Extreme Heterogeneity in Genomic Differentiation between Phenotypically Divergent Songbirds: A Test of Mitonuclear Co-introgression

Ellen Nikelski[1, *], Alexander S. Rubtsov[2], and Darren Irwin[1]

[1] Department of Zoology, and Biodiversity Research Centre, 6270 University Blvd., University of British Columbia, Vancouver, BC, Canada

[2] State Darwin Museum, Moscow, Russia

[*] Present address: Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada

Corresponding Author:

Ellen Nikelski, Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, ON, Canada. Email: ellen.nikelski@mail.utoronto.ca

Keywords:

mitonuclear co-introgression, mtDNA, mitonuclear gene, genetic differentiation, chromosome Z, Aves

Abstract:

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 the 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 mitonuclear 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 the extent of mitonuclear discordance and characterize nuclear differentiation patterns 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 tested for preferential introgression of mitonuclear genes and found statistical support for this process in yellowhammers. A role for mitonuclear coevolution in this system is supported by a stronger signal of co-introgression in genes coding for subunits of the mitoribosome and electron transport chain complexes. Altogether, our study emphasizes the extreme variation seen in differentiation across genomic components and study systems as well as highlights the ramifications of this variation in species evolution.

51 ***Introduction***

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

66 In most cases, mtDNA shows strong differentiation between speciating populations (e.g.
67 Hebert et al. 2004; Kerr et al. 2007). These patterns are partially driven by the mitochondrial
68 genome's uniparental inheritance and haploid nature which decrease its effective population size
69 to ¼ that of autosomal DNA (Moore, 1995). Combined with a relatively high mutation rate
70 (Lynch et al. 2006), this low effective population size leads to the rapid fixation of mitochondrial
71 mutations via genetic drift and strong mtDNA differentiation between taxa. Additionally, a lack
72 of recombination across the mitochondrial genome can further contribute to mtDNA divergence
73 through rampant genetic hitchhiking (Hill, 2020). Here, positive selection for an adaptive

74 mtDNA mutation causes the fixation of genetic variants across the mitochondrial genome
75 resulting in strong differentiation between speciating taxa.

76 Sex chromosomes—specifically the Z (Borge et al. 2005; Ruegg et al. 2014; Sackton et
77 al. 2014) and X chromosomes (Thorton & Long, 2002; Lu & Wu, 2005; Harr, 2006)—show
78 moderate levels of genetic differentiation between speciating populations, but often much greater
79 genetic differentiation than autosomes. This trend is driven by two important characteristics of
80 Z/X chromosomes. First, on Z/X chromosomes, beneficial recessive mutations are immediately
81 exposed to selection in the heterogametic sex allowing these advantageous variants to fix more
82 rapidly than they would on autosomes—a phenomenon known as the "faster Z/X effect" (Meisel
83 & Connallon, 2013; Irwin, 2018). Second, because Z/X chromosomes are inherited as either one
84 or two copies depending on the sex, these chromosomes have a lower effective population size
85 than autosomes (Mank et al. 2010; Irwin, 2018). This lower effective population size allows for
86 the fixation of a greater number of neutral and slightly deleterious mutations due to less effective
87 purifying selection and a larger role of genetic drift. Working in tandem, these two
88 characteristics drive genetic divergence of Z/X chromosomes between speciating populations.

89 Across autosomes, genetic differentiation between speciating taxa is relatively modest
90 compared to mtDNA and sex chromosomes and, interestingly, it is often characterized by
91 "islands of differentiation"—peaks of high relative differentiation that appear within a
92 background of low relative differentiation (Harr, 2006; Nadeau et al. 2012; Hejase et al. 2020).
93 Explanations for these "islands" often invoke reduced gene flow between speciating taxa during
94 secondary contact (Wu, 2001) and/or repeated bouts of selection prior to and following
95 secondary contact (Cruickshank and Hahn, 2014; Irwin et al. 2018). In the former scenario,
96 differentiation peaks are hypothesized to house the loci responsible for reproductive barriers

97 between taxa making them resistant to the homogenizing influence of gene flow (Wu, 2001). In
98 the latter scenario, differentiation islands are described as genomic areas that have experienced
99 recurrent selection or selective sweeps (Cruickshank and Hahn, 2014; Irwin et al. 2018). These
100 events reduce genetic diversity first in the ancestral population and then in both daughter
101 populations producing relative peaks in genetic differentiation between taxa.

102 Despite the general differentiation trends discussed above, an increasing number of
103 studies have reported differentiation patterns that greatly diverge from what is normally seen
104 between speciating populations (Irwin et al. 2009; Yannic et al. 2010; Bryson et al. 2012). In
105 these studies, genomic components that are normally highly divergent between taxa show
106 remarkably low differentiation when compared to other genomic components and observable
107 phenotypes. In particular, cases of "mitonuclear discordance", where mtDNA shows
108 dramatically low differentiation compared to nuclear DNA, have been identified across lineages
109 (Toews & Brelsford, 2012). One hypothesis that may explain mitonuclear discordance is
110 introgression of mitochondrial haplotypes from one population into another following secondary
111 contact and hybridization. Introgression of mtDNA can occur neutrally due to various processes
112 such as sex-biased dispersal or hybrid zone movement (Toews & Brelsford, 2012); however, the
113 extreme degree of discordance observed in certain systems (e.g. Alves et al. 2008; Irwin et al.
114 2009), suggests that this process may also be driven by selection and occur adaptively by
115 providing a fitness advantage to the receiving population.

116 Introgression of foreign mtDNA is thought to provide fitness advantages through two
117 major avenues. First, because variation in mtDNA has been associated with variation in
118 mitochondrial efficiency under different abiotic conditions (e.g. Ballard et el. 2007), mtDNA
119 introgression may allow the receiving population to better adapt and survive within a changing or

120 novel environment (Hulsey et al. 2016; Sloan et al. 2017; Hill, 2019). For example, mtDNA
121 introgression has been tentatively associated with thermal adaptation in populations of rabbits
122 (Alves et al. 2008) and cichlids (Hulsey et al. 2016). The second way that mtDNA introgression
123 can provide a fitness advantage to a receiving population is by replacing a mitochondrial genome
124 with a high mutational load (Sloan et al. 2017; Hill, 2019). As described earlier, the low effective
125 population size (Moore, 1995) and high mutation rate (Lynch et al. 2006) of mtDNA can result
126 in rapid fixation of mitochondrial mutations, including deleterious mutations. This tendency
127 combined with genetic hitchhiking of deleterious mutations (Hill, 2020), can lead to a high
128 mtDNA mutational load which may decrease mitochondrial efficiency. Through introgression of
129 a foreign mitochondrial haplotype with a lower mutational load, the receiving population may be
130 able to regain greater mitochondrial function which translates into higher fitness (Llopart et al.
131 2014; Hulsey et al. 2016; Sloan et al. 2017; Hill, 2019).

132 Adaptive mtDNA introgression presents a compelling hypothesis for the appearance of
133 extreme "mitonuclear discordance" between speciating taxa. However, this idea becomes
134 complicated when we consider recent work that suggests strong coevolution between the
135 mitochondrial and nuclear genomes (Hill, 2019). In most bilaterian animals, the mitochondrial
136 genome has been reduced to 37 genes. As a result of this low gene content, all mitochondrial
137 processes—including oxidative phosphorylation and transcription, translation and replication of
138 mtDNA—are reliant on more than 1000 proteins encoded by specific "mitonuclear genes" across
139 the nuclear genome (Calvo & Mootha, 2010; Lotz et al. 2014). Tight associations between
140 mitochondrial and mitonuclear proteins imply tight coevolution between the mitochondrial and
141 nuclear genomes due to selection for mitochondrial efficiency: changes in one create selective
142 pressure for compatible changes in the other (Gershoni et al. 2009; Burton & Barreto, 2012; Hill,

143 2019). This association also suggests that hybridization and gene flow between speciating
144 populations will be selected against due to recombination of coevolved mitochondrial and
145 mitonuclear genes exposing genetic incompatibilities (mitonuclear incompatibilities) within
146 hybrid individuals. In this way, mitonuclear coevolution has the potential to contribute towards
147 reproductive isolation between taxa and to strongly select against mtDNA introgression.
148 Nevertheless, if mtDNA introgression is highly adaptive such that its fitness advantages
149 outweigh the fitness disadvantages of mitonuclear incompatibilities, this process may still be
150 able to occur. In this scenario, researchers hypothesize that mtDNA introgression will select for
151 co-introgression of mitonuclear genes to maintain coevolved genotypes and optimum
152 mitochondrial function (Sloan et al. 2017; Hill, 2019). Evidence for mitonuclear co-introgression
153 has been found in some systems (Beck et al. 2015; Morales et al. 2018; Wang et al. 2021), but
154 the significance of this process in genome-wide nuclear differentiation and general speciation
155 trends is still up for debate.

156 The yellowhammer (Passeriformes: Emberizidae: *Emberiza citrinella*) and pine bunting
157 (*E. leucocephalos*) system (Figure 1) is one that may capture the complex interplay between
158 genetic differentiation and mitonuclear coevolution. Thought to have diverged in isolation during
159 the Pleistocene glaciations (Irwin et al. 2009), this Eurasian avian sister pair is highly divergent
160 in plumage and moderately divergent in song and ecology (Panov et al. 2003; Rubtsov &
161 Tarasov, 2017). Yet, despite their differences, yellowhammers and pine buntings hybridize
162 extensively in a large and seemingly expanding secondary contact zone in central and western
163 Siberia (Panov et al. 2003; 2007; Rubtsov, 2007; Irwin et al. 2009; Rubtsov & Tarasov, 2017).
164 Previous genomic work has identified mitonuclear discordance between allopatric
165 yellowhammers and pine buntings (Irwin et al. 2009) as they are nearly identical in mtDNA but

166 show moderate differentiation in nuclear AFLP (Amplified Fragment Length Polymorphism)
167 markers. To explain these results, Irwin et al. (2009) suggested that mtDNA introgressed
168 adaptively from one species into the other during a previous selective sweep, and this hypothesis
169 was supported by several statistical tests performed on the mtDNA haplotype network. Based on
170 mitonuclear theory, such mtDNA introgression could select for similar introgression at
171 mitonuclear genes to maintain optimum mitochondrial function (Sloan et al. 2017; Hill, 2019).
172 The resulting lack of mitonuclear incompatibilities between taxa due to this co-introgression
173 could facilitate their continued hybridization and hinder the build-up of reproductive barriers.
174 With this opposition between the observed strong nuclear differentiation and the potential for
175 mitonuclear co-introgression, the fate of the yellowhammer and pine bunting system remains
176 uncertain. Depending on which way the scales tip, yellowhammers and pine buntings may
177 continue to diverge and speciate or they may continue to hybridize and eventually collapse into
178 one interbreeding population.

179 Here, we present a largescale comparison of DNA sequence variation across the nuclear
180 genomes of allopatric yellowhammers and pine buntings. With this data, we address key
181 questions regarding genetic differentiation and mitonuclear coevolution in this system. First,
182 what is the degree and structure of genetic differentiation between yellowhammers and pine
183 buntings across the nuclear genome? Earlier AFLP analyses identified clear differentiation of
184 nuclear markers between yellowhammers and pine buntings (Irwin et al. 2009), but those results
185 were not based on actual DNA sequences and only captured a small portion of the nuclear
186 genome. Comparing patterns of differentiation across the nuclear genome enables better
187 understanding of the extent of mitonuclear discordance between taxa and of which genomic
188 regions or components (e.g. sex chromosomes) show particularly high differentiation. Second, is

189 there an over-representation of known mitonuclear genes within genomic regions that show the
190 strongest signals of being introgressed–a pattern consistent with mitonuclear co-introgression?
191 Evidence of mitonuclear co-introgression and the resulting lack of mitonuclear incompatibilities
192 could explain the continued, extensive hybridization seen between yellowhammers and pine
193 buntings and implicate this process as a force that counters divergence and the evolution of
194 strong reproductive barriers between groups. By answering these questions, we hope to provide
195 insight on the evolutionary trajectory of yellowhammers and pine buntings (i.e. whether it is one
196 of continued population divergence or of population merging) and also to explore how the
197 interplay between genetic differentiation and mitonuclear coevolution influences the speciation
198 process more generally.

199

200 ***Materials and Methods***

201 **Sampling**

202 We included 109 blood and tissue samples in this study: 53 phenotypic yellowhammers,
203 42 phenotypic pine buntings, and 14 other members of Emberizidae (one *Emberiza aureola*
204 [yellow-breasted bunting], one *Emberiza calandra* [corn bunting], one *Emberiza cioides*
205 [meadow bunting], one *Emberiza hortulana* [ortolan bunting], four *Emberiza stewarti* [white-
206 capped bunting], and six *Emberiza cirlus* [cirl bunting]) to put variation between yellowhammers
207 and pine buntings into a deeper phylogenetic context (Figure 1A; Table 1; Supplementary Table
208 1). A total of 91 of our samples were included in the AFLP analysis of Irwin et al. (2009) while
209 18 samples were examined for the first time as part of the present research.

210 When possible, body measurements and photographs were taken of live birds or museum
211 skins. Yellowhammer and pine bunting males were scored phenotypically and sorted into
212 phenotypic classes based on the protocols presented in Panov et al. (2003) and Rubtsov &
213 Tarasov (2017). Briefly, each male received a score from 0-7 for three plumage traits, where 0
214 represents a phenotypically pure yellowhammer and 7 represents a phenotypically pure pine
215 bunting. These three traits were (1) the amount of chestnut plumage (vs. yellow or white) at the
216 brow, (2) the amount of chestnut plumage (vs. yellow or white) at the throat, and (3) background
217 plumage colour, meaning the strength of yellow—ranging from bright yellow to pure white—in
218 head and body plumage that did not show brown or black streaking. Phenotypic classes included:
219 pure *citrinella* (PC), almost *citrinella* (SC, for "semi-*citrinella*"), *citrinella* hybrid (CH), yellow
220 hybrid (YH), white hybrid (WH), *leucocephalos* hybrid (LH), almost *leucocephalos* (SL, for
221 "semi-*leucocephalos*") and pure *leucocephalos* (PL) (Rubtsov & Tarasov, 2017). Any SC and SL
222 individuals that appeared in the allopatric zones were grouped together with PC and PL
223 individuals respectively and treated as phenotypic yellowhammers and phenotypic pine buntings
224 in subsequent analyses (Figure 1B).

225 **DNA extraction and genotyping-by-sequencing**

226 DNA was extracted from samples using a standard phenol-chloroform method. We then
227 divided the DNA samples into four genotyping-by-sequencing (GBS) libraries (Elshire et al.
228 2011). The 109 samples included in this study were sequenced together with 226 yellowhammer,
229 pine bunting and hybrid DNA samples collected near and within the sympatric zone as part of a
230 larger project (Nikelski et al. in prep). The libraries were prepared as per the protocol described
231 by Alcaide et al. (2014) with modifications specified by Geraldes et al. (2019) except that we
232 maintained a 300-400 bp fragment size during size selection. Paired-end sequencing was

233 completed by Genome Québec using an Illumina HiSeq 4000 system, producing more than 1.2
234 billion reads, each 150 bp in length, across the four GBS libraries.

235 **Genotyping-by-sequencing data filtering**

236 We processed the reads following Irwin et al. (2016; 2018), as summarized here. Reads
237 were demultiplexed using a custom perl script designed by Baute et al. (2016). Next, reads were
238 trimmed for quality using Trimmomatic version 0.36 (Bolger et al. 2014) with the parameters:
239 TRAILING:3, SLIDINGWINDOW:4:10, MINLEN:30. Trimmed reads were aligned to the zebra
240 finch reference genome (*Taeniopygia guttata* version 3.2.4; Warren et al. 2010) using the
241 program BWA-MEM (Li & Durbin, 2009) and a BAM file of this information was created for
242 each individual using the programs Picard (http://broadinstitute.github.io/picard/) and SAMtools
243 (Li et al. 2009). BAM files were converted into GVCF files using the HaplotypeCaller command
244 as part of GATK version 3.8 (McKenna et al. 2010). We then combined information from the
245 individuals in two ways to create 1) a genome-wide "variant site" VCF file containing only
246 variant site information, and 2) a series of chromosome-specific "info site" VCF files containing
247 information on both variant and invariant sites with sufficient coverage.

248 To create the genome-wide "variant site" VCF file, we used the GenotypeGVCFs
249 command in GATK to identify and isolate single nucleotide polymorphisms (SNPs) among
250 individuals. This command also converted the variant site information into a single VCF file
251 encompassing the entire nuclear genome. Using a combination of VCFtools (Danecek et al.
252 2011) and GATK, we filtered the VCF file to remove indels and non-biallelic SNPs. We also
253 discarded loci with QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, or ReadPosRankSum < -8.0.
254 Finally, loci with more than 60% missing genotypes were removed. The average coverage of
255 variable sites in the resulting VCF file was 16.59.

256 To convert GVCF files into "info site" VCF files, we similarly employed the
257 GenotypeGVCFs command in GATK with the addition of the -allSites and -L flags to retain
258 invariant sites and split the information into chromosome-specific files. The resulting VCF files
259 were filtered using VCFtools and GATK to remove indels, sites with more than two alleles, sites
260 with more than 60% missing data, sites with MQ values lower than 20 and sites with
261 heterozygosities greater than 60% (to avoid potential paralogs).

262 **Variant site analyses**

263 The genome-wide "variant site" VCF file was analyzed using modified versions of the R
264 scripts described in Irwin et al. (2018), and all of our analyses used R version 3.6.2 (R Core
265 Team, 2014). A total of 374,780 SNPs were identified among allopatric yellowhammers and pine
266 buntings. For each of these SNPs, we calculated sample size, allele frequency, and Weir and
267 Cockerham's F_{ST} (Weir & Cockerham, 1984). Genetic differentiation between yellowhammers
268 and pine buntings was then visualized using a principal components analysis (PCA) generated
269 with the pca command and the svdImpute method to account for any missing genomic data using
270 the pcaMethods package (Stacklies et al. 2007). A Manhattan plot of F_{ST} values for 349,807
271 SNPs with known genomic locations was created using the package qqman (Turner, 2018).

272 **Differentiation across the genome**

273 To thoroughly investigate nuclear differentiation between allopatric yellowhammers and
274 pine buntings, we performed further analysis on both variant and invariant loci within "info site"
275 VCF files using R scripts described in Irwin et al. (2018).

276 We calculated Weir and Cockerham's F_{ST}, between-group nucleotide differentiation (π_B)
277 and within-group nucleotide variation (π_W) within nonoverlapping windows of available

278 sequence data across each chromosome. The first window was positioned at the "start" of each
279 chromosome as described in the zebra finch reference genome (Warren et al. 2010). We used a
280 window size of 2,000 bp of sequenced data rather than 10,000 bp (as in Irwin et al. 2018), to
281 visualize narrow peaks in relative and absolute differentiation within our dataset. We hereafter
282 refer to these windows as "genomic windows."

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

291 **Phylogenetic comparison with other Emberizidae species**

292 We employed whole-genome averages of π_B between allopatric yellowhammers and pine
293 buntings as well as among these focal species and six other Emberizidae species (*Emberiza*
294 *aureola, Emberiza calandra, Emberiza cioides, Emberiza cirlus, Emberiza hortulana and*
295 *Emberiza stewarti*) to estimate a phylogeny. A list of average π_B values for each species pair was
296 converted into a distance matrix and used to create an unrooted neighbour-joining tree. This tree
297 was constructed using the ape package (Paradis & Schliep, 2019) and the BioNJ algorithm
298 (Gascuel, 1997) with *Emberiza aureola* set as the outgroup (Alström et al. 2008). The phylogeny
299 was created to provide further support for the sister relationship between yellowhammers and
300 pine buntings that had previously been hypothesized using mitochondrial markers (Alström et al.

301 2008; Irwin et al. 2009) but that was questioned in some studies (Rubtsov & Opeav, 2012). In
302 creating this phylogeny we were also able to better investigate the degree of mitonuclear
303 discordance within the system using a greater amount of nuclear data.

304 **Signals of mitonuclear co-introgression**

305 To test for preferential mitonuclear gene introgression (compared to other genomic
306 regions) between allopatric yellowhammers and pine buntings, we asked whether there was an
307 association between a list of known mitonuclear genes and a list of genomic windows that have
308 statistical characteristics most consistent with introgression. We call the latter "putative
309 introgression windows" (PIWs), and we explain their identification in detail below.

310 The mitonuclear genes that we analyzed for signals of introgression were all protein
311 coding with products that interacted directly with mtDNA or an immediate product of the
312 mitochondrial genome (i.e. protein or RNA). For these nuclear-encoded genes, any change in
313 mtDNA including those caused by introgression would likely cause selection for co-
314 introgression of compatible alleles (Gershoni et al. 2009; Burton & Barreto, 2012; Hill, 2019).
315 Mitonuclear genes that met these criteria included those that encode protein subunits of ATP
316 synthase or the first, third and fourth complex of the electron transport chain (ETC), assembly
317 and ancillary proteins involved in the formation of the ETC, and proteins that are part of the
318 transcription, translation or DNA replication machinery within the mitochondria (Diodato et al.
319 2013; Greber and Ban 2016; Hill, 2019). After removing any genes that were not annotated in
320 the zebra finch reference genome or that lacked a specific location on the reference genome, a
321 total of 162 mitonuclear genes remained for analysis (Supplementary Table 2).

322 Given the ongoing hybridization observed between yellowhammers and pine buntings
323 and the evidence for mtDNA introgression and replacement, it is reasonable to assume that
324 regions of the nuclear genome have also moved between the species. For each taxon, we
325 identified a subset of genomic windows that showed statistical properties most consistent with
326 substantial introgression, although we cannot be certain of introgression for any particular PIW.
327 We identified PIWs as those that possessed a low π_B average and a low Tajima's D average.
328 Low π_B indicates high similarity between the nucleotide sequences of the two groups as would
329 be expected if alleles had introgressed from one taxon into the other. Low Tajima's D suggests a
330 past selective sweep within a population which would also be expected if an adaptive allele had
331 introgressed from a separate taxon and swept throughout the receiving population. For this
332 analysis, our quantitative criteria for a PIW were a Tajima's D value within the lowest 5% of the
333 available windowed values and a π_B value within the lowest 30% of the available windowed
334 values. The Tajima's D threshold was kept relatively low to capture windows with particularly
335 strong signals of selection that could be associated with adaptive introgression while the π_B
336 threshold was left higher to identify an appreciable number of PIWs for analysis.

337 Using a custom R script, each mitonuclear gene was assigned to the genomic window
338 that minimized the absolute difference between the location of the mitonuclear gene centre and
339 the location of the window centre. We then determined the number of mitonuclear genes that
340 occurred within the PIWs of each taxon.

341 We conducted a Fisher's Exact test for both yellowhammers and pine buntings to
342 determine whether the proportion of mitonuclear genes within PIWs was significantly different
343 from what would be expected based on the total proportion of protein coding genes appearing
344 within these windows. To complete this analysis, a list of the 14,008 protein-coding genes

345 annotated in the zebra finch reference genome (Warren et al. 2010) was compiled. We removed
346 our 162 mitonuclear genes from this list and assigned the remaining 13,846 non-mitonuclear
347 genes to genomic windows using the methodology described above. The proportion of
348 mitonuclear and non-mitonuclear genes appearing in PIWs were then compared separately in
349 yellowhammers and pine buntings.

350

351 ***Results***

352 When comparing allopatric yellowhammers and pine buntings, we identified 374,780
353 variable SNPs within our "variant site" VCF file and 13,703,455 invariant and 699,122 variant
354 sites across thirty autosomes and the Z chromosome within our "info site" VCF files following
355 filtering. In the latter "info site" files, we designated a total of 7,187 genomic windows (of 2000
356 sequenced bp each) across the genome, with each window covering an average distance of about
357 139 kilobases.

358 **Phylogenetic comparison with other Emberizidae species**

359 An unrooted neighbour-joining tree of average π_B values between yellowhammers, pine
360 buntings and six other Emberizidae species (Figure 2) depicted similar species relationships as
361 were identified previously using mitochondrial markers (Alström et al. 2008; Irwin et al. 2009).
362 Relative branch lengths were also similar, with the major exception being the branch length
363 between yellowhammers and pine buntings which was much longer in our analysis using nuclear
364 DNA. To put this into context, we determined the relative genetic distance between
365 yellowhammers and pine buntings versus the genetic distance between *E. stewarti* and either
366 member of the yellowhammer/pine bunting clade for our nuclear phylogeny and for the

367 previously calculated mitochondrial phylogeny (Irwin et al. 2009). The nuclear ratio was 11.4
368 times greater than the mitochondrial ratio which corroborates the presence of strong mitonuclear
369 discordance between yellowhammers and pine buntings. These results also support the
370 hypothesis of an extended period of divergence between yellowhammers and pine buntings
371 followed by adaptive mtDNA introgression.

372 **Overall genetic differentiation**

373 Based on 374,780 SNPs, the genome-wide F_{ST} estimate was 0.0232 between allopatric
374 yellowhammers and pine buntings. Despite this low average, a PCA based on the same SNP
375 genotypes separated yellowhammers and pine buntings into tight genetic clusters (Figure 3).
376 Two pine buntings were outliers along PC1, while the remaining yellowhammers and pine
377 buntings separated into distinct groups mainly along PC2. Further investigation into these
378 outliers revealed that they were males from the same location, but a kinship analysis completed
379 as part of a separate study did not find close kinship between the two pine buntings that could
380 explain their position (Nikelski et al. in prep). We also examined the PC1 loadings and found
381 that the signal for PC1 position was broadly distributed across the nuclear genome, rather than
382 being concentrated in a few highly influential regions (Supplementary Figure 1). Finally, we
383 temporarily removed one of the outliers and re-ran the PCA. This caused the other outlier to fall
384 into the pine bunting cluster, but revealed a further yellowhammer outlier (Supplementary Figure
385 2). Removal of this yellowhammer outlier in addition to one member of the pine bunting outlier
386 pair in turn revealed another yellowhammer outlier (Supplementary Figure 3). It is unclear what
387 is responsible for these outliers, but the distinct yellowhammer and pine bunting genetic clusters
388 remained intact in all the PCAs considered.

389 **Differentiation across the genome**

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

395 Patterns of between-group nucleotide differentiation (π_B) and within-group nucleotide
396 variation (π_W) were also heterogenous across the genome and comparable to each other in
397 magnitude: genome-wide π_B = 0.0041; genome-wide π_W for both taxa= 0.0040 (Figure 5;
398 Supplementary Figure 4). Because between-group and within-group nucleotide differentiation
399 are so intimately related in their evolution and calculation, it is expected that windowed averages
400 of these two statistics will show a highly positive relationship. In fact, most windowed π_B and
401 π_W averages fell near a 1:1 association line (Figure 6C) which is equivalent to no or little
402 differentiation. However, some genomic windows showed much reduced π_W compared to π_B;
403 these were the windows with high F_{ST}. Additionally, we detected a weak but highly significant
404 negative correlation between the windowed averages of F_{ST} and π_B (Spearman's Rank
405 Correlation: -0.1196, $p < 2.2 \times 10^{-16}$; Figure 6D) as is hypothesized if peaks in relative
406 differentiation are products of repeated selective events (Cruickshank & Hahn, 2014; Irwin et al.
407 2018).

408 Finally, we found that Tajima's D varied across the genome but was mostly negative
409 (Figure 5; Supplementary Figure 4), consistent with a history of population growth and/or

410 selective sweeps. The average genome-wide Tajima's D was similar between populations: -1.377
411 for yellowhammers and -1.335 for pine buntings.

412 **Signals of mitonuclear co-introgression**

413 Of the 7187 genomic windows identified across the nuclear genome, we classified 244
414 (3.4%) as PIWs within yellowhammers and 222 (3.1%) as PIWs within pine buntings. Average
415 values of π_B and Tajima's D in yellowhammer PIWs were 0.0016 and -2.3751 respectively, and
416 0.0019 and -2.3369 in pine bunting PIWs respectively. In non-PIWs, the average values of π_B
417 and Tajima's D were 0.0042 and -1.3416 in yellowhammers and 0.0042 and -1.3031 in pine
418 buntings. Of the PIWs identified in yellowhammer and pine bunting populations, 71 were shared
419 between the taxa. It should be noted that sharing of some PIWs is expected given that the
420 contribution of π_B to window selection was identical for both taxa (in contrast, Tajima's D was
421 calculated separately for yellowhammers and pine buntings).

422 Our examination of the gene content within yellowhammer PIWs revealed that they
423 contained a higher percentage (7.4%) of mitonuclear genes (12 of the 162 genes considered) than
424 of non-mitonuclear genes (4.1%; 574 of the 13,846 genes considered). This difference was
425 statistically significant (Fisher's Exact test: p = 0.04714), providing evidence for mitonuclear
426 genes preferentially appearing within yellowhammer PIWs. Pine bunting PIWs contained 4.3%
427 of the mitonuclear genes (7 of the 162 genes considered) and 3.3% of the non-mitonuclear genes
428 (455 of the 13,846 genes considered), a difference that was not statistically significant (Fisher's
429 Exact test: p = 0.3806).

430 The twelve mitonuclear genes that appeared within yellowhammer PIWs were: APOPT1,
431 COX5A, COX17, LARS2, MRPL1, MRPL27, MRPL32, MRPS7, MRPS25, NDUFC1, SSBP1

432 and UQCR11 (Table 2). Five of these genes encode protein subunits of the mitoribosome, three
433 encode structural subunits of the ETC, two encode assembly factors of the ETC, one encodes a
434 mitochondrial aminoacyl-tRNA synthetase and one encodes a single-stranded DNA-binding
435 protein involved in mtDNA replication. Two genes each appear on chromosomes 2, 4 and 18
436 while the rest appear on separate chromosomes. Interestingly, three of the five putatively
437 introgressed genes associated with the ETC are specifically associated with complex IV.

438 The seven mitonuclear genes that appeared within pine bunting PIWs were: ATP5H,
439 COX5A, MRPL2, MRPL58, MRPS7, MRPS14 and NDUFB4 (Table 2). Four of these genes
440 encode protein subunits of the mitoribosome and three encode structural subunits of the ETC.
441 Three genes appear on chromosome 18—with two genes sharing the same genomic window—
442 while the rest of the genes appear on separate chromosomes. The COX5A and MRPS7 genes
443 was found in both yellowhammer and pine bunting PIWs.

444

445 ***Discussion***

446 Yellowhammers and pine buntings show negligible mtDNA differentiation (Irwin et al.
447 2009) but are well differentiated phenotypically (Panov et al. 2003, Rubtsov & Tarasov, 2017)
448 and moderately differentiated in AFLP nuclear markers (Irwin et al. 2009). In the wake of this
449 mitonuclear discordance, Irwin et al. (2009) proposed that mtDNA adaptively introgressed
450 between taxa following a period of allopatric isolation; however, due to the limited information
451 provided by AFLP analyses, the extent of mitonuclear discordance and of nuclear differentiation
452 between yellowhammers and pine buntings remained unknown lending some uncertainly to this
453 hypothesis. In the present study, analysis of genetic variation across the nuclear genome

454 identified heterogeneous nuclear differentiation between allopatric populations with strong
455 differentiation peaks that separated taxa into distinct genetic clusters. This result supports
456 yellowhammers and pine buntings experiencing a period of separate evolution followed by the
457 hybridization within their current contact zone in western and central Siberia. Our phylogenetic
458 analysis showing a longer branch length between yellowhammers and pine buntings based on
459 nuclear markers—when compared to a phylogeny based on mtDNA—also corroborates a
460 hypothesis of recent mtDNA introgression and mitochondrial haplotype replacement in this
461 system likely driven by selection (Irwin et al. 2009). Additionally, our finding that mitonuclear
462 genes are over-represented in putative introgression windows in yellowhammers is consistent
463 with mitonuclear co-introgression from pine buntings into yellowhammers.

464 Though genetically distinct, the genome-wide F_{ST} between allopatric yellowhammers and
465 pine buntings (0.0232) was comparable to or sometimes lower than the averages seen between
466 avian subspecies (e.g., subspecies of barn swallow: 0.017-0.026 [Scordato et al. 2017]; myrtle
467 warbler and Audubon's warbler: 0.077-0.106 [Irwin et al. 2018]; yellow- and red-shafted
468 northern flickers: 0.098 [Manthey et al. 2016]). This low genome-wide F_{ST} contrasts with the
469 moderate F_{ST} averages reported from an analysis of AFLP markers performed on the same
470 populations: 0.078 based on allele frequencies and 0.140 based on band frequencies (Irwin et al.
471 2009). However, the present study also revealed that relative differentiation was highly
472 heterogeneous across the nuclear genome with F_{ST} peaks on various chromosomes. It is possible
473 that the previous AFLP analysis captured a disproportionate number of loci within these
474 differentiation peaks, thereby inflating F_{ST} estimates. This comparison highlights the caution that
475 should be taken when interpreting genome-wide averages as highly variable genetic

476 differentiation landscapes can cause large variability in F_{ST} estimates when they are based on a
477 limited and non-random sample of loci.

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

497 Of the "islands of differentiation" identified between taxa, the tallest and widest was
498 found on the Z chromosome. Greater differentiation on the Z chromosome compared to

499 autosomes is a common observation when comparing closely related species (Borge et al. 2005;
500 Ruegg et al. 2014; Sackton et al. 2014) and is consistent with stronger positive selection (the
501 “faster Z/X effect") and with less efficient purifying selection on this chromosome (Mank et al.
502 2010; reviewed in Meisel & Connallon, 2013; reviewed in Irwin, 2018). However, the large
503 region of the Z chromosome that have F_{ST} values near zero suggests that additional factors are
504 involved in producing this island of differentiation.

505 One possible explanation for the large differentiation island on chromosome Z could be
506 that it corresponds with an area of low recombination—a region of connected loci that tend to be
507 inherited together, leading to linked selection of nearby loci. Strong divergent selection acting on
508 one SNP within this region would act similarly on all the loci that are linked to it such that a
509 wide, highly divergent genomic block would become fixed and appear as an “island” between
510 taxa (reviewed in Cutter & Payseur, 2013). Areas of low recombination and linkage are often
511 associated with inversion polymorphisms (reviewed in Smukowski & Noor, 2011) as different
512 orientations of an inversion experience little successful recombination (reviewed in Kirkpatrick,
513 2010). Further research is warranted to characterize the nature of this differentiated region as
514 well as whether it houses an inversion polymorphism.

515 While numerous “islands of differentiation” were observed between yellowhammers and
516 pine buntings implying moderate genetic divergence, mtDNA introgression has the potential to
517 homogenize the nuclear genomes of these taxa at mitonuclear genes by selecting for co-
518 introgression of compatible alleles (Beck et al. 2015; Sloan et al. 2017; Morales et al. 2018). In
519 the yellowhammer and pine bunting system, we found statistical support for the preferential
520 introgression of mitonuclear genes into yellowhammers, but no statistical support for the
521 preferential introgression into pine buntings. More specifically, our analysis showed that the

522 proportion of introgressing mitonuclear genes was 1.7X higher in yellowhammers versus pine
523 buntings. While these results suggest mitonuclear introgression primarily in the direction of pine
524 buntings into yellowhammers, we also noted some consistency in the functions of putatively
525 introgressing mitonuclear genes in both taxa.

526 Three of the mitonuclear genes within yellowhammer PIWs and three within pine bunting
527 PIWs encode structural subunits of the ETC. The ETC is broken into five protein complexes
528 which, through a series of enzymatic reactions, perform oxidative phosphorylation to produce
529 ATP necessary for organism survival (reviewed in Ernster & Schatz, 1981). Four of the five ETC
530 complexes are made up of subunits encoded by both the nuclear and mitochondrial genome (Hill,
531 2019) and correct fit between differentially encoded subunits is essential for the flow of electrons
532 and protons across the ETC. To put this in perspective, changing even a single amino acid in one
533 subunit can significantly disrupt its ability to interact with other subunits within an ETC complex
534 (e.g. Gershoni et al. 2014). Because of the tight interactions within complexes and the
535 consequences of subunit incompatibility, introgression of mtDNA is expected to select for co-
536 introgression of mitonuclear genes encoding ETC structural subunits. Such co-introgression has
537 been detected between differentially adapted populations of eastern yellow robin where mtDNA
538 introgression between populations was followed by similar introgression of mitonuclear genes
539 encoding subunits of complex I (Morales et al. 2018) and between different species of
540 *Drosophila* where introgression and replacement of the mtDNA of one species during
541 hybridization selected for co-introgression of genes that encode subunits of complex IV (Beck et
542 al. 2015).

543 Of the ETC complexes, complex IV showed the strongest signal of co-introgression in
544 the yellowhammer and pine bunting system. Three of the genes within yellowhammer PIWs and

545 one gene within pine bunting PIWs were associated with this complex. As well, the gene
546 COX5A—a structural subunit of complex IV—appeared in both sets of PIWs. It is unlikely that
547 this gene introgressed in both directions, but it is possible that COX5A adaptively swept in both
548 populations. In this situation, a particularly adaptive allele may have appeared in one species and
549 swept to high frequency before co-introgressing into the other species following mtDNA
550 introgression. Interestingly, the COX5A gene was one of the subunits that co-introgressed in the
551 *Drosophila* example discussed above (Beck et al. 2015) lending some support to its particular
552 importance to mitonuclear coevolution. More generally, complex IV is often used as a model for
553 studying mitonuclear interactions due to its distinctive structure where a core of mitochondrial-
554 encoded subunits is surrounded by nuclear-encoded subunits (Saraste, 1999). With such an
555 excess of mitonuclear interactions, incompatibility involving complex IV has been investigated
556 and detected in several systems including within primate xenomitochondrial cybrids (Barrientos
557 et al. 2000) and between different species of *Drosophila* (Sackton et al. 2003). Furthermore,
558 work by Osada & Akashi (2012) has provided strong evidence for compensatory coevolution
559 between mitonuclear genes related to complex IV—including COX5A—and mtDNA among
560 primates particularly at interacting amino acids of differentially encoded subunits. Altogether,
561 these results suggest a crucial role for complex IV in mitonuclear coevolution as it may relate to
562 divergence and speciation between yellowhammers and pine buntings.

563 Another group of mitonuclear genes that appeared consistently within the yellowhammer
564 and pine bunting PIWs were those encoding subunits of the mitoribosome (five in
565 yellowhammer PIWs and four in pine bunting PIWs). MRPS7, like COX5A, appeared in both
566 yellowhammer and pine bunting PIWs suggesting that this gene may have adaptively swept
567 through both taxa. Unlike the protein-protein interactions occurring within ETC complexes,

568 mitonuclear interactions in the mitoribosome are between nuclear-encoded proteins and
569 mitochondrial-encoded RNA (Hill, 2019). Protein subunits associate closely with rRNA during
570 the formation of a mitoribosome, but also interact with mRNA and tRNA during the synthesis of
571 mitochondrial proteins (Greber & Ban, 2016). Currently, research is limited on the extent and
572 importance of interactions between mitoribosomal subunits and mitochondrial RNA. However,
573 the fact that interactions between components are extensive and necessary for the synthesis of
574 mitochondrial proteins suggests close coevolution between mtDNA and genes encoding
575 mitoribosomal subunits that could strongly select for mitonuclear co-introgression following
576 mtDNA introgression.

577 To summarize, yellowhammers and pine buntings are sister taxa that are divergent in
578 appearance, song, and ecology (Panov et al. 2003; Rubtsov & Tarasov, 2017) yet vary greatly in
579 their genomic differentiation from virtually none (at the mitochondrial genome) to nearly fixed
580 (the differentiation peak on the Z chromosome). These patterns are best explained by a period of
581 differentiation while geographically separated, followed by hybridization and mitochondrial
582 DNA introgression. We observed evidence for preferential mitonuclear gene introgression
583 (compared to introgression of other genes) from pine buntings into yellowhammers, as well as a
584 tendency for mitonuclear genes encoding structural components of the ETC and the
585 mitoribosome to appear within the PIWs of both taxa. One intriguing possibility is that
586 mitonuclear co-introgression could have resulted in reduced mitonuclear incompatibilities
587 between yellowhammers and pine buntings (Gershoni et al. 2009; Burton & Barreto, 2012; Hill,
588 2019), thereby contributing to their current extensive hybridization within central Siberia (Panov
589 et al. 2003; 2007; Rubtsov, 2007; Rubtsov & Tarasov, 2017). This idea leads into the question—
590 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.

Author contributions:

E. N., D.I., and A.S.R. conceived of this study. A.S.R. collected samples. E. N. and A.S.R. completed molecular techniques. E.G.M.N. conducted data analysis and wrote this manuscript with input from D.I. and A.S.R.

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 and Jessica Irwin. 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, 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 E.N, Discovery Grants RGPIN-2017-03919 and RGPAS-2017-507830 awarded to D.I.) and by the Werner and Hildegard Hesse research awards

(Research award in Ornithology and Fellowship in Ornithology awarded to E.G.M.N. by the University of British Columbia).

Data Accessibility Statement:

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 upon publication acceptance.

Conflict of Interest Statement:

None declared

Bibliography

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

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

Alves, P. C., Melo-Ferreira, J., Freitas, H., & Boursot, P. (2008). The ubiquitous mountain hare mitochondria: Multiple introgressive hybridization in hares, genus *Lepus*. *Philosophical Transactions. Biological Sciences, 363*(1505), 2831-2839. doi:10.1098/rstb.2008.0053

Avise, J. C. (2000). Phylogeography: The history and formation of species. Harvard University Press.

Ballard, J. W. O., Melvin, R. G., Katewa, S. D., & Maas, K. (2007). Mitochondrial DNA variation is associated with measurable differences in life-history traits and mitochondrial metabolism in *Drosophila simulans. Evolution, 61*(7), 1735-1747. doi:10.1111/j.1558-5646.2007.00133.x

Barrientos, A., Müller, S., Dey, R., Wienberg, J., & Moraes, C. T. (2000). Cytochrome c oxidase assembly in primates is sensitive to small evolutionary variations in amino acid sequence. *Molecular Biology and Evolution, 17*(10), 1508-1519. doi:10.1093/oxfordjournals.molbev.a026250

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

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

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

Bryson, J., Robert W, De Oca, A. N., 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

673 Calvo, S. E., & Mootha, V. K. (2010). The mitochondrial proteome and human disease. *Annual*
674 *Review of Genomics and Human Genetics, 11*(1), 25-44.
675 doi:10.1146/annurev-genom-082509-141720

676

677 Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

678

679 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of
680 speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology,*
681 *23*(13), 3133-3157. doi:10.1111/mec.12796

682

683 Cutter, A. D., & Payseur, B. A. (2013). Genomic signatures of selection at linked sites: Unifying
684 the disparity among species. *Nature Reviews. Genetics, 14*(4), 262-274.
685 doi:10.1038/nrg3425

686

687 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R.
688 E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R. & 1000 Genomes
689 Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics,*
690 *27*(15), 2156-2158. doi:10.1093/bioinformatics/btr330

691

692 Diodato, D., Ghezzi, D., & Tiranti, V. (2014). The mitochondrial aminoacyl tRNA synthetases:
693 Genes and syndromes. *International Journal of Cell Biology, 2014*(2014), 1-11.
694 doi:10.1155/2014/787956

695

696 Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell,
697 S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high
698 diversity species. *PloS One, 6*(5), e19379. doi:10.1371/journal.pone.0019379

699

700 Ernster, L., & Schatz, G. (1981). Mitochondria: A historical review. *The Journal of Cell*
701 *Biology, 91*(3), 227s-255s. doi:10.1083/jcb.91.3.227s

702

703 Gascuel, O. (1997). BIONJ: An improved version of the NJ algorithm based on a simple model
704 of sequence data. *Molecular Biology and Evolution, 14*(7), 685-695.
705 doi:10.1093/oxfordjournals.molbev.a025808

706

707 Geraldes, A., Askelson, K. K., Nikelski, E., Doyle, F. I., Harrower, W. L., Winker, K., & Irwin,
708 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., Levin, L., Ovadia, O., Toiw, Y., Shani, N., Dadon, S., Barzilai, N., Bergman, A., Atzmon, G., Wainstein, J., Tsur, A., Nijtmans, L. G. J., Glaser, B., & Mishmar, D. (2014). Disrupting mitochondrial-nuclear coevolution affects OXPHOS complex I integrity and impacts human health. *Genome Biology and Evolution, 6*(10), 2665-2680. doi:10.1093/gbe/evu208

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

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

Harr, B. (2006). Genomic islands of differentiation between house mouse subspecies. *Genome Research, 16*(6), 730-737. doi:10.1101/gr.5045006

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

Hill, G. E. (2019). *Mitonuclear Ecology* (First ed.). Oxford University Press.

Hill, G. E. (2020). Genetic hitchhiking, mitonuclear coadaptation, and the origins of mt DNA barcode gaps. *Ecology and Evolution, 10*(17), 9048-9059. doi:10.1002/ece3.6640

Hulsey, C. D., Bell, K. L., García-de-León, F. J., Nice, C. C., & Meyer, A. (2016). Do relaxed selection and habitat temperature facilitate biased mitogenomic introgression in a narrowly endemic fish? *Ecology and Evolution, 6*(11), 3684- 3698. doi:10.1002/ece3.2121

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., 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

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

Kirkpatrick, M. (2010). How and why chromosome inversions evolve. *PLoS Biology, 8*(9), e1000501. doi:10.1371/journal.pbio.1000501

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

Llopart A., Herrig D., Brud E., & Stecklein Z. (2014) Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. *Molecular Ecology, 23* (5), 1124–1136. doi:10.1111/mec.12678

783 Lotz, C., Lin, A. J., Black, C. M., Zhang, J., Lau, E., Deng, N., Wang, Y., Zong, N. C., Choi, J.
784 H., Xu, T., Liem, D. A., Korge, P., Weiss, J. N., Hermjakob, H., Yates, J. R., III.,
785 Apweiler, R., & Ping, P. (2014). Characterization, design, and function of the
786 mitochondrial proteome: From organs to organisms. *Journal of Proteome Research,*
787 *13*(2), 433-446. doi:10.1021/pr400539j

788

789 Lu, J., & Wu, C. (2005). Weak selection revealed by the whole-genome comparison of the X
790 chromosome and autosomes of human and chimpanzee. *Proceedings of the National*
791 *Academy of Sciences, 102*(11), 4063-4067. doi:10.1073/pnas.0500436102

792

793 Lynch, M., Koskella, B., & Schaack, S. (2006). Mutation pressure and the evolution of organelle
794 genomic architecture. *Science, 311*(5768), 1727-1730. doi:10.1126/science.1118884

795

796 Mank, J. E., Nam, K., & Ellegren, H. (2010). Faster-Z evolution is predominantly due to genetic
797 drift. *Molecular Biology and Evolution, 27*(3), 661-670. doi:10.1093/molbev/msp282

798

799 Manthey, J. D., Geiger, M., & Moyle, R. G. (2017). Relationships of morphological groups in
800 the northern flicker superspecies complex (*Colaptes auratus* & *C. chrysoides*).
801 *Systematics and Biodiversity, 15*(3), 183-191. doi:10.1080/14772000.2016.1238020

802

803 McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella,
804 K., Altshuler, D., Gabriel, S., Daley, M., & DePristo, M. A. (2010). The genome analysis
805 toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
806 *Genome Research, 20*(9), 1297-1303. doi:10.1101/gr.107524.110

807

808 Meisel, R. P., & Connallon, T. (2013). The faster-X effect: Integrating theory and data. *Trends in*
809 *Genetics, 29*(9), 537-544. doi:10.1016/j.tig.2013.05.009

810

811 Moore, W. S. (1995). Inferring phylogenies from mtDNA variation: Mitochondrial-gene trees
812 versus nuclear-gene trees. *Evolution, 49*(4), 718-726. doi:10.2307/2410325

813

814 Morales, H. E., Pavlova, A., Amos, N., Major, R., Kilian, A., Greening, C., & Sunnucks, P.
815 (2018). Concordant divergence of mitogenomes and a mitonuclear gene cluster in bird
816 lineages inhabiting different climates. *Nature Ecology & Evolution, 2*(8), 1258-1267.
817 doi:10.1038/s41559-018-0606-3

818

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

Osada, N., & Akashi, H. (2012). Mitochondrial-nuclear interactions and accelerated compensatory evolution: Evidence from the primate cytochrome c oxidase complex. *Molecular Biology and Evolution, 29*(1), 337-346. doi:10.1093/molbev/msr211

Panov, E. N., Rubtsov, A. S., & Monzikov, 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

Price, T. (2008). *Speciation in birds*. Roberts and Co.

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/

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 of the Russian Academy of Sciences, 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

Sackton, T. B., Haney, R. A., & Rand, D. M. (2003). Cytonuclear coadaptation in *Drosophila*: Disruption of cytochrome c oxidase activity in backcross genotypes. *Evolution, 57*(10), 2315-2325. doi:10.1111/j.0014-3820.2003.tb00243.x

Saraste, M. (1999). Oxidative phosphorylation at the fin de siècle. *Science, 283*(5407), 1488-1493. doi:10.1126/science.283.5407.1488

Scordato, E. S. C., Wilkins, M. R., Semenov, G., Rubtsov, A. S., Kane, N. C., & Safran, R. J. (2017). Genomic variation across two barn swallow hybrid zones reveals traits associated with divergence in sympatry and allopatry. *Molecular Ecology, 26*(20), 5676-5691. doi:10.1111/mec.14276

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

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

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics, 123*(3), 585-595.

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

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

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

Wang, S., Ore, M. J., Mikkelsen, E. K., Lee-Yaw, J., Toews, D. P. L., Rohwer, S., & Irwin, D. (2021). Signatures of mitonuclear coevolution in a warbler species complex. *Nature Communications*, in press. doi:10.1038/s41467-021-24586-8

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

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

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. *Molecular Phylogenetics and Evolution, 57*(3), 1062-1071. doi:10.1016/j.ympev.2010.09.015

Tables

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

Sampling Location	Latitude (°N)	Longitude (°E)	*E. citrinella* sample size	*E. leucocephalos* 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**

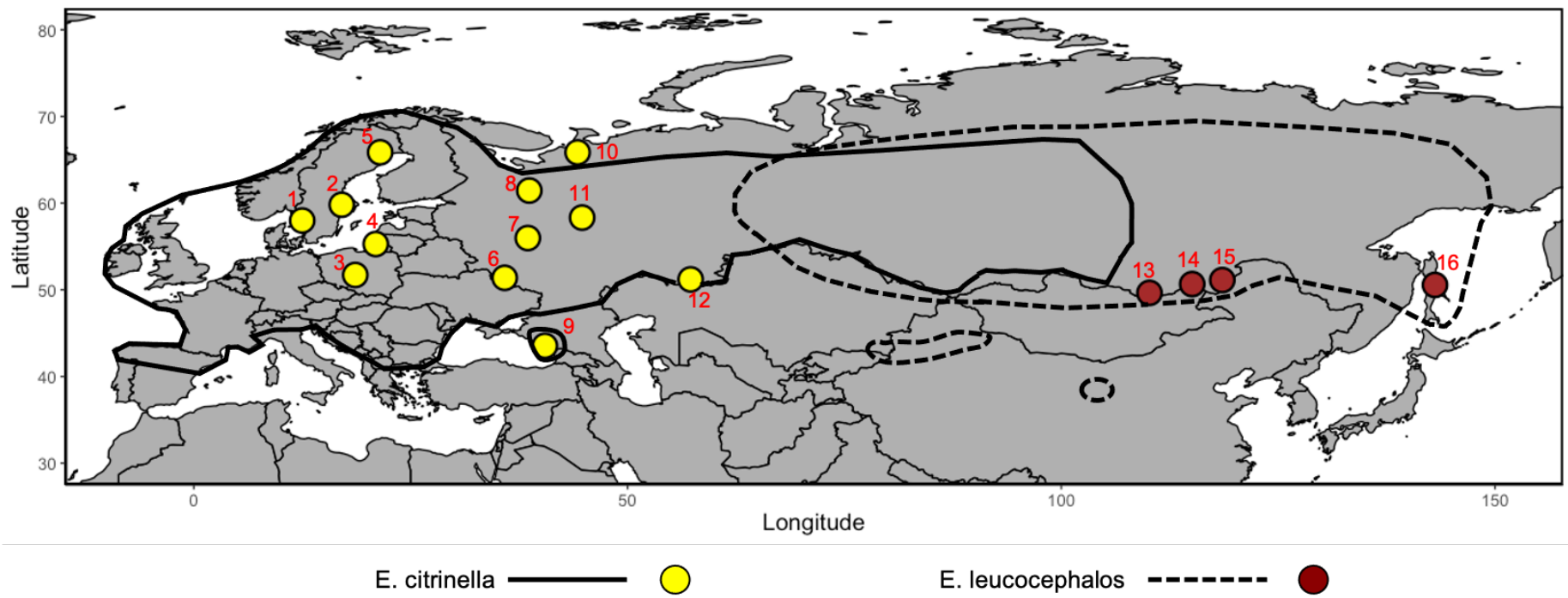
Table 2. Identities, chromosomal locations, windowed Tajima's D values and functions of mitonuclear genes that appeared within 244 yellowhammer PIWs or within 222 pine bunting PIWs. In the "Mitonuclear Gene Function" column, ETC stands for "Electron Transport Chain". The "*" indicates a mitonuclear gene that appeared in both yellowhammer and pine bunting PIWs. The "†" indicates two genes that appeared within the same PIW.

Mitonuclear Gene	Chromosome where mitonuclear gene is found	Windowed Tajima's D Value	Mitonuclear Gene Function
		Yellowhammer (E. citrinella)	
APOPT1	5	-2.207	Assembly factor/ancillary protein for ETC complex IV
COX5A*	10	-2.420	Structural subunit of ETC complex IV
COX17	1	-2.509	Assembly factor/ancillary protein for ETC complex IV
LARS2	2	-2.238	Mitochondrial aminoacyl-tRNA synthetase
MRPL1	4	-2.207	Mitochondrial large ribosomal subunit protein
MRPL27	18	-2.214	Mitochondrial large ribosomal subunit protein
MRPL32	2	-2.306	Mitochondrial large ribosomal subunit protein
MRPS7*	18	-2.628	Mitochondrial small ribosomal subunit protein
MRPS25	12	-2.323	Mitochondrial small ribosomal subunit protein
NDUFC1	4	-2.399	Structural subunit of ETC complex I
SSBP1	1A	-2.362	Single stranded DNA-binding protein
UQCR11	28	-2.499	Structural subunit of ETC complex III
		Pine bunting (E. leucocephalos)	
ATP5H†	18	-2.601	Structural subunit of ETC complex V
COX5A*	10	-2.545	Structural subunit of ETC complex IV

MRPL2	3	-2.247	Mitochondrial large ribosomal subunit protein
MRPL58†	18	-2.601	Mitochondrial large ribosomal subunit protein
MRPS7*	18	-2.252	Mitochondrial small ribosomal subunit protein
MRPS14	8	-2.173	Mitochondrial small ribosomal subunit protein
NDUFB4	1	-2.304	Structural subunit of ETC complex I

Figures

A.



B.

Emberiza citrinella

Pure Citrinella (PC)

Almost Citrinella (SC)

Emberiza leucocephalos

Pure Leucocephalos (PL)

Almost Leucocephalos (SL)

Figure 1. 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 1. Sampling location points are coloured 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.

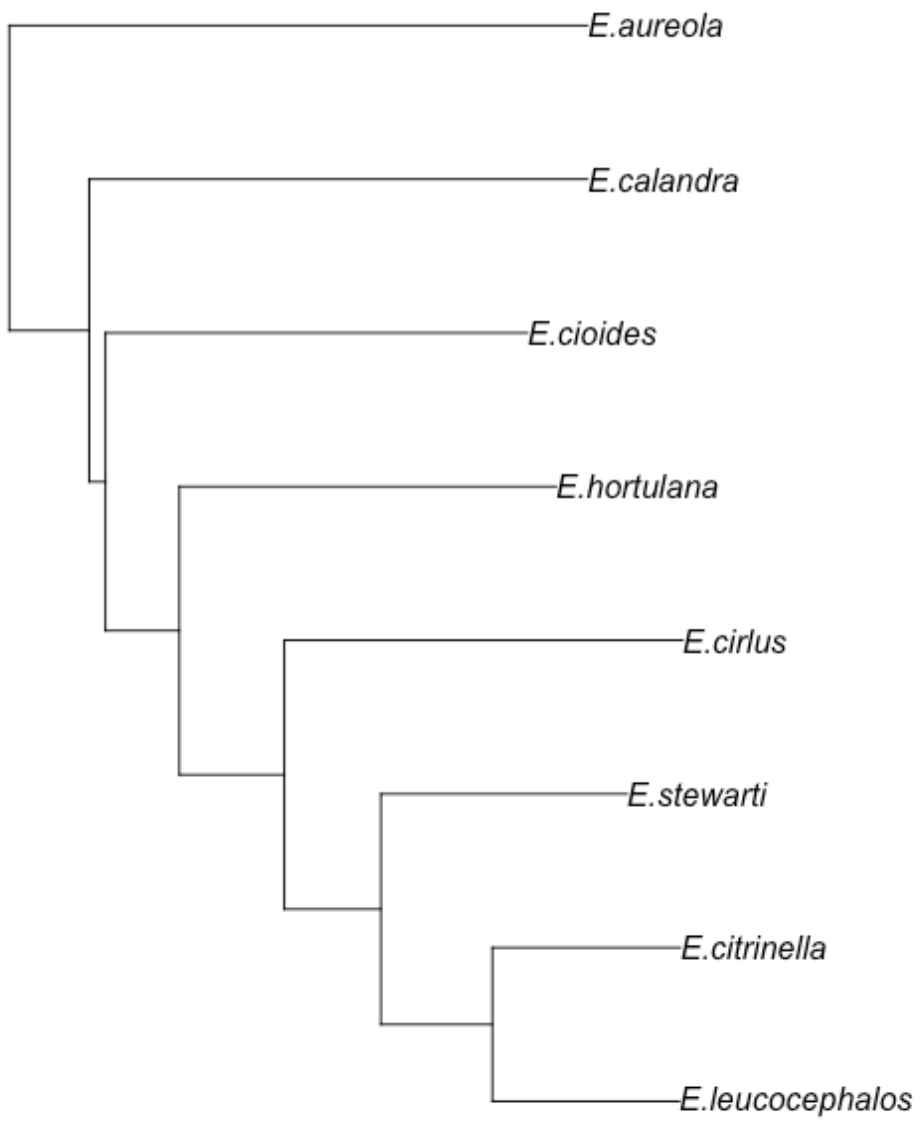


Figure 2. Unrooted neighbour-joining tree of Emberizidae species constructed based on average absolute between-population nucleotide diversity (π_B). Sample sizes for each species are as follows: *E. aureola* = 1, *E. calandra* = 1, *E. cioides* = 1, *E. hortulana* = 1, *E. cirlus* = 6, *E. stewarti*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

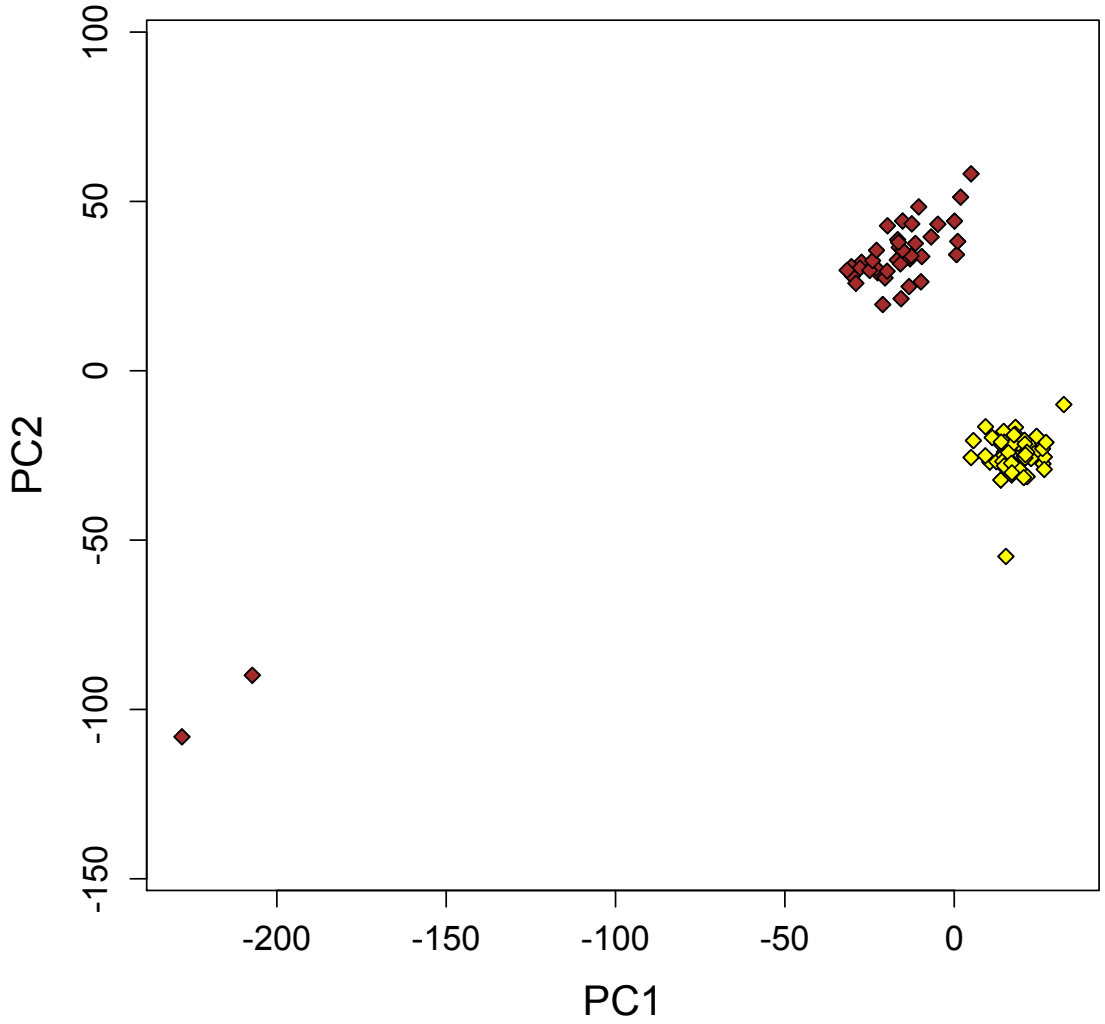


Figure 3. PCA of genetic variation between allopatric yellowhammers (yellow; n = 53) and allopatric pine buntings (brown; n = 42), based on 374,780 genome-wide SNPs. PC1 and PC2 explain 3.6% and 2.9%, respectively, of the variation among individuals.

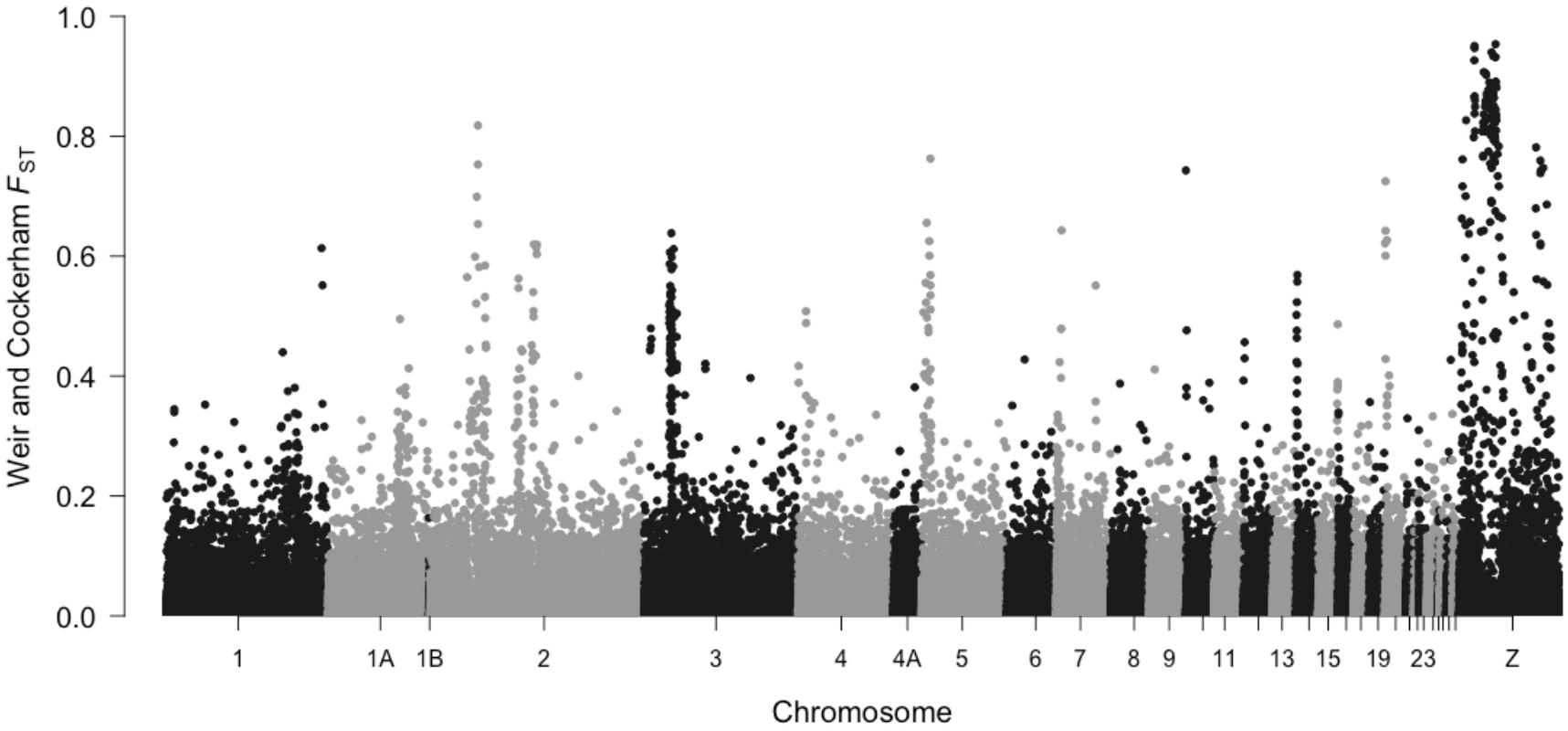


Figure 4. Relative differentiation (F_{ST}) of 349,807 genome-wide SNPs identified between allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42), with chromosomes represented with alternating black and grey. Narrow regions of elevated differentiation can be seen on many autosomes, and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

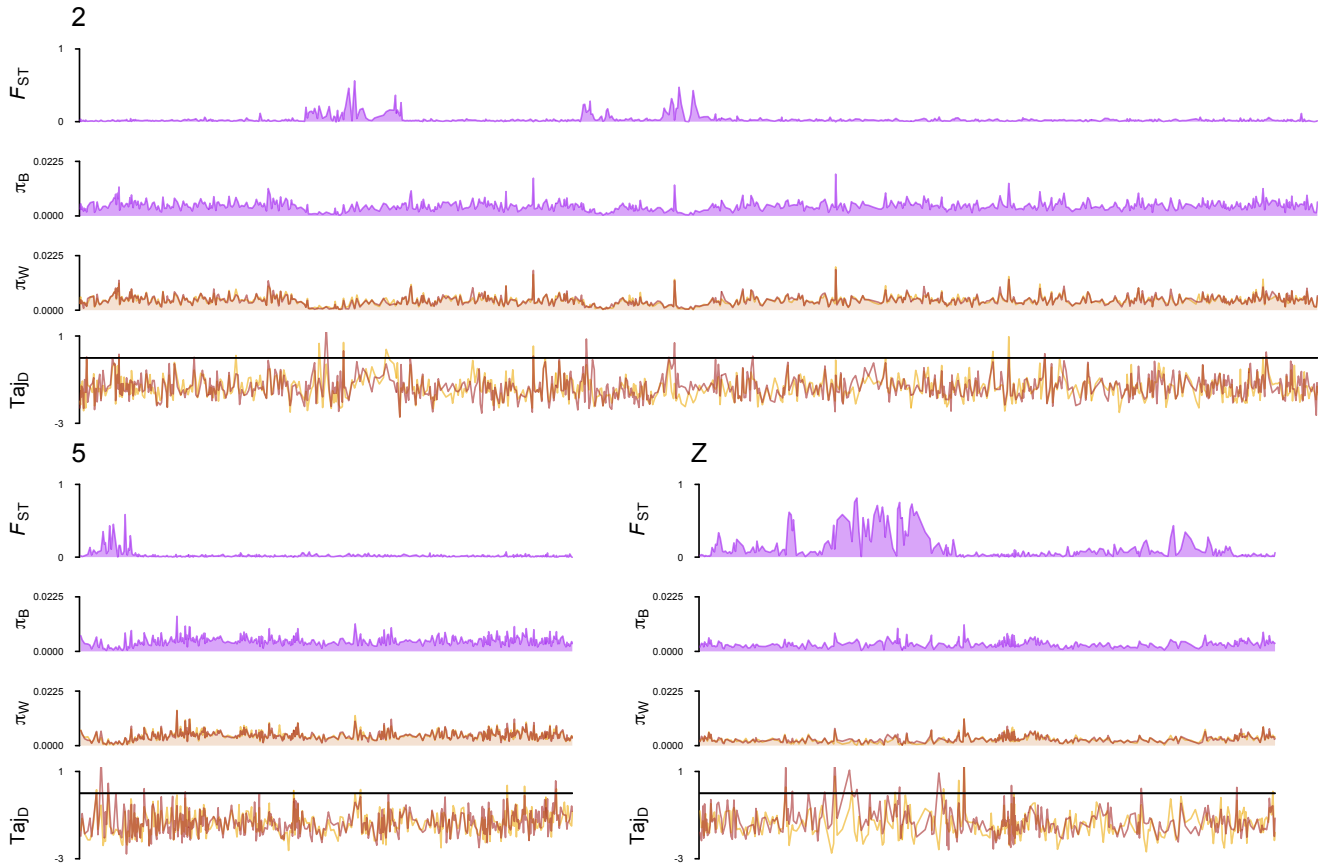


Figure 5. Patterns of genetic variation comparing allopatric yellowhammers (n = 53) and allopatric pine buntings (n = 42) across three chromosomes (2, 5 and Z). Relative nucleotide differentiation (F_{ST}), absolute between-population nucleotide differentiation (π_B), absolute within-population nucleotide 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.

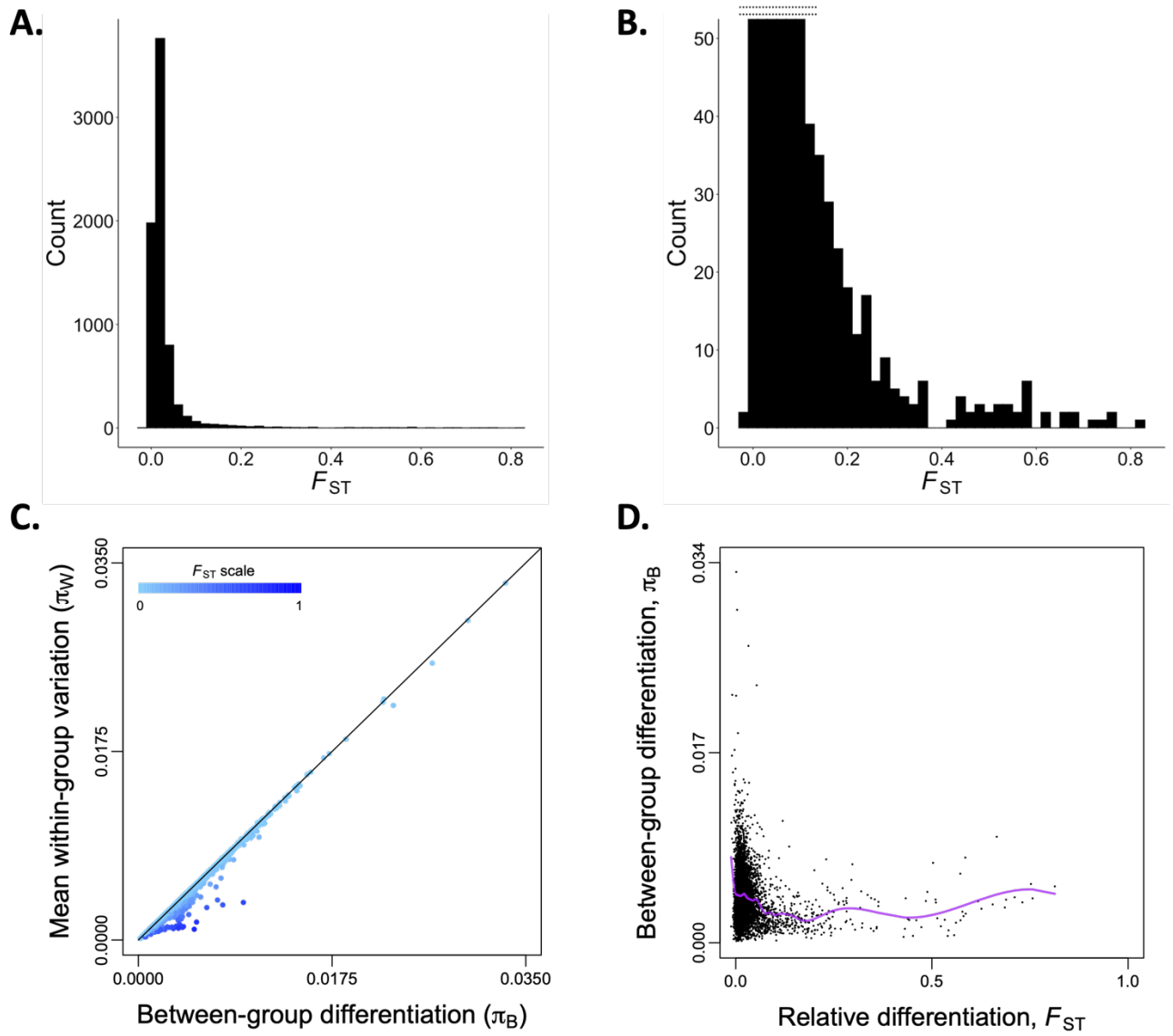


Figure 6. A) A histogram of average relative differentiation (F_{ST}) values calculated for 2000 bp windows across the nuclear genome when comparing allopatric yellowhammers (n=53) with allopatric pine buntings (n=42). **B)** A truncated version of Figure 6A that depicts the high F_{ST} tail of the histogram. The limits of the y-axis were set as (0,50) with the double dotted line indicating bars of the histogram that have been truncated. **C)** 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 2000 bp window of sequenced data 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. **D)** Association between relative differentiation and absolute between-group nucleotide differenation 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.