuclear coevolution. Thought to have diverged in isolation during the Pleistocene glaciations (Irwin et al. 2009), this Eurasian avian sister pair is highly divergent in plumage and moderately divergent in song and ecology (Panov et al. 2003; Rubtsov and Tarasov 2017). Yet, despite their differences, yellowhammers and pine buntings hybridize extensively in a large and apparently expanding secondary contact zone in central and western Siberia (Panov et al. 2003, 2007; Rubtsov 2007; Irwin et al. 2009; Rubtsov and Tarasov 2017). Previous genomic work has identified discordant genetic differentiation patterns between allopatric yellowhammers and pine buntings (Irwin et al. 2009) as they are nearly identical in mtDNA but show moderate differentiation in nuclear AFLP (Amplified Fragment Length Polymorphism) markers. To explain these results, Irwin et al. (2009) suggested that mtDNA introgressed adaptively from one species into the other during a previous selective sweep, and this hypothesis was supported by several statistical tests performed on the mtDNA haplotype network. Alternative hypotheses of shared ancestral polymorphism or neutral introgression and mtDNA

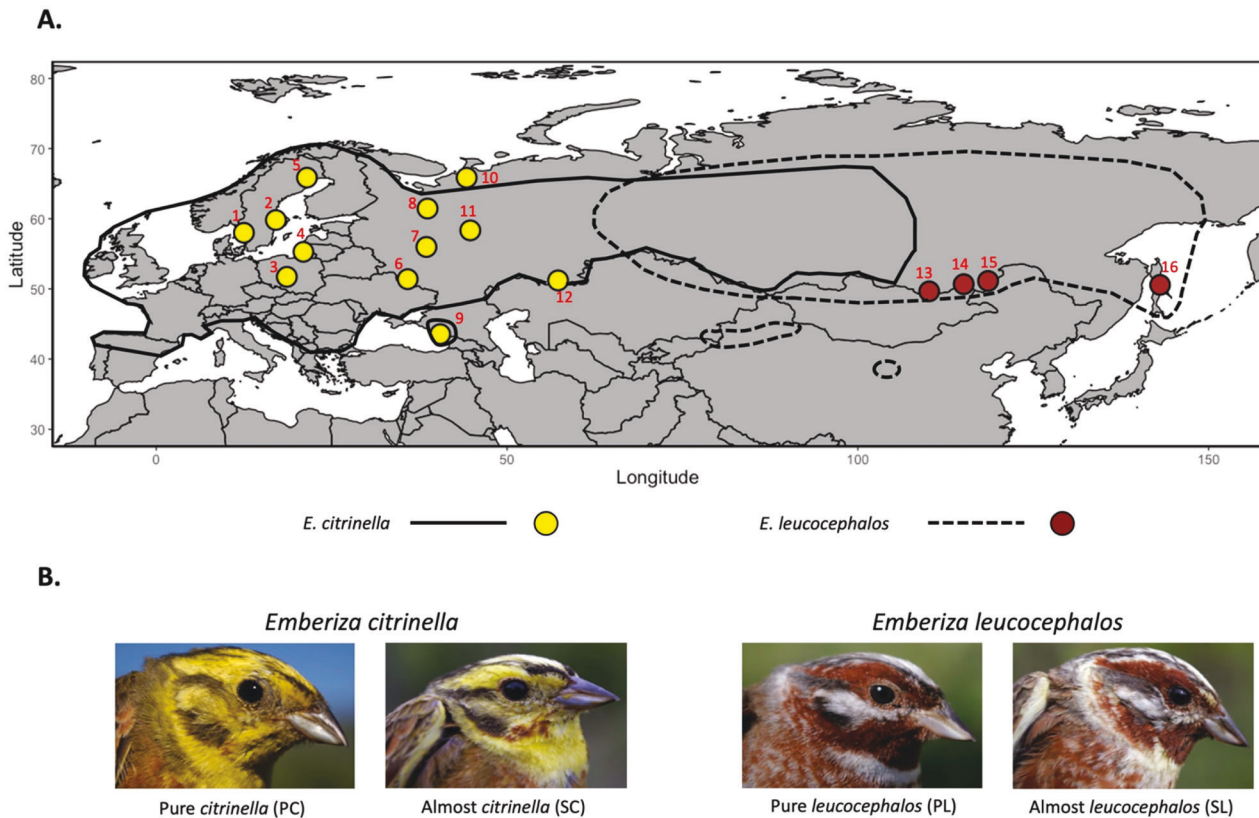


Fig. 1 Geographic distribution and phenotypic variation of sampled yellowhammers and pine buntings. **A** Map of sampling locations included in this study. Red numbers accompanying each location correspond to the sampling location numbers appearing in Table 1 which also describes sample sizes. Sampling locations may include multiple sites that appeared too close together to be shown in detail in this figure. Full details for the sites included in each sampling location can be found in Supplementary Table S1. Sampling location points are colored based on the taxon caught in each area: yellowhammer (*Emberiza citrinella*; yellow) and pine bunting (*Emberiza leucocephalos*; 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 of plumage variation between yellowhammers and pine buntings. Each photo represents one of four phenotypic classes: PC, SC, PL and SL. Individuals with a PC and SC phenotypic class were grouped together as *Emberiza citrinella* and individuals with a PL and SL phenotypic class were grouped together as *Emberiza leucocephalos*. All photos are credited to Dr. Alexander Rubtsov.

replacement were considered unlikely given the depth of differentiation observed in nuclear DNA compared to mtDNA, the inferred length of time the two taxa apparently evolved in allopatry, and the presently large population size. Based on mitonuclear theory, mtDNA introgression between yellowhammers and pine buntings could select for similar introgression at mitonuclear alleles to maintain mitochondrial function (Sloan et al. 2017; Hill 2019b). The resulting lack of mitonuclear incompatibilities between taxa as a result of co-introgression could facilitate their continued hybridization and hamper the build-up of reproductive barriers. With this opposition between the observed strong nuclear differentiation and the potential for mitonuclear co-introgression, the fate of the yellowhammer and pine bunting system remains uncertain. Depending on which way the scales tip, yellowhammers and pine buntings may continue to diverge and speciate or they may continue to hybridize and eventually collapse into one interbreeding population.

Here, we present a large-scale comparison of DNA sequence variation across the nuclear genomes of allopatric yellowhammers and pine buntings. With this data, we address key questions regarding genetic differentiation and mitonuclear coevolution in this system. First, what is the degree and structure of genetic differentiation between yellowhammers and pine buntings across the nuclear genome? Earlier AFLP analyses identified clear differentiation of nuclear markers between yellowhammers and pine buntings (Irwin et al. 2009), but those results were not based

on actual DNA sequences and only captured a small portion of the nuclear genome. Comparing patterns of differentiation across the nuclear genome enables better understanding of the extent of discordance between mitochondrial and nuclear differentiation and of which genomic regions show particularly high differentiation. Second, is there an over-representation of known mitonuclear genes within genomic regions showing genetic patterns suggestive of mitonuclear co-introgression? Support for mitonuclear co-introgression and a resulting lack of mitonuclear incompatibilities could explain the extensive hybridization seen between yellowhammers and pine buntings and implicate this process as a force that counters divergence and the evolution of strong reproductive barriers between groups. By answering these questions, we hope to provide insight on the evolutionary trajectory of yellowhammers and pine buntings (i.e., whether it is one of continued population divergence or of population merging) and also to explore how the interplay between genetic differentiation and mitonuclear coevolution influences the speciation process more generally.

MATERIALS AND METHODS

Sampling

We included 109 blood and tissue samples in this study: 53 phenotypic yellowhammers, 42 phenotypic pine buntings, and 14 other members of *Emberizidae* (one *Emberiza aureola* [yellow-breasted bunting], one

Table 1. Geographic locations and sample sizes of the sites included in this study.

Sampling Location	Latitude (°N)	Longitude (°E)	<i>E. citrinella</i> sample size	<i>E. leucocephalos</i> sample size
1	57.99	12.49	1	0
2	59.81	17.05	1	0
3	51.71	18.61	1	0
4	55.28	20.97	5	0
5	65.86	21.48	2	0
6	51.38	35.84	3	0
7	55.97	38.50	18	0
8	61.45	38.67	12	0
9	43.54	40.47	1	0
10	65.85	44.24	1	0
11	58.33	44.76	1	0
12	51.20	57.27	7	0
13	49.64	110.17	0	2
14	50.66	115.09	0	17
15	51.12	118.56	0	15
16	50.56	143.08	0	8
		Total	53	42

Sampling locations may include multiple sites that appeared too close together to be shown in detail in Fig. 1A. Full details for the sites included in each sampling locations can be found in Supplementary Table S1. The sampling location numbers that appear in the “Sampling Location” column correspond to those that appear in red in Fig. 1A. The “Sample Size” columns describes the total number of samples collected from a particular location.

Emberiza calandra [corn bunting], one *Emberiza cioides* [meadow bunting], one *Emberiza hortulana* [ortolan bunting], four *Emberiza stewarti* [white-capped bunting], and six *Emberiza cirius* [cirl bunting]) to put variation between yellowhammers and pine buntings into a deeper phylogenetic context (Fig. 1A; Table 1; Supplementary Table S1). These samples were part of a larger dataset that included individuals collected within and near the yellowhammer and pine bunting sympatric zone in western and central Siberia (Fig. 1A). All individuals examined in the present study (referred to as “allopatric”) were found at least 400 km from the border of the sympatric zone to account for recent and continued expansion of this region (Panov et al. 2003, 2007; Rubtsov 2007; Irwin et al. 2009; Rubtsov and Tarasov 2017). All male birds included in this study had clear phenotypes consistent with these phenotypic classes described by Rubtsov and Tarasov (2017): yellowhammers of the pure *citrinella* (PC) or semi-*citrinella* (SC) phenotypes; and pine buntings of the pure *leucocephalos* (PL) or semi-*leucocephalos* (SL) phenotypes (Fig. 1B). Because females are phenotypically similar between species, we did not put them into phenotypic classes. A total of 91 of our samples were included in the AFLP analysis of Irwin et al. (2009) while 18 samples were examined for the first time as part of the present research.

DNA extraction and genotyping-by-sequencing

DNA was extracted from samples using a standard phenol-chloroform method. We then divided the DNA samples into four genotyping-by-sequencing (GBS) libraries (Elshire et al. 2011). The 109 samples included in this study were sequenced together with 226 yellowhammer, pine bunting and hybrid samples collected near and within the sympatric zone (Nikelski et al. in prep). The libraries were prepared as per the protocol described by Alcaide et al. (2014) with modifications specified by Geraldine et al. (2019) except that we maintained a 300–400 bp DNA fragment size during size selection. Paired-end sequencing was completed by Genome Québec using an Illumina HiSeq 4000 system, producing more than 1.2 billion reads, each 150 bp in length, across the four GBS libraries.

Genotyping-by-sequencing data filtering

We processed the reads following Irwin et al. (2016; 2018), as summarized here. Reads were demultiplexed using a custom perl script designed by Baute et al. (2016). Next, 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 and Durbin

2009) and a BAM file of this information was created for each individual using the programs Picard (<http://broadinstitute.github.io/picard/>) and SAMtools (Li et al. 2009). The zebra finch reference genome was utilized for its comprehensive gene annotation which was essential to conducting our mitonuclear analysis discussed below. BAM files were converted into GVCF files using the HaplotypeCaller command as part of GATK version 3.8 (McKenna et al. 2010). We then combined information from the individuals to create (1) a genome-wide “variant site” VCF file containing only variant site information with sufficient coverage, and (2) a series of chromosome-specific “all sites” VCF files which contained information on both variant and invariant sites with sufficient coverage.

To create the genome-wide “variant site” VCF file, we used the GenotypeGVCFs command in GATK to identify single nucleotide polymorphisms (SNPs) among the 336 individuals included in the pipeline. This command also converted the variant site information into a single VCF file of SNP information encompassing the entire nuclear genome. Using a combination of VCFtools (Danecek et al. 2011) and GATK, we filtered the VCF file to remove indels and non-biallelic SNPs. To avoid erroneous SNP calls based on poorly mapped reads, we discarded loci with $QD < 2.0$, $MQ < 40.0$, $FS > 60.0$, $SOR > 3.0$, or $ReadPosRankSum < -8.0$. Finally, loci with more than 60% missing genotypes were removed. The average coverage of variable sites in the resulting VCF file was 16.59.

To convert GVCF files into “all sites” VCF files, we similarly employed the GenotypeGVCFs command in GATK with the addition of the -allSites and -L flags to retain invariant sites and split the information into chromosome-specific files respectively. The resulting VCF files were filtered using VCFtools and GATK to remove indels, sites with more than two alleles, sites with more than 60% missing data, sites with MQ values lower than 20 and sites with heterozygosities $> 60\%$ (to avoid potential paralogs). Some of these filters were more permissive than those applied to the “variant site” file in order to increase the resolution of windowed statistics whose calculation are discussed below.

Variant site analyses

The genome-wide “variant site” VCF file was analyzed using modified versions of the R scripts described in Irwin et al. (2018), and all of our analyses used R version 3.6.2 (R Core Team 2014). A total of 374,780 SNPs were identified among allopatric yellowhammers and pine buntings. For each of these SNPs, we calculated sample size, allele frequency, and Weir and Cockerham’s F_{ST} (Weir and Cockerham 1984). Genetic differentiation between yellowhammers and pine buntings was then visualized using a principal components analysis (PCA) generated with the pca command

and the svtImpute method to account for any missing genomic data using the pcaMethods package (Stacklies et al. 2007). A Manhattan plot of F_{ST} values for 349,807 SNPs identified among allopatric yellowhammers and pine buntings with known genomic locations was created using the package qqman (Turner 2018).

Differentiation across the genome

To thoroughly investigate nuclear differentiation between allopatric yellowhammers and pine buntings, we performed further analyses on both variant and invariant loci within “all sites” VCF files using R scripts described in Irwin et al. (2018). We calculated Weir and Cockerham’s F_{ST} and between-group nucleotide differentiation (π_B) comparing allopatric yellowhammers and pine buntings as well as within-group nucleotide variation (π_W) for each population. All statistics were calculated within nonoverlapping windows of available sequence data across each chromosome. The first window was positioned at the “start” of each chromosome as described in the zebra finch reference genome (Warren et al. 2010) and each window contained 2000 bp of sequence data which included multiple small sequenced regions separated by unsequenced regions. We used a window size of 2000 bp of sequenced data rather than 10,000 bp (as in Irwin et al. 2018), to visualize narrow peaks in relative and absolute differentiation within our dataset. We hereafter refer to these windows as “genomic windows.”

We developed a new R script to calculate a Tajima’s D value (Tajima 1989) for each of the genomic windows. Values of Tajima’s D were used to identify areas of the genome where patterns of variation in yellowhammer and pine bunting populations deviated from a neutral model. Significantly negative Tajima’s D implies that the ratio of common versus rare alleles is lower than expected under neutrality, likely because of a selective sweep or population expansion following a bottleneck. Significantly positive Tajima’s D indicates that the ratio of common versus rare alleles is higher than expected under neutrality, potentially stemming from balancing selection or a rapid population contraction.

Phylogenetic comparison with other Emberizidae species

We employed whole-genome averages of π_B between allopatric yellowhammers and pine buntings as well as among these focal species and six other Emberizidae species (*Emberiza aureola*, *Emberiza calandra*, *Emberiza cioides*, *Emberiza cirius*, *Emberiza hortulana* and *Emberiza stewarti*) to estimate a phylogeny. A list of average π_B values for each species pair was converted into a distance matrix and used to create an unrooted neighbor-joining tree. This tree was constructed using the ape package (Paradis and Schliep 2019) and the BioNJ algorithm (Gascuel 1997) with *Emberiza aureola* set as the outgroup (Alström et al. 2008). The phylogeny was created to provide further support for the sister relationship between yellowhammers and pine buntings that had previously been hypothesized using mitochondrial markers (Alström et al. 2008; Irwin et al. 2009) but that was questioned in some studies (Rubtsov and Opaev 2012). In creating this phylogeny, we were also able to investigate discordance between mitochondrial and nuclear differentiation within the system using a greater amount of nuclear data.

Signals of mitonuclear co-introgression

To investigate potential mitonuclear co-introgression between allopatric yellowhammers and pine buntings, we asked whether there was an association between a list of known mitonuclear genes and a list of genomic windows with statistical characteristics suggestive of introgression. Due to a lack of resolution in our genomic dataset, we were unable to perform statistical tests to directly characterize regions of introgression between taxa and, instead, identified genomic windows with low π_B and low Tajima’s D in each taxon. Low π_B indicates high similarity between the nucleotide sequences of allopatric populations, which would be expected if alleles introgressed from one taxon into the other. Low Tajima’s D suggests a past selective sweep within a population which would be expected if an adaptive allele introgressed from a separate taxon and swept throughout the receiving population. We call these windows “Between Population Putative Sweep Windows” (BPPSWs). Our quantitative criteria for a BPPSW were a Tajima’s D value within the lowest 5% of the available windowed values and a π_B value within the lowest 30% of the available windowed values. The Tajima’s D threshold was kept relatively low to capture windows with particularly strong signals of selection that could be associated with an adaptive selective sweep and introgression between taxa while the π_B threshold was left higher to identify an appreciable number of BPPSWs for analysis.

The mitonuclear genes that we investigated in our co-introgression analysis were all protein coding with products that interact directly with mtDNA or an immediate product of the mitochondrial genome. Theory predicts that these genes experience strong mitonuclear coevolution and that changes in mtDNA caused by introgression would drive selection for co-introgression of compatible alleles (Gershoni et al. 2009; Burton and Barreto 2012; Hill 2019b). By consulting various resources on mitochondrial and mitonuclear dynamics, we produced a list of these mitonuclear genes that encode: protein subunits of complexes I, III, IV, and V of the ETC, assembly and ancillary proteins involved in the formation of these ETC complexes, and proteins that are part of the transcription, translation or DNA replication machinery within mitochondria (Diodato et al. 2014; Greber and Ban 2016; Hill 2019b). 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 162 mitonuclear genes remained for analysis (Supplementary Table S2). Using a custom R script, we assigned each mitonuclear gene to the genomic window that minimized the absolute difference between the location of the mitonuclear gene centre and the location of the window centre. We then determined the number of mitonuclear genes that occurred within the BPPSWs identified for each taxon.

We conducted a Fisher’s Exact test for both yellowhammers and pine buntings to determine whether the proportion of mitonuclear genes within BPPSWs was significantly different from what would be expected based on the total proportion of protein coding genes appearing within these windows. To complete this analysis, a list of the 14,008 protein-coding genes annotated in the zebra finch reference genome (Warren et al. 2010) was compiled. We removed our 162 mitonuclear genes from this list and assigned the remaining 13,846 non-mitonuclear genes to genomic windows using the methodology described above. Because previous research has reported that mitonuclear genes are non-randomly distributed between autosomes and sex chromosomes in some systems (Drown et al. 2012; Dean et al. 2014)—a situation that could bias our mitonuclear analysis—we investigated whether the mitonuclear genes we examined were over- or under-represented on the Z chromosome. A Fisher’s Exact showed no significant difference in the proportions of mitonuclear versus non-mitonuclear genes positioned on chromosome Z versus on autosomes (p value = 0.7182). Following this non-significant result, we proceeded with comparing the proportion of mitonuclear and non-mitonuclear genes appearing in yellowhammer and pine bunting BPPSWs.

Following the completion of the above analyses, we performed additional analyses to determine whether windows containing mitonuclear genes differed significantly in average F_{ST} , π_B , π_W and Tajima’s D from windows not containing mitonuclear genes. Average values were calculated based on 155 genomic windows with mitonuclear genes and 7032 genomic windows without. A series of Welch’s t -tests were used to compare these values with results shown in detail in Supplementary Table S3.

RESULTS

When comparing allopatric yellowhammers and pine buntings, we identified 374,780 variable SNPs within our “variant site” VCF file and 13,703,455 invariant and 699,122 variant sites across thirty autosomes and the Z chromosome within our “all sites” VCF files (the different numbers of variable sites result from differences in filtering parameters; see Methods). In the latter “all sites” files, we designated a total of 7187 genomic windows (of 2000 sequenced bp each) across the genome, with each window covering an average distance of about 139 kilobases (with the reference genome being 1.0 Gb in length).

Phylogenetic comparison with other Emberizidae species

An unrooted neighbor-joining tree of average π_B values between yellowhammers, pine buntings and six other Emberizidae species (Fig. 2) depicted similar species relationships as estimated previously using mitochondrial markers (Alström et al. 2008; Irwin et al. 2009). Relative branch lengths were also similar, with the major exception being the branch length between yellowhammers and pine buntings which was much longer in our analysis using nuclear DNA. To put this into context, we determined the relative genetic distance between yellowhammers

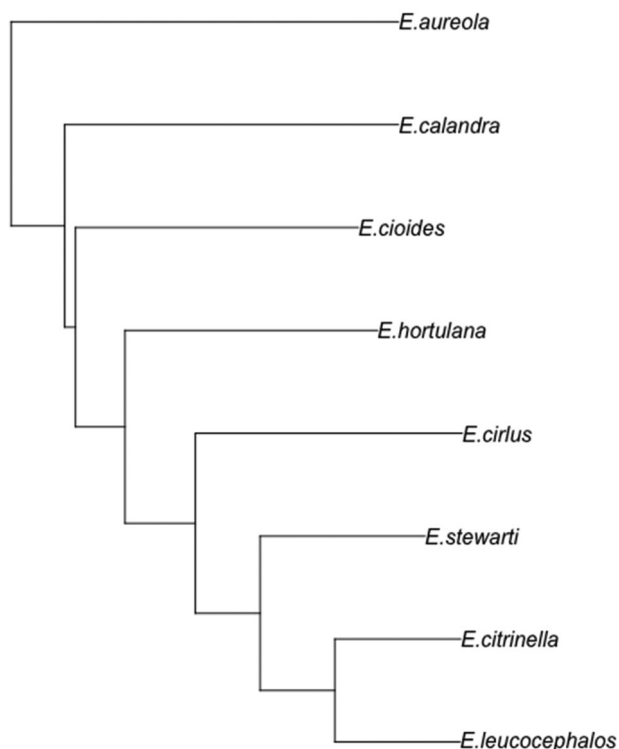


Fig. 2 Unrooted neighbor-joining tree of Emberizidae species constructed based on average absolute between-population nucleotide diversity (π_B). Sample sizes for each species are as follows: *E. aureola* = 1, *E. calandra* = 1, *E. cioides* = 1, *E. hortulana* = 1, *E. cirius* = 6, *E. stewarti* = 4, *E. citrinella* = 53 and *E. leucocephalos* = 42.

and pine buntings versus the genetic distance between *E. stewarti* and either member of the yellowhammer/pine bunting clade for our nuclear phylogeny and for the previously calculated mitochondrial phylogeny (Irwin et al. 2009). The nuclear ratio was 11.4 times greater than the mitochondrial ratio which corroborates the presence of strong discordance in genetic differentiation between yellowhammers and pine buntings. These results also support the hypothesis of an extended period of divergence between yellowhammers and pine buntings followed by adaptive mtDNA introgression.

Overall genetic differentiation

Based on 374,780 SNPs, the genome-wide F_{ST} estimate was 0.0232 between allopatric yellowhammers and pine buntings. Despite this low average, a PCA based on the same SNP genotypes separated yellowhammers and pine buntings into tight genetic clusters (Fig. 3). Two pine buntings were outliers along PC1, while the remaining yellowhammers and pine buntings separated into distinct groups mainly along PC2. Further investigation into these outliers revealed that they were males from the same location, but a kinship analysis completed as part of a separate study did not find close kinship between the two pine buntings that could explain their position (Nikelski et al. in prep). We also examined the PC1 loadings and found that the signal for PC1 position was broadly distributed across the nuclear genome, rather than being concentrated in a few highly influential regions (Supplementary Fig. S1). We then temporarily removed one of the outliers and re-ran the PCA. This caused the other outlier to fall into the pine bunting cluster, but revealed a further yellowhammer outlier (Supplementary Fig. S2). Removal of this yellowhammer outlier in addition to one member of the pine bunting outlier pair in turn revealed another yellowhammer outlier (Supplementary Fig. S3). It is unclear what is responsible for these outliers, but the distinct

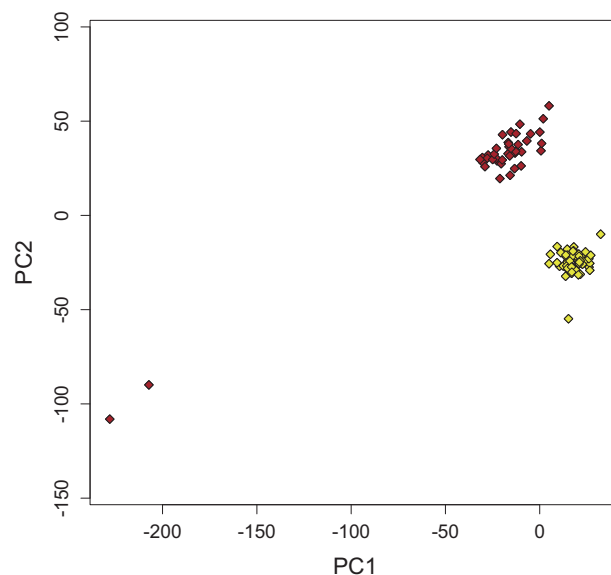


Fig. 3 PCA of genetic variation between allopatric yellowhammers (yellow; $n = 53$) and allopatric pine buntings (brown; $n = 42$), based on 374,780 genome-wide SNPs. PC1 and PC2 explain 3.6% and 2.9%, respectively, of the variation among individuals.

yellowhammer and pine bunting genetic clusters remained intact in all the PCAs considered.

Differentiation across the genome

Relative differentiation between allopatric yellowhammers and pine buntings was highly heterogeneous across the nuclear genome with peaks in F_{ST} seen on most of the larger chromosomes (Figs. 4, 5, 6A; Supplementary Figs. S4, S5). Chromosome Z in particular showed a large peak in F_{ST} with several SNPs possessing values close to one. In fact, F_{ST} for the Z chromosome was 0.1246—more than five times larger than the genome-wide F_{ST} .

Patterns of between-group nucleotide differentiation (π_B) and within-group nucleotide variation (π_W) were also heterogeneous across the genome and comparable to each other in magnitude: genome-wide $\pi_B = 0.0041$; genome-wide π_W for both taxa = 0.0040 (Fig. 5; Supplementary Fig. S4). 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 fact, most windowed π_B and π_W averages fell near a 1:1 association line (Spearman's Rank Correlation: 0.9943, $p < 2.2 \times 10^{-16}$; Fig. 6B) which is equivalent to no or little differentiation. However, some genomic windows showed much reduced π_W compared to π_B ; these were the windows with high F_{ST} . In addition, we detected a weak but highly significant negative correlation between the windowed averages of F_{ST} and π_B (Spearman's Rank Correlation: -0.1196 , $p < 2.2 \times 10^{-16}$; Fig. 6C) as is hypothesized if peaks in relative differentiation are products of repeated selective events (Cruickshank and Hahn 2014; Irwin et al. 2018).

Finally, we found that Tajima's D varied across the genome but was mostly negative (Fig. 5; Supplementary Fig. S4), consistent with a history of population growth and/or selective sweeps. The average genome-wide Tajima's D was similar between populations: -1.377 for yellowhammers and -1.335 for pine buntings.

Signals of mitonuclear co-introgression

Of the 7187 genomic windows identified across the nuclear genome, we classified 244 (3.4%) as BPPSWs within yellowhammers

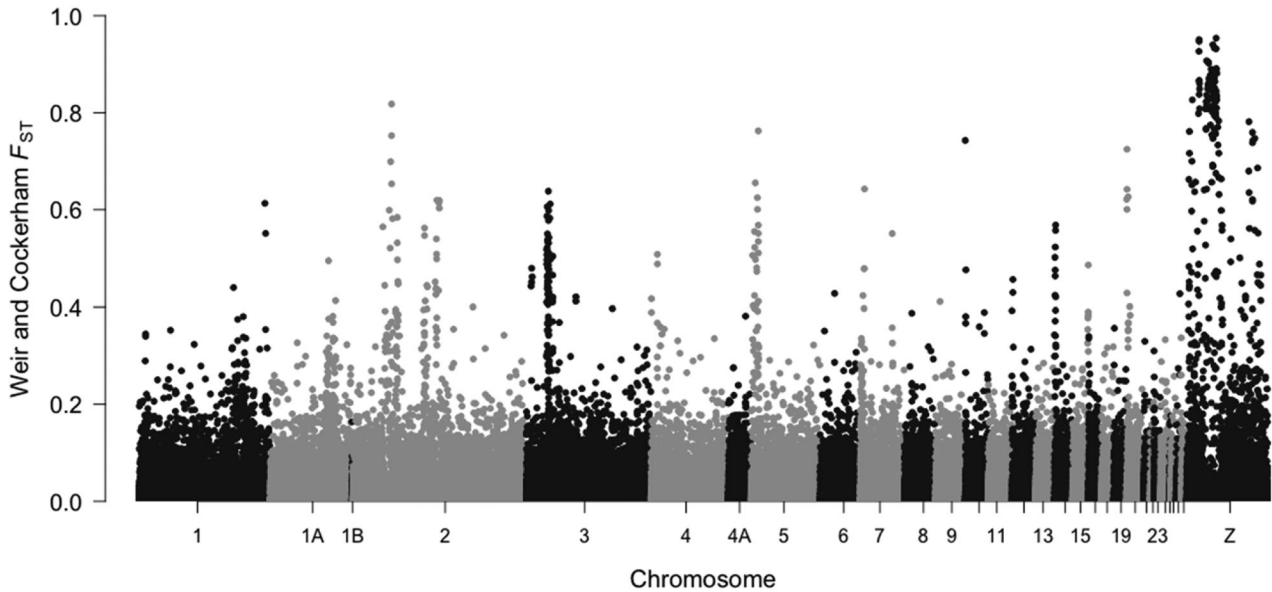


Fig. 4 Relative differentiation (F_{ST}) of 349,807 genome-wide SNPs identified among allopatric yellowhammers ($n = 53$) and allopatric pine buntings ($n = 42$), with chromosomes represented with alternating black and gray. Narrow regions of elevated differentiation can be seen on many autosomes, and there are broad regions of high differentiation on the Z chromosome.

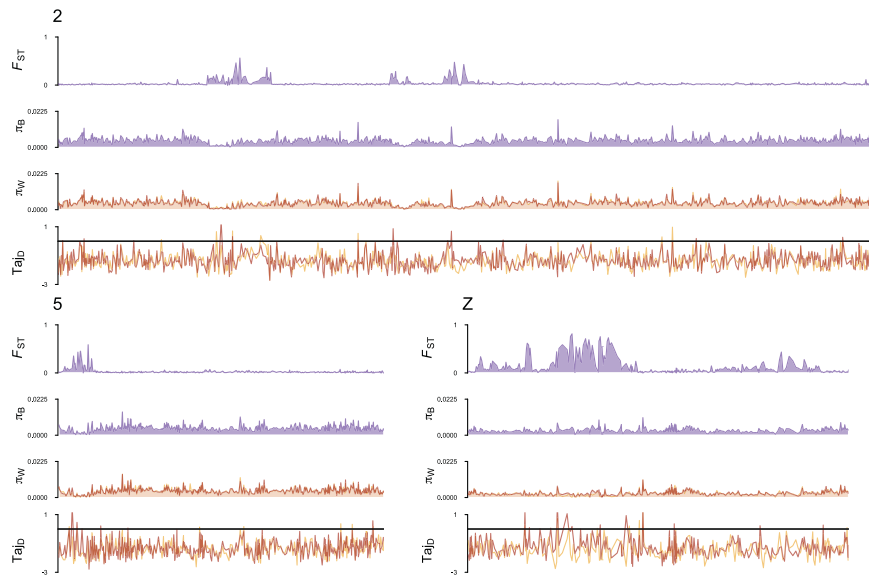


Fig. 5 Patterns of genetic variation comparing allopatric yellowhammers ($n = 53$) and allopatric pine buntings ($n = 42$) across chromosomes 2, 5 and Z shown as examples of general genomic patterns in this system (similar plots for all chromosomes can be found in Supplementary Fig. S4). Relative nucleotide differentiation (F_{ST}), absolute between-population nucleotide differentiation (π_B), absolute within-population nucleotide variation (π_W) and Tajima's D (Taj_D) are shown as 2000 bp windowed averages across each chromosome. F_{ST} and π_B are shown as purple lines to indicate that values were calculated as a comparison between allopatric yellowhammers and pine buntings. π_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.

and 222 (3.1%) as BPPSWs within pine buntings. Average values of π_B and Tajima's D in yellowhammer BPPSWs were 0.0016 and -2.3751 respectively, and 0.0019 and -2.3369 in pine bunting BPPSWs respectively. In non-BPPSWs, the average values of π_B and Tajima's D were 0.0042 and -1.3416 in yellowhammers and 0.0042 and -1.3031 in pine buntings. Of the BPPSWs identified in yellowhammer and pine bunting populations, 71 were shared between the taxa. It should be noted that sharing of some BPPSWs is expected given that the contribution of π_B to window selection was identical for both taxa (in contrast, Tajima's D was calculated separately for yellowhammers and pine buntings).

Our examination of the gene content within yellowhammer BPPSWs revealed that they contained a higher percentage (7.4%) of mitonuclear genes (12 of the 162 genes considered) than of non-mitonuclear genes (4.1%; 574 of the 13,846 genes considered). This difference was statistically significant (Fisher's Exact test: $p = 0.04714$), providing evidence for mitonuclear genes preferentially appearing within yellowhammer BPPSWs. Pine bunting BPPSWs contained 4.3% of the mitonuclear genes (7 of the 162 genes considered) and 3.3% of the non-mitonuclear genes (455 of the 13,846 genes considered), a difference that was not statistically significant (Fisher's Exact test: $p = 0.3806$).

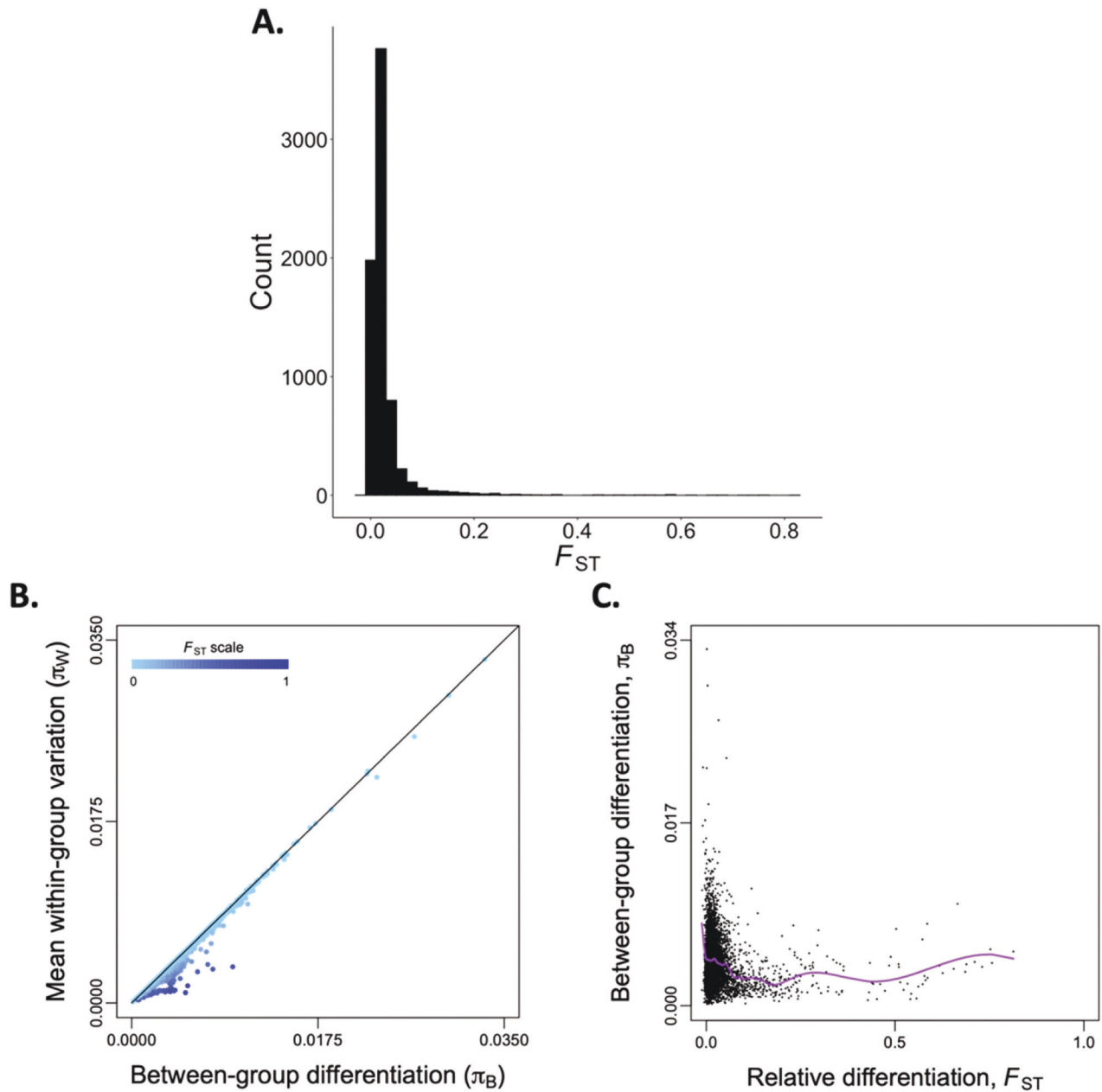


Fig. 6 **A** summary of genetic statistics calculated for allopatric populations of yellowhammers and pine buntings. **A** A histogram of average relative differentiation (F_{ST}) values calculated for windows of 2000 sequenced nucleotides across the nuclear genome when comparing allopatric yellowhammers ($n = 53$) with allopatric pine buntings ($n = 42$). A truncated version of this plot that shows the high F_{ST} tail more clearly can be seen in Supplementary Fig. S5. **B** Mean absolute within-group nucleotide variation (π_W) of allopatric yellowhammers and allopatric pine buntings plotted against absolute between-group nucleotide differentiation (π_B). Each dot represents the average value taken from a window of 2000 sequenced bases across the nuclear genome. The black line indicates where mean within-group nucleotide variation equals between-group nucleotide differentiation. Increasing values of relative differentiation (F_{ST}) calculated for each window are shown in darker shades of blue. **C** Association between relative differentiation (F_{ST}) and absolute between-group nucleotide differentiation (π_B) of allopatric yellowhammers and allopatric pine buntings. Each black dot represents average values calculated from a 2000 bp window of sequenced data. A cubic spline fit between the variables is shown as a purple line.

The twelve mitonuclear genes that appeared within yellowhammer BPPSWs were: *APOPT1*, *COX5A*, *COX17*, *LARS2*, *MRPL1*, *MRPL27*, *MRPL32*, *MRPS7*, *MRPS25*, *NDUFC1*, *SSBP1* and *UQCR11* (Table 2). Five of these genes encode protein subunits of the mitoribosome, three encode structural subunits of the ETC, two encode assembly factors of the ETC, one encodes a mitochondrial aminoacyl-tRNA synthetase and one encodes a single-stranded DNA-binding protein involved in mtDNA replication. Two genes each appear on chromosomes 2, 4 and 18 while the rest appear

on separate chromosomes. Interestingly, three of the five putatively introgressed genes associated with the ETC are specifically associated with complex IV.

The seven mitonuclear genes that appeared within pine bunting BPPSWs were: *ATP5H*, *COX5A*, *MRPL2*, *MRPL58*, *MRPS7*, *MRPS14* and *NDUFB4* (Table 2). Four of these genes encode protein subunits of the mitoribosome and three encode structural subunits of the ETC. Three genes appear on chromosome 18—with two genes sharing the same genomic window—while the

Table 2. Identities, chromosomal locations, windowed Tajima's D values and functions of mitonuclear genes that appeared within 244 yellowhammer BPPSWs and within 222 pine bunting BPPSWs.

Mitonuclear gene	Chromosome where mitonuclear gene is found	Windowed Tajima's D Value	Mitonuclear gene function
<i>Yellowhammer (E. citrinella)</i>			
<i>APOPT1</i>	5	−2.207	Assembly factor/ancillary protein for ETC complex IV
<i>COX5A^a</i>	10	−2.420	Structural subunit of ETC complex IV
<i>COX17</i>	1	−2.509	Assembly factor/ancillary protein for ETC complex IV
<i>LARS2</i>	2	−2.238	Mitochondrial aminoacyl-tRNA synthetase
<i>MRPL1</i>	4	−2.207	Mitochondrial large ribosomal subunit protein
<i>MRPL27</i>	18	−2.214	Mitochondrial large ribosomal subunit protein
<i>MRPL32</i>	2	−2.306	Mitochondrial large ribosomal subunit protein
<i>MRPS7^a</i>	18	−2.628	Mitochondrial small ribosomal subunit protein
<i>MRPS25</i>	12	−2.323	Mitochondrial small ribosomal subunit protein
<i>NDUFC1</i>	4	−2.399	Structural subunit of ETC complex I
<i>SSBP1</i>	1A	−2.362	Single stranded DNA-binding protein
<i>UQCR11</i>	28	−2.499	Structural subunit of ETC complex III
<i>Pine bunting (E. leucocephalos)</i>			
<i>ATP5H^b</i>	18	−2.601	Structural subunit of ETC complex V
<i>COX5A^a</i>	10	−2.545	Structural subunit of ETC complex IV
<i>MRPL2</i>	3	−2.247	Mitochondrial large ribosomal subunit protein
<i>MRPL58^b</i>	18	−2.601	Mitochondrial large ribosomal subunit protein
<i>MRPS7^a</i>	18	−2.252	Mitochondrial small ribosomal subunit protein
<i>MRPS14</i>	8	−2.173	Mitochondrial small ribosomal subunit protein
<i>NDUFB4</i>	1	−2.304	Structural subunit of ETC complex I

In the "Mitonuclear Gene Function" column, ETC stands for "Electron Transport Chain".

^aIndicates a mitonuclear gene that appeared in both yellowhammer and pine bunting BPPSWs.

^bIndicates two genes that appeared within the same BPPSW.

rest of the genes appear on separate chromosomes. The *COX5A* and *MRPS7* genes were found in both yellowhammer and pine bunting BPPSWs.

In comparisons of genomic windows that did versus did not contain mitonuclear genes, we found that average F_{ST} , π_B , π_W and Tajima's D did not differ significantly between these two categories (Supplementary Table S3). Only yellowhammer π_W approached significance ($p = 0.05106$) with average π_W being lower within genomic windows containing mitonuclear genes (0.00373) compared to those not containing mitonuclear genes (0.00400). This finding is consistent with our results showing that mitonuclear genes appear preferentially in yellowhammer BPPSWs, as selective sweeps of mitonuclear genes into a population would lower nucleotide diversity in the associated genomic windows.

DISCUSSION

Yellowhammers and pine buntings show negligible mtDNA differentiation (Irwin et al. 2009) but are well differentiated phenotypically (Panov et al. 2003; Rubtsov and Tarasov 2017) and

moderately differentiated in AFLP nuclear markers (Irwin et al. 2009). In the wake of this discordance in differentiation, Irwin et al. (2009) proposed that mtDNA adaptively introgressed between taxa following a period of allopatric isolation; however, due to the limited information provided by AFLP analyses, the extent of discordance and of nuclear differentiation between yellowhammers and pine buntings remained unknown lending some uncertainty to this hypothesis. In the present study, analysis of genetic variation identified heterogeneous nuclear differentiation between allopatric populations with strong differentiation peaks that separated taxa into distinct genetic clusters. This result supports yellowhammers and pine buntings experiencing a period of separate evolution followed by hybridization within their current contact zone in western and central Siberia (Panov et al. 2003; 2007; Rubtsov 2007; Rubtsov and Tarasov 2017). Our phylogenetic analysis showing a longer branch length between yellowhammers and pine buntings based on nuclear markers—when compared to the very short branch length in a phylogeny based on mtDNA—also corroborates a hypothesis of recent mtDNA introgression and mitochondrial haplotype replacement in this system likely driven by selection (Irwin et al. 2009). In addition,

our finding that mitonuclear genes are over-represented in BPPSWs in yellowhammers lends support for mitonuclear co-introgression from pine buntings into yellowhammers.

Though genetically distinct, the genome-wide F_{ST} between allopatric yellowhammers and pine buntings (0.0232) was comparable to or sometimes lower than the averages seen between avian subspecies (e.g., subspecies of barn swallow: 0.017–0.026 (Scordato et al. 2017); myrtle warbler and Audubon's warbler: 0.077–0.106 (Irwin et al. 2018); yellow- and red-shafted northern flickers: 0.098 (Manthey et al. 2017)). This low genome-wide F_{ST} contrasts with the moderate F_{ST} averages reported from an analysis of AFLP markers performed on the same populations: 0.078 based on allele frequencies and 0.140 based on band frequencies (Irwin et al. 2009). However, the present study also revealed that relative differentiation was highly heterogeneous across the nuclear genome with F_{ST} peaks on various chromosomes. It is possible that the previous AFLP analysis captured a disproportionate number of loci within these differentiation peaks, thereby inflating F_{ST} estimates. This comparison highlights the caution that should be taken when interpreting genome-wide averages, because highly variable genetic differentiation landscapes can cause large variability in F_{ST} estimates when they are based on a limited and non-random sample of loci.

The F_{ST} peaks seen between yellowhammers and pine buntings on larger autosomes and most significantly on the Z chromosome are consistent with the “islands of differentiation” often noted in comparisons of closely related taxa (Harr 2006; Nadeau et al. 2012; Irwin et al. 2018). In contrast to these islands, large regions of close similarity in π_B and π_W suggests high gene flow between taxa across much of the nuclear genome. This scenario is consistent with the observed extensive hybridization between these taxa (Panov et al. 2003, 2007; Rubtsov 2007; Rubtsov and Tarasov 2017). Nevertheless, the high F_{ST} islands—those with much reduced π_W compared to π_B —can be explained by divergent selection causing low gene flow in these regions. It is unlikely that this pattern is the result of genetic drift over an extended period of geographic separation, as this would result in most genomic regions deviating slightly from $\pi_B = \pi_W$ congruence rather than the observed pattern of extreme heterogeneity. Instead, this trend suggests that selection acted in a way that lowered π_W relative to π_B within “islands of differentiation”. Considering that high F_{ST} regions were associated with relatively low values of π_B , we propose that differentiation islands in this system are most consistent with a model invoking repeated bouts of selection that lower nucleotide diversity (Cruickshank and Hahn 2014; Irwin et al. 2018). A sweep-before-differentiation model (Irwin et al. 2018), where F_{ST} peaks are produced by adaptive selective sweeps between populations followed by adaptive selection at the same regions in local populations, is particularly in line with the extensive hybridization presently observed between yellowhammers and pine buntings (Panov et al. 2003, 2007; Rubtsov 2007; Rubtsov and Tarasov 2017).

Of the “islands of differentiation” identified between taxa, the tallest and widest was found on the Z chromosome. Greater differentiation on the Z chromosome 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 stronger positive selection and with less efficient purifying selection on this chromosome (Mank et al. 2010; reviewed in Meisel and Connallon 2013; reviewed in Irwin 2018). However, the large regions of the Z chromosome that have F_{ST} values near zero suggest that additional factors are involved in producing this island of differentiation.

One possible explanation for the large differentiation island on chromosome Z could be that it corresponds with an area of low recombination—a region of connected loci that tend to be inherited together, leading to linked selection of nearby loci. Strong divergent selection acting on one SNP within this region

would act similarly on loci that are linked to it such that a wide, highly divergent genomic block would become fixed and appear as an “island” between taxa (reviewed in Cutter and Payseur 2013). Areas of low recombination and linkage are often associated with inversion polymorphisms (reviewed in Smukowski and Noor 2011) as different orientations of an inversion experience little successful recombination (reviewed in Kirkpatrick 2010). Further research is warranted to characterize the nature of this differentiated region as well as whether it houses an inversion polymorphism.

While numerous “islands of differentiation” were observed between yellowhammers and pine buntings implying moderate genetic divergence, mtDNA introgression has the potential to homogenize the nuclear genomes of these taxa at mitonuclear genes by selecting for co-introgression of compatible alleles (Beck et al. 2015; Sloan et al. 2017; Morales et al. 2018). Because our genomic dataset lacked the resolution to directly test for introgression between taxa, we identified genomic windows with low between-population differentiation and strong signals of selective sweeps in each population (BPPSWs) as a proxy and tested for an association between these regions and mitonuclear genes predicted to coevolve with mtDNA. We found statistical support for mitonuclear genes appearing preferentially in yellowhammer BPPSWs, but no such statistical support in pine bunting BPPSWs. Specifically, our analysis showed that the proportion of mitonuclear genes within yellowhammer BPPSWs was 1.7 times higher than within pine bunting BPPSWs. In other words, mitonuclear genes showing highly similar nucleotide sequences when comparing allopatric populations were more likely to have selectively swept through the yellowhammer population than the pine bunting population. This could suggest that mitonuclear gene introgression, if it occurred, proceeded in the direction of pine buntings into yellowhammers.

We noted additional support for mitonuclear gene introgression in this system in the fact that mitonuclear genes appearing within yellowhammer and pine bunting BPPSWs encode proteins with interrelated mitochondrial functions. Three of the mitonuclear genes within yellowhammer BPPSWs and three within pine bunting BPPSWs encode structural subunits of the ETC. Four of the five ETC complexes are made up of subunits encoded by both the nuclear and mitochondrial genome (Hill 2019b); correct fit between differentially encoded subunits is essential for the flow of electrons and protons across the ETC during oxidative phosphorylation. To put this in perspective, changing even a single amino acid in one subunit can significantly disrupt its ability to interact with other subunits within an ETC complex (e.g., Gershoni et al. 2014). Because of the tight interactions within complexes and the consequences of subunit incompatibility, introgression of mtDNA is expected to select for co-introgression of mitonuclear genes encoding ETC structural subunits. Such co-introgression has been detected between differentially adapted populations of eastern yellow robin where mtDNA introgression between populations was followed by similar introgression of mitonuclear genes encoding subunits of complex I (Morales et al. 2018) and between different species of *Drosophila* where introgression and replacement of the mtDNA of one species during hybridization selected for co-introgression of genes that encode subunits of complex IV (Beck et al. 2015).

Of the ETC complexes, complex IV showed the strongest signal of potential co-introgression in the yellowhammer and pine bunting system. Three of the genes within yellowhammer BPPSWs and one gene within pine bunting BPPSWs were associated with this complex. As well, the gene *COX5A*—a structural subunit of complex IV—appeared in both sets of BPPSWs. It is unlikely and inconsistent with mitonuclear theory that this gene introgressed in both directions in the yellowhammer and pine bunting system; however, it is possible that *COX5A* adaptively swept in both populations which, due to the nature of BPPSW identification, would cause it to appear in both sets of BPPSWs. In such a

situation, we can predict that a particularly adaptive mitonuclear allele may have appeared in one species and swept to high frequency before co-introgressing and sweeping across the other species following mtDNA introgression. Interestingly, the *COX5A* gene was one of the subunits that co-introgressed in the *Drosophila* example discussed above (Beck et al. 2015) lending some support to its particular importance to mitonuclear coevolution. More generally, complex IV is often used as a model for studying mitonuclear interactions due to its distinctive structure where a core of mitochondrial-encoded subunits is surrounded by nuclear-encoded subunits (Saraste 1999). With such an excess of mitonuclear interactions, incompatibility involving complex IV has been detected in several systems including within primate xenomitochondrial cybrids (Barrientos et al. 2000) and between different species of *Drosophila* (Sackton et al. 2003). Furthermore, work by Osada and Akashi (2012) has provided strong evidence for compensatory coevolution between mitonuclear genes related to complex IV—including *COX5A*—and mtDNA among primates at interacting amino acids of differentially encoded subunits. Altogether, these results suggest a crucial role for complex IV in mitonuclear coevolution and co-introgression between pine buntings and yellowhammers.

Another group of mitonuclear genes that appeared consistently within the yellowhammer and pine bunting BPPSWs were those encoding subunits of the mitoribosome (five in yellowhammer BPPSWs and four in pine bunting BPPSWs). *MRP57*, like *COX5A*, appeared in both yellowhammer and pine bunting BPPSWs suggesting that this gene may have adaptively swept through both taxa. Unlike the protein-protein interactions occurring within ETC complexes, mitonuclear interactions in the mitoribosome are between nuclear-encoded proteins and mitochondrial-encoded RNA (Hill 2019b). Protein subunits associate closely with rRNA during the formation of a mitoribosome, but also interact with mRNA and tRNA during the synthesis of mitochondrial proteins (Greber and Ban 2016). Currently, research is limited on the extent and importance of interactions between mitoribosomal subunits and mitochondrial RNA (but see: Sloan et al. 2014; Barreto et al. 2018). However, the fact that interactions between components are extensive and necessary for the synthesis of mitochondrial proteins suggests close coevolution between mtDNA and genes encoding mitoribosomal subunits that could strongly select for mitonuclear co-introgression following mtDNA introgression.

Our mitonuclear analysis provides support for mitonuclear gene introgression in the direction of pine buntings into yellowhammers; nevertheless, we acknowledge that there are shortcomings in this methodology and that our results must be interpreted with caution. Most importantly, our BPPSWs are an imperfect measure of introgression between pine buntings and yellowhammers that utilize diversity and differentiation statistics as a proxy for this process. As such, it is possible that the mitonuclear genes tend to occur in genomic regions that have other characteristics that are associated with BPPSWs, such that the association between mitonuclear genes and BPPSWs is not directly causal. Knowing the direction of historical mtDNA introgression in this system would allow us to determine whether the direction of mitonuclear gene introgression proposed here matches the direction of mtDNA introgression (Gershoni et al. 2009; Burton and Barreto 2012; Hill 2019b). Unfortunately, introgression has apparently eliminated the original mitochondrial haplotype in the receiving population (Irwin et al. 2009) such that determining the direction of mtDNA introgression between yellowhammers and pine buntings is not possible using available data and methodology. Despite these weaknesses in our analysis, we found what we believe is compelling support for mitonuclear gene introgression from pine buntings into yellowhammers which warrants further investigation.

To summarize, yellowhammers and pine buntings are sister taxa that are divergent in appearance, song, and ecology (Panov et al.

2003; Rubtsov and Tarasov 2017) yet vary greatly in their genomic differentiation from virtually none (at the mitochondrial genome) to nearly fixed (the differentiation peak on the Z chromosome). These patterns are best explained by a period of differentiation while geographically separated, followed by hybridization and mtDNA introgression. We found tentative support for mitonuclear gene introgression (compared to introgression of other genes) from pine buntings into yellowhammers, as well as a tendency for mitonuclear genes encoding structural components of the ETC and the mitoribosome to appear within the BPPSWs of both taxa. One intriguing possibility is that mitonuclear co-introgression has resulted in reduced mitonuclear incompatibilities between yellowhammers and pine buntings (Gershoni et al. 2009; Burton and Barreto 2012; Hill 2019b), thereby contributing to their current extensive hybridization within central Siberia (Panov et al. 2003, 2007; Rubtsov 2007; Rubtsov and Tarasov 2017). This idea leads into the question—which can be addressed through a close examination of genomic variation within the hybrid zone—of whether the observed islands of differentiation on the Z and autosomes are sufficient in stabilizing yellowhammers and pine buntings as separate entities where they hybridize, or whether the two taxa are gradually merging into a single species.

DATA AVAILABILITY

Raw DNA sequencing reads are available on the NBCI Sequence Read Archive (BioProject PRJNA768601). Read processing codes, barcodes, genotype data and R codes associated with statistical analyses will be made available on Dryad.

REFERENCES

- Alcaide M, Scordato ESC, Price TD, Irwin DE (2014) Genomic divergence in a ring species complex. *Nature* 511:83–85
- Alström P, Olsson U, Lei F, Wang H, Gao W, Sundberg P (2008) Phylogeny and classification of the old world *Emberizini* (Aves, *Passeriformes*). *Mol Phylogenet Evol* 47:960–973
- Alves PC, Melo-Ferreira J, Freitas H, Boursot P (2008) The ubiquitous mountain hare mitochondria: Multiple introgressive hybridization in hares, genus *Lepus*. *Philos Trans R Soc B* 363:2831–2839
- Avise JC (2000) *Phylogeography: The history and formation of species*. Harvard University Press, Cambridge, Mass
- Ballard JWO, Melvin RG, Katewa SD, Maas K (2007) Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans*. *Evolution* 61:1735–1747
- Barreto FS, Watson ET, Lima TG, Willett CS, Edmands S, Li W et al. (2018) Genomic signatures of mitonuclear coevolution across populations of *Tigriopus californicus*. *Nat Ecol Evol* 2:1250–1257
- Barrientos A, Müller S, Dey R, Wienberg J, Moraes CT (2000) Cytochrome c oxidase assembly in primates is sensitive to small evolutionary variations in amino acid sequence. *Mol Biol Evol* 17:1508–1519
- Baute GJ, Owens GL, Bock DG, Rieseberg LH (2016) Genome-wide genotyping-by-sequencing data provide a high-resolution view of wild *Helianthus* diversity, genetic structure, and interspecies gene flow. *Am J Bot* 103:2170–2177
- Beck EA, Thompson AC, 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* 69:1973–1986
- Bolger AM, Lohse M, Usadel B (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120
- Borge T, Webster MT, Andersson G, Saetre G (2005) Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two *Ficedula* flycatcher species. *Genetics* 171:1861–1873
- Burton RS, Barreto FS (2012) A disproportionate role for mtDNA in Dobzhansky–Muller incompatibilities? *Mol Ecol* 21:4942–4957
- Bryson J, Robert W, De Oca AN, Jaeger JR, Riddle BR (2010) Elucidation of cryptic diversity in a widespread Nearctic treefrog reveals episodes of mitochondrial gene capture as frogs diversified across a dynamic landscape. *Evolution* 64:2315–2330
- Calvo SE, Mootha VK (2010) The mitochondrial proteome and human disease. *Annu Rev Genom Hum Genet* 11:25–44
- Coyne JA, Orr HA (2004) *Speciation*. Sinauer Associates, Sunderland, Mass
- Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol Ecol* 23:3133–3157

- Cutter AD, Payseur BA (2013) Genomic signatures of selection at linked sites: unifying the disparity among species. *Nat Rev Genet* 14:262–274
- Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA et al. (2011) The variant call format and VCFtools. *Bioinformatics* 27:2156–2158
- Dean R, Zimmer F, Mank JE (2014) The potential role of sexual conflict and sexual selection in shaping the genomic distribution of mito-nuclear genes. *Genome Biol Evol* 6:1096–1104
- Diodato D, Ghezzi D, Tiranti V (2014) The mitochondrial aminoacyl tRNA synthetases: Genes and syndromes. *Int J Cell Biol* 2014:1–11
- Drown DM, Preuss KM, Wade MJ (2012) Evidence of a paucity of genes that interact with the mitochondrion on the X in mammals. *Genome Biol Evol* 4:763–880
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES et al. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* 6:e19379
- Ernster L, Schatz G (1981) Mitochondria: A historical review. *J Cell Biol* 91:227s–255s
- Gascuel O (1997) BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Mol Biol Evol* 14:685–695
- Geraldes A, Askelson KK, Nikelski E, Doyle FI, Harrower WL, Winker K et al. (2019) Population genomic analyses reveal a highly differentiated and endangered genetic cluster of northern goshawks (*Accipiter gentilis laingi*) in Haida Gwaii. *Evol Appl* 12:757–772
- Gershoni M, Levin L, Ovadia O, Toiw Y, Shani N, Dadon S et al. (2014) Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biol Evol* 6:2665–2680
- Gershoni M, Templeton AR, Mishmar D (2009) Mitochondrial bioenergetics as a major motive force of speciation. *BioEssays* 31:642–650
- Greber BJ, Ban N (2016) Structure and function of the mitochondrial ribosome. *Annu Rev Biochem* 85:103–132
- Harr B (2006) Genomic islands of differentiation between house mouse subspecies. *Genome Res* 16:730–737
- Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptus fuligator*. *Proc Natl Acad Sci USA* 101:14812–14817
- Hejase HA, Salman-Minkov A, Campagna L, Hubisz MJ, Lovette IJ, Gronau I et al. (2020) Genomic islands of differentiation in a rapid avian radiation have been driven by recent selective sweeps. *Proc Natl Acad Sci USA* 117:30554–30565
- Hill GE (2019a) Reconciling the mitonuclear compatibility species concept with rampant mitochondrial introgression. *Integr Comp Biol* 59:912–924
- Hill GE (2019b) Mitonuclear Ecology (First ed.). Oxford University Press, Oxford
- Hill GE (2020) Genetic hitchhiking, mitonuclear coadaptation, and the origins of mt DNA barcode gaps. *Ecol Evol* 10:9048–9059
- Hulsey CD, Bell KL, García-de-León FJ, Nice CC, Meyer A (2016) Do relaxed selection and habitat temperature facilitate biased mitogenomic introgression in a narrowly endemic fish? *Ecol Evol* 6:3684–3698
- Irwin DE (2018) Sex chromosomes and speciation in birds and other ZW systems. *Mol Ecol* 27:3831–3851
- Irwin DE, Alcaide M, Delmore KE, Irwin JH, Owens GL (2016) Recurrent selection explains parallel evolution of genomic regions of high relative but low absolute differentiation in a ring species. *Mol Ecol* 25:4488–4507
- Irwin DE, Milá B, Toews DPL, Brelsford A, Kenyon HL, Porter AN et al. (2018) A comparison of genomic islands of differentiation across three young avian species pairs. *Mol Ecol* 27:4839–4855
- Irwin DE, Rubtsov AS, Panov EN (2009) Mitochondrial introgression and replacement between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza leucocephalos*) (Aves: Passeriformes). *Biol J Linn Soc* 98:422–438
- Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN (2007) Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes* 7:535–543
- Kirkpatrick M (2010) How and why chromosome inversions evolve. *PLoS Biol* 8:e1000501
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25:1754–1760
- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N et al. (2009) The sequence alignment map format and SAMtools. *Bioinformatics* 25:2078–2079
- Llopart A, Herrig D, Brud E, Stecklein Z (2014) Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Mol Ecol* 23:1124–1136
- Lotz C, Lin AJ, Black CM, Zhang J, Lau E, Deng N et al. (2014) Characterization, design, and function of the mitochondrial proteome: From organs to organisms. *J Proteome Res* 13:433–446
- Lu J, Wu C (2005) Weak selection revealed by the whole-genome comparison of the X chromosome and autosomes of human and chimpanzee. *Proc Natl Acad Sci* 102:4063–4067
- Lynch M, Koskella B, Schaack S (2006) Mutation pressure and the evolution of organellar genomic architecture. *Science* 311:1727–1730
- Mank JE, Nam K, Ellegren H (2010) Faster-Z evolution is predominantly due to genetic drift. *Mol Biol Evol* 27:661–670
- Manthey JD, Geiger M, Moyle RG (2017) Relationships of morphological groups in the northern flicker superspecies complex (*Colaptes auratus* & *C. chrysoides*). *Syst Biodivers* 15:183–191
- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A et al. (2010) The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 20:1297–1303
- Meisel RP, Connallon T (2013) The faster-X effect: Integrating theory and data. *Trends Genet* 29:537–544
- Moore WS (1995) Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees versus nuclear-gene trees. *Evolution* 49:718–726
- Morales HE, Pavlova A, Amos N, Major R, Kilian A, Greening C et al. (2018) Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird lineages inhabiting different climates. *Nat Ecol Evol* 2:1258–1267
- Nadeau NJ, Whibley A, Jones RT, Davey JW, Dasmahapatra KK, Baxter SW et al. (2012) Genomic islands of divergence in hybridizing *Heliconius* butterflies identified by large-scale targeted sequencing. *Philos Trans R Soc B* 367:343–353
- Osada N, Akashi H (2012) Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Mol Biol Evol* 29:337–346
- Panov EN, Rubtsov AS, Monzikov DG (2003) Hybridization between yellowhammer and pine bunting in Russia. *Dutch Bird* 25:17–31
- Panov EN, Rubtsov AS, Mordkovich MV (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. *Zool Zh* 86:1362–1378
- Paradis E, Schliep K (2019) Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35:526–528
- Price T (2008) Speciation in birds. Roberts and Co, Greenwood Village, Colo
- R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, <http://www.r-project.org/>
- Rubtsov AS (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. *Zool Zh* 86:863–876
- Rubtsov AS, Opaev AS (2012) Phylogeny reconstruction of the yellowhammer (*Emberiza citrinella*) and pine bunting (*Emberiza leucocephala*) based on song and morphological characters. *Biol Bull Russ Acad Sci* 39:715–728
- Rubtsov AS, Tarasov VV (2017) Relations between the yellowhammer (*Emberiza citrinella*) and the pine bunting (*Emberiza leucocephalos*) in the forested steppe of the Trans-Urals. *Biol Bull* 44:1059–1072
- Ruegg K, Anderson EC, Boone J, Pouls J, Smith TB (2014) A role for migration-linked genes and genomic islands in divergence of a songbird. *Mol Ecol* 23:4757–4769
- Sackton TB, Corbett-Detig RB, Nagaraju J, Vaishna L, Arunkumar KP, Hartl DL (2014) Positive selection drives faster-z evolution in silkworms. *Evolution* 68:2331–2342
- Sackton TB, Haney RA, Rand DM (2003) Cytonuclear coadaptation in *Drosophila*: Disruption of cytochrome c oxidase activity in backcross genotypes. *Evolution* 57:2315–2325
- Saraste M (1999) Oxidative phosphorylation at the fin de siècle. *Science* 283:1488–1493
- Scordato ESC, Wilkins MR, Semenov G, Rubtsov AS, Kane NC, Safran RJ (2017) Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Mol Ecol* 26:5676–5691
- Sloan DB, Havird JC, Sharbrough J (2017) The on-again, off-again relationship between mitochondrial genomes and species boundaries. *Mol Ecol* 26:2212–2236
- Sloan DB, Triant DA, Wu M, Taylor DR (2014) Cytonuclear interactions and relaxed selection accelerate sequence evolution in organelle ribosomes. *Mol Biol Evol* 31:673–682
- Smukowski CS, Noor MAF (2011) Recombination rate variation in closely related species. *Heredity* 107:496–508
- Stacklies H, Redestig H, Scholz M, Walthert D, Selbig J (2007) PcaMethods—a bio-conductor package providing PCA methods for incomplete data. *Bioinformatics* 23:1164–1167
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Thornton K, Long M (2002) Rapid divergence of gene duplicates on the *Drosophila melanogaster* X chromosome. *Mol Biol Evol* 19:918–925
- Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Mol Ecol* 21:3907–3930
- Turner S (2018) Qqman: An R package for visualizing GWAS results using Q-Q and Manhattan plots. *J Open Source Softw* 3:731
- Wang S, Ore MJ, Mikkelsen EK, Lee-Yaw J, Toews DPL, Rohwer S et al. (2021) Signatures of mitonuclear coevolution in a warbler species complex. *Nat Commun* 12:4279–4279

- Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Künstner A et al. (2010) The genome of a songbird. *Nature* 464:757–762
- Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370
- Wu C (2001) The genic view of the process of speciation. *J Evol Biol* 14:851–865
- Yannic G, Dubey S, Hausser J, Basset P (2010) Additional data for nuclear DNA give new insights into the phylogenetic position of *Sorex granarius* within the *Sorex araneus* group. *Mol Phylogenet Evol* 57:1062–1071

ACKNOWLEDGEMENTS

For providing valuable feedback, we thank Dolph Schluter, Eric Taylor, Judith Mank, Elizabeth Natola, Rashika Ranasinghe, Kenneth Askelson, Finola Fogarty, Quinn McCallum, Ana Barreira, Jamie Clarke, Armando Geraldes, Jessica Irwin and several anonymous reviewers. For their kindness and support during field work, we thank the Tazeev family and Madelyn Ore. For providing additional samples, we thank 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 and their accompanying personnel. Major research funding was provided by the Natural Sciences and Engineering Research Council of Canada (NSERC CGSM award to EN, Discovery Grants RGPIN-2017-03919 and RGPAS-2017-507830 awarded to DI) and by the Werner and Hildegard Hesse research awards (Research award in Ornithology and Fellowship in Ornithology awarded to EN by the University of British Columbia).

AUTHOR CONTRIBUTIONS

EN, DI, and ASR conceived of this study. ASR collected samples. EN and ASR completed molecular techniques. EN conducted data analysis and wrote this paper with input from DI and ASR.

COMPETING INTERESTS

The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41437-022-00580-8>.

Correspondence and requests for materials should be addressed to Ellen Nikelski.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.