Cytonuclear discordance in the crowned-sparrows, *Zonotrichia atricapilla* and *Zonotrichia leucophrys*: a mitochondrial selective sweep?

Rebecca S. Taylor^{a,1}, Ashley C. Bramwell^{a,2}, Rute Clemente-Carvalho^a, Nicholas A. Cairns^a, Frances Bonier^a, Stephen C. Lougheed^{a,*}

^a Department of Biology, Queen's University, Kingston, Ontario, K7L 3N6, Canada

*Corresponding author at: Department of Biology, Queen's University, Kingston, Ontario, Canada.

E-mail address: steve.lougheed@queensu.ca (S. C. Lougheed).

¹ Present address: Biology Department, Trent University, Peterborough, Ontario, K9J 7B8, Canada

² Present address: Department of Biological Sciences, University of Toronto Scarborough, 1265 Military Trail, Toronto, Ontario M1C 0B1

ABSTRACT

The golden-crowned (*Zonotrichia atricapilla*) and white-crowned (*Z. leucophrys*) sparrows have been presented as a compelling case for rapid speciation. They display divergence in song and plumage with overlap in their breeding ranges implying reproductive isolation, but have almost identical mitochondrial genomes. Previous research proposed hybridization and subsequent mitochondrial introgression as an alternate explanation, but lacked robust nuclear gene trees to distinguish between introgression and incomplete lineage sorting. We test for signatures of these processes between *Z. atricapilla* and *Z. leucophrys*, and investigate the relationships among *Z. leucophrys* subspecies, using mitochondrial sequencing and a reduced representation nuclear genomic dataset. Contrary to the paraphyly evident in mitochondrial gene trees, we confirmed the reciprocal monophyly of *Z. atricapilla* and *Z. leucophrys* using large panels of single nucleotide polymorphism (SNPs). The pattern of cytonuclear discordance is consistent with limited, historical hybridization and mitochondrial introgression, rather than a recent origin and incomplete lineage sorting between recent sister species. We found evidence of nuclear phylogeographic structure within *Z. leucophrys* with two distinct clades. Altogether, our results support the true species status of *Z. atricapilla* and *Z. leucophrys*, and indicate deeper divergences between the two species than inferred using mitochondrial markers. Our results demonstrate the limitations of relying solely on mitochondrial DNA for taxonomy, and raise questions about the possibility of selection on the mitochondrial genome during temperature oscillations (e.g. during the Pleistocene). Historical mitochondrial introgression facilitated by past environmental changes could cause erroneous dating of lineage splitting in other taxa when based on mitochondrial DNA alone.

Keywords:

crowned-sparrows, ddRADSeq, mitochondrial introgression, incomplete lineage sorting, speciation

1. Introduction

Speciation is typically construed as an incremental process that requires millions of years to complete. For example, in sister species of birds, divergence in allopatry followed by secondary sympatry where the two taxa remain reproductively isolated despite possibilities for interbreeding, typically requires separation of at least 1.7 million years (Weir and Price, 2011). Against this backdrop, examples of very rapid speciation (hundreds of thousands rather than millions of years) can help to illuminate the ecological and genetic factors that underpin diversification and the evolution of reproductive isolation. There are various examples of purported rapid speciation in birds including goldfinches (*Carduelis*; Arnaiz-Villena et al., 1998), canaries (*Serinus*; Arnaiz-Villena et al., 1999), Darwin's finches (Geospizinae; Grant, 1999), red crossbills (*Loxia curvirostra* complex; Benkman, 2003), indigobirds (*Vidua*; Sorenson et al., 2003), pheasants (Shen et al., 2010), and the southern capuchino seedeaters (*Sporophila*; Campagna et al., 2012). Rapid speciation in birds is typically accompanied by well-developed prezygotic isolating mechanisms (e.g. based on plumage, songs), despite sister species being genetically similar across most of their genome (Edwards et al., 2005). For example, the southern capuchinos comprise nine broadly sympatric species found in grasslands of South America that exhibit marked differentiation in male plumage and song, yet show virtually no molecular divergence, implying very recent origins within the second half of the Pleistocene (Campagna et al., 2012). Genomic evidence suggests that this remarkable diversification of male plumage is underlain by selection on regulatory elements of genes involved in melanogenesis (Campagna et al., 2017). The

rapidity of divergence from a common ancestor in all of these avian systems means that intrinsic postzygotic isolation may evolve relatively more slowly making hybridization possible, and that prezygotic isolation may arise before we see complete reciprocal monophyly.

Golden-crowned (*Zonotrichia atricapilla*) and white-crowned (*Z. leucophrys*) sparrows have been touted as an example of exceedingly rapid speciation, perhaps splitting as recently as 50,000 years ago (Johnson and Cicero, 2004). The evidence for this is not unequivocal. Early genealogies based on allozyme data (Zink, 1982) and mitochondrial DNA (mtDNA) (Zink et al., 1991) implied that *Z. atricapilla* and *Z. leucophrys* were independently-evolving sister species. However, subsequent analyses of mtDNA sequences suggested that they are not reciprocally monophyletic, even sharing some haplotypes (Weckstein et al., 2001; McKereghan-Dares, 2008;), and that their low global mitochondrial nucleotide diversity is comparable to a single, homogenous avian population rather than two well diverged species. Weckstein et al. (2001) presented alternate hypotheses to explain this pattern: 1) *Z. atricapilla* and *Z. leucophrys* originated in allopatry, followed by contact, historical hybridization and subsequent introgression, with ultimate replacement of an one ancestral mitochondrial genome in one of the species with that of the other (mtDNA capture; see Good et al., 2008); or 2) these two species originated very recently, with insufficient time for complete lineage sorting to have occurred.

The white-crowned sparrow has a geographic range spanning most of North America, while the golden-crowned is generally confined to regions west of the Rocky Mountains (Cortopassi and Mewaldt, 1965; Fig. 1). Both species are migratory, breeding in the northern parts of their ranges and overwintering in the south, although *Z. leucophrys* have some year-round resident populations along the coast of Washington, Oregon and California. Adults are distinct in their plumages, breeding behaviours, and vocalizations (del Hoyo et al., 2011).

Playback experiments have shown that both sexes of *Z. leucophrys* react more strongly to conspecific than heterospecific songs (Pleszczynska, 1979), and to natal versus non-natal vocal dialects (Baker et al., 1982). Female *Z. l. nuttallii* mate assortatively with males of their natal dialect population (Tomback and Baker, 1984). Finally, while some reports of *Z. atricapilla*-*Z. leucophrys* hybrids exist (Miller, 1940; Morton and Mewaldt, 1960; McCarthy, 2006), these represent a comparatively small number given the extensive overlap in their breeding distributions across British Columbia, the Yukon, and Alaska (Cortopassi and Mewaldt, 1965). All existing evidence thus suggests that these are good biological species.

Within *Z. leucophrys* there are five named subspecies (*Z. l. nuttallii*, *Z. l. pugetensis*, *Z. l. gambelii*, *Z. l. oriantha* and *Z. l. leucophrys*). A recent genomic study of nuclear DNA investigated population structure across a contact zone of *Z. l. nuttallii*, and *Z. l. pugetensis*. Differentiation was found between the two subspecies, driven by differences in advertising song, indicating that they are separate evolutionary significant units (Lipshutz et al., 2017). However, secondary contact has led to ongoing hybridization between the two subspecies (Lipshutz et al., 2017; Toews, 2017). Genetic relationships based on nuclear DNA among the other subspecies are largely unknown.

The five crowned-sparrow species (golden-crowned and white-crowned sparrows along with Harris' sparrow, *Z. querula*, white-throated sparrow, *Z. albicollis*, and the rufous-collared sparrow, *Z. capensis*) have contributed significantly to our current understanding of speciation with research on reproductive isolation (e.g. Danner et al., 2011), mating behaviour (e.g. Tomback and Baker, 1984; Houtman and Falls, 1994; Lipshutz et al., 2017), phenology (e.g. Handford, 1980; Moore et al., 2005), migration (e.g. Landys et al., 2004; McFarlan et al., 2009; Price et al., 2010), vocal dialects, song learning, and function (e.g. Handford and Nottebohm, 1976; Loughheed and Handford, 1992; MacDougall-Shackleton and MacDougall-Shackleton, 2001; Shizuka, 2014; Toews, 2017), and phylogeography,

geographic range expansions, and secondary contact zones (Lougheed et al., 2013; Campagna et al., 2014). Despite their prominence in evolutionary and ecological studies, there is phylogenetic uncertainty about the relationships among the species (Carson and Spicer, 2003).

Here, we use a genome-wide panel of nuclear Single Nucleotide Polymorphisms (SNPs) combined with mitochondrial DNA sequencing to investigate relationships between *Z. atricapilla* and *Z. leucophrys* to test the alternate propositions presented by Weckstein et al. (2001). If the species originated, as in the first hypothesis, with allopatric divergence and the introgression of mitochondrial DNA from one species into another, we expect signatures of genetic differentiation and monophyly in the nuclear DNA. Conversely, under the second hypothesis with recent speciation we predict low genetic differentiation and a lack of reciprocal monophyly in the nuclear DNA. We include samples from all subspecies within *Z. leucophrys*, and from all five crowned sparrow species so that we can assess the phylogenetic affinities of these taxa.

2. Materials and methods

2.1. Sampling

From museum collections and our own sample archive, we assembled 73 tissue and blood samples of *Z. atricapilla* and *Z. leucophrys* from across their respective geographic ranges, including individuals from all five subspecies of *Z. leucophrys* (Fig. 1; Table 1) for ddRAD sequencing. For mtDNA sequencing, 74 samples were used (Table 1). We added eight specimens from the remaining three *Zonotrichia* species (two from *Z. albicollis*, three from *Z. capensis*, two from *Z. querula*; Table 1; Fig. S1). Tissue samples were preserved in

lysis buffer (4.0M urea, 0.2M NaCl, 10mM EDTA, 0.5% n-Lauroyl Sarcosine, 100mM Tris-CHl; pH 8.0) or 99% ethanol until extraction.

2.2. DNA extraction

Total genomic DNA was extracted either using the DNeasy® Blood and Tissue Kit (QIAGEN, Mississauga, Ontario, Canada) following the manufacturer's protocols, or a standard phenol-chloroform method (Sambrook & Russell, 2001). For ethanol preserved samples, excess ethanol was allowed to evaporate before extraction. Samples fixed in lysis buffer were washed in a 1x phosphate-buffered saline (PBS; 10mM, pH 7.4) solution. All newly extracted samples were treated with RNase A (10 mg/mL) and incubated at 37°C for 30 minutes. We checked the purity of our DNA using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies Inc., Wilmington, DE, USA). If needed, due to low concentration or purity, samples were re-precipitated using a standard ethanol precipitation protocol with sodium acetate (3M, pH 5.2) as the carrier molecule, and then re-suspended in microfiltered autoclaved water and stored at -20°C until subsequent analyses. We quantified the concentration of double-stranded DNA in each sample using a Qubit® 2.0 Fluorometer (Life Technologies, Grand Island, NY, USA).

2.3. Mitochondrial DNA sequencing

Mitochondrial DNA sequence variation was assayed from control region (approximately 413bp) and cytochrome oxidase I (COI, 666bp) using the primers, PCR cocktail recipes, and cycling parameters given in Loughheed et al. (2013; primers: LCA1271-R and ZNGluF3) and Kerr et al. (2009; primers BirdF1 and COIbirdR2), respectively. The

amplification products were visualized on 1.0% agarose gels using RedSafe™ Nucleic Acid Staining Solution, and purified using AMPure XP 2X Solid Phase Reversible Immobilization (SPRI) following the manufacturer's instructions. Purified products were sent to The Centre for Applied Genomics, SickKids Hospital, Toronto, ON, Canada) for sequencing on an ABI 3730XL using the BigDye™ terminator kit. Mitochondrial sequences will be deposited in Genbank upon acceptance.

2.4. Double-digest restriction associated DNA (ddRAD) sequencing

Following Poland et al. (2012) double-digest RAD-Seq libraries were constructed at the Institute for Integrative Systems Biology (IBIS; at Laval University) using the restriction enzyme pairing of SbfI and MspI. Uniquely barcoded samples were pooled into a library and then size selected for roughly 100-300 bp using a BluePippin® DNA size selection system (Sage Science Inc., Beverly, MA, USA). The library was forwarded to Génome Québec (Génome Québec Innovation Centre) and assessed using a LabChip GX (Perkin-Elmer, MA, USA) instrument to determine library size quality. The library was then sequenced as single-end, 100 base-pair reads on a single Illumina HiSeq2000 lane (Illumina, San Diego, CA, USA). All ddRADseq data will be archived on the Dryad Digital Repository upon acceptance.

2.5. ddRADseq read filtering

Raw reads were de-multiplexed sequentially by barcode length using Stacks 2.3 (Catchen et al., 2013). Low quality reads (with a Phred score below 10), and those with uncalled bases, were filtered out and reads were truncated to 92bp. De-multiplexed reads

were aligned to the *Zonotrichia albicollis* genome from NCBI (GCF000385455, Bioproject PRJNA197293) using the ‘very_sensitive’ setting in Bowtie2 (Langmead and Salzberg, 2012). Resulting SAM files were converted to BAM files and sorted using Samtools 1.1 (Li et al., 2009). We then used the Refmap and Populations modules in Stacks randomly selecting one SNP per locus, filtering out loci with a minor allele frequency of less than 0.05. After an initial run, seven individuals were found to have high levels of missing data and were removed, leaving 66 individuals (Table 1). We re-ran Refmap and Populations without the low-quality samples using two filtering settings, one allowing for 10% missing data (all individuals as one population and setting r to 0.9) and the other without any missing data (setting r to 1.0). We also did the same but without the outgroup taxa to create input files with only *Z. atricapilla* and *Z. leucophrys*, leaving 58 individuals (18 *Z. atricapilla*, 18 *Z. l. nuttalli*, 9 *Z. l. pugetensis*, 7 *Z. l. gambelli*, 4 *Z. l. oriantha*, and 2 *Z. l. leucophrys*). All analyses were conducted with both datasets (with no missing data and with 10% missing data).

2.6. mtDNA network and phylogenetics

We used the concatenated COI and control region sequences (total 1211 bp) to make a statistical parsimony haplotype network in PopArt (Clement et al., 2000; Leigh and Bryant, 2015; PopArt: <http://popart.otago.ac.nz>) using individuals from *Z. leucophrys* and *Z. atricapilla*. We performed Bayesian phylogenetic analyses on a dataset comprising all individuals and a dataset with only individual haplotypes identified in our network analysis. The best partitioning scheme and models of evolution for each dataset were selected using PartitionFinder2 (Lanfear et al., 2016). For the full dataset the best partition scheme and associated models of evolution were: HKY+I for control region, HKY+I for COI codon

position 1, F81 for COI codon position 2, and HKY+I for COI codon position 3. For the haplotype dataset the best partition scheme and associated of models of evolution were: HKY+I for control region, HKY+I for COI codon position 1, F81 for COI codon position 2, and HKY for COI codon position 3.

We inferred phylogenetic relationships using the full and haplotype mtDNA datasets using Bayesian inference with MrBayes v3.2.6 x64 (Huelsenbeck and Ronquist, 2001, Ronquist et al., 2012). For all models, partitions were unlinked for substitution rates and proportion of invariant sites (I). allowing parameter values to vary independently. Each Bayesian analyses included two simultaneous runs each with random starting trees, one cold and three incrementally heated Markov chains, and default priors for all parameters. Each analysis was allowed to run until the standard deviation of split frequencies < 0.01, sampling every 100 generations (full dataset: 2 million generations; haplotype dataset: 1 million generations). We visually assessed that our runs had achieved stationarity by evaluating the likelihoods in the program Tracer version 1.7.1 (Rambaut et al., 2018). For each analysis, we constructed a 50% majority rule consensus tree (discarding the first 25% of trees as burnin).

2.7. Phylogenomic reconstruction using SNPs

Using the PHYLIP files produced in Stacks, we ran RAxML version 8 (Stamatakis, 2014) to generate a maximum likelihood tree. We used a General Time Reversible model with a gamma distribution of substitution rates and a proportion of invariant sites (GTRGAMMAI), and used the rapid bootstrap algorithm using 1000 replicates. We ran the analysis with all species, but because of the long branch lengths (see Results), we also ran the analysis using *Z. querula* as an outgroup with only *Z. atricapilla* and *Z. leucophrys* individuals to make relationships clearer.

We ran Treemix 1.13 (Pickrell and Pritchard, 2012) to infer a maximum likelihood tree and to investigate any migration between *Z. atricapilla* and/or *Z. leucophrys* sub-species. We ran Treemix from 0-9 migration events, running it without grouping SNPs for three iterations, and grouping SNPs into windows of 20 for three iterations to account for possible linkage and because the OptM package (below) needs different likelihood scored between iterations. We plotted the results in RStudio 1.0.136 (RStudio Team, 2015) and used the R package OptM (available here: <https://cran.r-project.org/web/packages/OptM/index.html>) to calculate the ad hoc statistic delta M, which is the second order rate of change in the log-likelihood of the different migration events, to help infer how many migration events to visualize.

2.8. Population structure

All population structure analyses were done using a dataset comprising only *Z. atricapilla* and *Z. leucophrys*. We ran an ADMIXTURE analysis (version 1.3; Alexander et al., 2009) for population assignment. To do this, we converted the VCF file into a BED file using Plink 1.9 (Purcell et al., 2007). We pruned the dataset to remove any SNPs with a correlation coefficient (R^2) of 0.1 or greater. We then ran ADMIXTURE from K 1 to 8, and plotted the results in RStudio. We also did a principal component analysis in RStudio using the packages ADEGENET (Jombart, 2008) and vcfR (Knaus and Grünwald, 2017). We did an additional PCA on a subset of data with *Z. l. nuttalli* and *Z. l. pugetensis* only, and another PCA on *Z. l. gambelli*, *Z. l. oriantha*, and *Z. l. leucophrys* only to investigate possible substructure (see Results). We calculated pairwise F_{st} between *Z. atricapilla* and all *Z. leucophrys* subspecies, as well as between the three major lineages uncovered in the analyses (see Results) in RStudio using the HIERFSTAT package in R (Goudet, 2005) and the Weir

and Cockerham calculation (WC84; Weir and Cockerham, 1984).

3. Results

3.1. ddRADseq data quality

We recovered 18.4 million single-end Illumina reads. After removing individuals with low quality reads, we had 3,469 SNPs in the VCF file with no missing data and 7,096 SNPs in the VCF file allowing 10% missing data for the datasets with all five crowned sparrow species, and 4,139 SNPs in the VCF file with no missing data and 6,626 SNPs in the VCF file allowing 10% missing data in the files with *Z. atricapilla* and *Z. leucophrys* only. The depth was high for all retained individuals, ranging from 41.6 – 183.1X (aside from one *Z. capensis* individual with 481.4 X) for the file with outgroups and no missing data, and between 39.0 – 183.3X (aside from one *Z. capensis* individual which has 464.1 X) in the file allowing 10% missing data. These datasets were used for the RAxML analyses. Similarly, for the dataset comprising *Z. atricapilla* and *Z. leucophrys* only the depth per individual ranged from 42.7 – 183.0X for the file with no missing data and between 39.7 – 169.3 X for the file allowing 10% missing data. These datasets were used for all subsequent nuclear SNP analyses, however all results presented are from the analyses with no missing data as all results were the same.

3.2. Evolutionary relationships apparent from analysis of mitochondrial DNA

The most obvious feature of both the mtDNA network (Fig. 2) and our Bayesian tree from analysis of distinct mtDNA haplotypes (Fig. 3) or of all individuals (Fig. S2) is that *Z.*

atricapilla and *Z. leucophrys* share mtDNA haplotypes and do not form distinct, reciprocally monophyletic clusters/clades. We also find no obvious phylogeographic structure among *Z. leucophrys* subspecies, although some haplotypes are unique to particular subspecies. *Zonotrichia albicollis* and *Z. querula* comprise sister lineages, that are in turn, sister to the *Z. atricapilla*/*Z. leucophrys* clade. The North American crowned sparrow clade is reciprocally monophyletic with the rufous-collared sparrow, *Z. capensis*.

3.3. Interspecific relationships and population structure in *Zonotrichia* with nuclear DNA

In contrast to our results from analysis of mtDNA, *Z. atricapilla* and *Z. leucophrys* are well-supported, reciprocally monophyletic sister taxa in the RAxML phylogenomic analysis (Fig. 4A and 4B). The *Z. leucophrys* lineage was further divided into two clades: *Z. l. nuttallii* and *Z. l. pugetensis* (NP), and *Z. l. gambelii*, *Z. l. oriantha*, and *Z. l. leucophrys* (GOL). *Zonotrichia queula* was sister to *Z. atricapilla* and *Z. leucophrys*, followed by *Z. albicollis*, with *Z. capensis* recovered as sister to all North American congeners (Fig. 4A and 4B).

The Treemix maximum likelihood tree was similar, with *Z. atricapilla* well separated from all *Z. leucophrys* subspecies, and showing a large drift parameter (Fig. 5A). The NP clade remains distinct from the GOL subspecies, and shows evidence of some drift (Fig. 5A). Treemix stopped adding migration events after 5, and OptM indicated that one migration event which shows migration between the ancestor of the NP clade into *Z. l. leucophrys* (Fig. 5B; Fig S3). When adding three migration events, the analysis did infer migration from *Z. atricapilla* into *Z. l. leucophrys* (Fig. 5D); however adding three migration events did not increase the variance explained compared with adding one or two migration event by a lot (Fig S3).

The ADMIXTURE analysis for $K = 2$ (which had the lowest cross validation error) shows *Z. atricapilla* as a different population from *Z. leucophrys*, although with some *Z. atricapilla* ancestry apparent in the GOL clade (Fig. 6A). The results for $K = 3$ split the three clades seen in the phylogenetic analyses (*Z. atricapilla*, NP, and GOL) with a small amount of GOL ancestry showing in the NP group, especially in *Z. l. pugetensis* individuals (Fig. 6B). For $K = 4$ we see some individuals within the GOL clade splitting off, and for $K = 5$ we see *Z. l. pugetensis* splitting off with admixed individuals in *Z. l. nuttalli* (Fig. 6C and 6D). The PCA reflected the same pattern, with the same three major groups, *Z. atricapilla*, NP, and GOL, well separated clusters (Fig. 7A). Results from the PCA on only individuals from the NP clade, shows *Z. l. nuttalli* and *Z. l. pugetensis* to be separate clusters (Fig. 7B). Similarly, when analyzing individuals from the GOL clade alone, we see *Z. l. gambelli*, *Z. l. oriantha*, and *Z. l. leucophrys* comprising unique clusters (although sampling is limited for *Z. l. leucophrys*; Fig. 7C). Pairwise F_{ST} between *Z. atricapilla* and all *Z. leucophrys* subspecies were high ranging from 0.27 to 0.31 (Table 2). The values were high between *Z. l. nuttalli* and all other subspecies except *Z. l. pugetensis*, and were low between the subspecies within the GOL clade (Table 2). Pairwise F_{ST} values between the three major clades (*Z. atricapilla*, NP, and GOL) exceed 0.10 (Table 3).

4. Discussion

4.1. Mitonuclear discordance in crowned sparrows

Our study is the first to investigate the phylogenetic relationships among the five *Zonotrichia* species using Next Generation Sequencing data. Contrary to inferences using mtDNA exclusively that indicated paraphyly between white-crowned and golden-crowned

sparrows (Fig. 2 and 3) (Weckstein et al., 2001; McKereghan-Dares, 2008), we find *Z. atricapilla* and *Z. leucophrys* comprise well-supported, reciprocally-monophyletic sister species, with all five *Zonotrichia* species monophyletic (Fig. 4A and 4B). *Zonotrichia atricapilla* and *Z. leucophrys* comprise distinct clusters in ADMIXTURE and PCA (Fig. 6 and 7) with high F_{ST} values between *Z. atricapilla* and all *Z. leucophrys* subspecies (Tables 2 and 3). Our results suggest that *Z. atricapilla* and *Z. leucophrys* are more deeply diverged in their nDNA than in their mtDNA, and that their origins probably markedly predate the previously reported 50,000 years given the reciprocal monophyly of the species using genome wide SNP data. Consistent with Weckstein et al. (2001), our analyses imply that mitochondrial introgression resulting from hybridization is the primary cause of discordance between the two genomes for the *Z. atricapilla*-*Z. leucophrys* clade, and this strong discordance between nuclear and mitochondrial DNA generally rules out incomplete lineage sorting (Toews and Brelsford, 2012). We also found genetic structuring within *Z. leucophrys*, uncovering two reciprocally-monophyletic phylogenetic groups (Fig. 4A and 4B), which form distinct clusters in ADMIXTURE and PCAs (Fig. 6 and 7).

4.2. Causes of mitonuclear discordance

Because the vertebrate mitochondrial genome has an effective population size that is approximately one-quarter that of nDNA (Funk and Omland, 2003), we expect the rate of stochastic lineage sorting to be commensurably faster in mtDNA, and thus to exhibit differentiation more quickly in good, reproductively-isolated species (Funk and Omland, 2003). However, when hybridization and subsequent mitochondrial introgression has occurred then we can see the opposite pattern. For example, if female hybrids backcross with males of their paternal lineage over multiple generations this could result in populations

where individuals bear the paternal species' nuclear genome, and the maternal species mtDNA (Funk and Omland, 2003; see Weckstein et al., 2001).

For taxa where the fitness of heterogametic hybrids is significantly reduced, the Z (or X) chromosome is largely thought to be responsible (Haldane's rule; Haldane, 1922; see Schilthuizen et al., 2011). This can promote speciation by preventing hybridization and fusion of incipient species ("large-Z effect"; Turelli and Orr, 1995), and thus limit opportunities for mtDNA to cross species boundaries via heterogametic hybrids (see Rheindt and Edwards, 2011). However, absolute inviability in hybrids of incipient bird species is relatively rare, and it generally takes a minimum of five million years for complete post-zygotic isolation in birds to evolve (Lijtmaer et al., 2003). As natural hybrids between these *Z. atricapilla* and *Z. leucophrys* are so rare, we have little information on their fitness in the wild.

Weckstein et al. (2001) hypothesized that, at one point in their histories, hybridization took place between *Z. leucophrys* and *Z. atricapilla* with subsequent mitochondrial introgression from one into the other. Our Treemix analysis did indicate an introgression event from *Z. atricapilla* into *Z. l. leucophrys* when adding three migration events (Fig. 5D). Under this scenario, hybridization followed by continued backcrossing of hybrids would explain the discordance between the *Zonotrichia* mitochondrial gene and species trees (Zink et al., 1991; Weckstein et al., 2001). The originally reported divergence time estimate of 50,000 years thus reflects a period of historical hybridization that essentially "reset" the mitochondrial molecular clock for the two species (see Rheindt and Edwards, 2011).

Recent evidence suggests mitochondrial introgression may be more common than previously appreciated, occurring across a wide range of taxa (Toews and Brelsford, 2012). Several processes can facilitate mtDNA introgression once hybridization has occurred. Neutral processes, such as small population sizes and genetic drift (for example during

bottlenecks or founder events), or large disparities in population sizes upon secondary contact, can facilitate mitochondrial introgression (Ballard and Whitlock, 2004; Toews and Brelsford, 2012). For example, mitochondrial introgression between chipmunk species in the genus *Tamias* is thought to have occurred due to low population densities during climate induced range shifts (Good et al., 2008), and differential population sizes may have promoted mitochondrial introgression in *Neodiprion* sawflies (Linnen and Farrell, 2007).

Mitochondrial introgression can also be caused by selection (i.e. adaptive introgression; Toews and Brelsford, 2012). Increasing evidence highlights the importance of selection on the mitochondrial genome, contradicting the long-held assumption that much of the variation in mtDNA can be regarded as neutral (Ballard and Whitlock, 2004; Toews and Brelsford, 2012). For example, selection on the mitochondrial genotype (or perhaps other cytoplasmic factors) is likely to have caused the introgression of mtDNA between *Drosophila teissieri* and the *D. yakuba*/*D. santomea* species pair (Bachtrog et al., 2006). Similarly, a mitochondrial selective sweep is thought to have caused mitonuclear discordance in yellowhammers (*Emberiza citrinella*) and pine buntings (*E. leucocephalos*; Irwin et al., 2009). Examples of adaptive introgression often invoke thermal adaptation as an explanation, due to the key role of mtDNA in efficient energy production and metabolism (Ballard and Whitlock, 2004; Lamb et al., 2018). For example, mitochondrial introgression between *Lepus* hare species (Melo-Ferreira et al., 2005), and between arctic char and lake trout (*Salvelinus alpinus* and *S. namaycush*; Wilson and Bernatchez, 1998), are potentially due to temperature adaptation. Mitochondrial variation in multiple Australian bird species is associated with historical climatic variation (Lamb et al., 2018).

It is unclear whether the mitochondrial introgression between *Z. atricapilla* and *Z. leucophrys* was caused by neutral or selective processes, or a combination thereof. The Pleistocene epoch was characterized by repeated cycles of global cooling and warming, with

the most rapid oscillations occurring from 60,000 to 25,000 years before present (Adams, 1997). In the northern hemisphere, colder periods resulted in a series of extensive glacial advances. At their maximum extent Arctic ice sheets covered most of present-day Canada (Adams, 1997). Cooling also compressed and pushed temperate and tropical climates towards the equator, and promoted the expansion of desert habitats (Hewitt, 2004).

While it is only informed speculation, during the Pleistocene populations of *Z. atricapilla* and *Z. leucophrys* could have been forced into common refugia by ice sheets and shifting vegetation, leading some sexually mature adults to seek out heterospecific mates. If the newly-introgressed mitochondrial haplotype conferred favorable metabolic or physiological traits, it could have become fixed throughout populations of both species due to adaptive introgression. Several rounds of bottlenecking due to climate change might have occurred during this period and similarly caused the fixation of one haplotype in both species, or acted in combination with selection to augment this process.

If effective population sizes were small, and the young species were defined by relatively few variable sites, the re-assortment of nuclear loci could have occurred quickly to erase traces of hybridization events. Premating isolation mechanisms may have also been well-developed to prevent the fusion of the species. Behavioural reproductive isolation through conspecific song recognition is well-documented in *Zonotrichia* species, with juveniles predisposed to learn their own species' songs (e.g. Shizuka, 2014). The distinct coloration of their crowns might have further distinguished them, but the importance of visual signaling in species recognition or mate selection remains to be explored in crowned sparrows.

The time required for biological speciation is taxon-dependent, with some species able to interbreed even after more than 20 million years of divergence (Lijtmaer et al., 2003; Coyne and Orr, 2004). Given that little gene flow is needed for mtDNA to move between

species successfully (Takahata and Slatkin, 1984), and complete reproductive isolation can take millions of years to develop, mitochondrial introgression could be far more common in birds than previously realized. Literature surveys suggest that 14-16% of bird species exhibit mitochondrial paraphyly, and it is estimated that at least 5.7% of these cases are due to introgression (Funk and Omland, 2003; McKay and Zink, 2010).

4.3. Implications of mitochondrial introgression

An increasing number of studies are revealing the importance of selection on the mitochondrial genome (Toews and Brelsford, 2012). Despite this, the drivers of mitochondrial introgression are not well-studied in natural systems. Future work should investigate the functional differences caused by different mitochondrial haplotypes, as was done in yellow-rumped warbler species (*Setophaga* spp.; Toews et al., 2014). In addition, the co-adaptation of mitochondrial and nuclear genes is of key importance for fitness (Hill, 2017). Cases such as the *Zonotrichia* sparrows, where one mitochondrial haplotype was completely replaced by another, indicate that in some instances the cytonuclear interactions are weaker than the strength of selection on the mtDNA (Bachtrog et al., 2006).

Regardless of whether mitochondrial introgression is caused by neutral or selective processes, there are implications in the use of mtDNA alone for inferring species histories. For example, the use of both nuclear and mitochondrial DNA in Asian box turtles (*Cuora* spp.) allowed the recognition of monophyletic taxon, which had been masked by complicated patterns of mitochondrial introgression, with strong implications for the choice of individuals to use in captive breeding programs (Spinks and Shaffer, 2007). In addition, there are obvious implications for taxonomy, with introgression confounding phylogenetic inference of species (e.g. crotaphytid lizards; McGuire et al., 2007), or masking hidden genetic structuring as we

have found both between *Z. atricapilla* and *Z. leucophrys*, and among *Z. leucophrys* subspecies.

4.4. Relationships between the five *Zonotrichia* species and the *Z. leucophrys* subspecies

In both the mitochondrial and nuclear phylogenetic reconstructions (Fig. 3 and 4), *Z. capensis* is the most distinct and sister to all other *Zonotrichia* species. However, we see different patterns between the nuclear and mitochondrial phylogenies with *Z. querula* and *Z. albicollis* which are recovered as sister species in the mitochondrial DNA tree (Fig. 3), but *Z. albicollis* is basal to *Z. querula* sister in nuclear DNA (Fig. 4). To resolve the relationship, it may be necessary to increase the sampling of these two species to increase the power of the analyses.

We found well-supported monophyletic clades within *Z. leucophrys* that divide the subspecies into two groups (GOL and NP; Fig. 4, 6, and 7). Further, we found support for genetic differences between *Z. l. nuttallii* and *Z. l. pugetensis* (Fig. 4A, 4B, and 7B), which is consistent with the differentiation found by Lipshutz et al. (2017), although we do not find them to be completely reciprocally monophyletic with one *Z. l. pugetensis* individual recovered within the *Z. l. nuttallii* clade (Fig. 4A and 4B). We likely find stronger structuring and less evidence of hybridization between *Z. l. nuttallii* and *Z. l. pugetensis* in our study because we did not use individuals from within or near to the contact zone (Fig. 1). Within the GOL clade, however, the subspecies are not recovered as distinct (Fig. 4A and 4B) with low F_{ST} values among three subspecies (Table 2). We do find that PCA separated them (Fig. 7C) suggesting some weak differentiation.

5. Conclusions

Our findings support previous assertions of historical mitochondrial introgression between the two sister species, and further emphasize the need to re-evaluate estimates of age of divergence. For example, future studies with more data could re-estimate divergence times using the relative branch lengths within the *Zonotrichia* phylogeny, as the age of other *Zonotrichia* species (e.g. *Z. capensis*) are better established (Lougheed et al., 2013). More work should also focus on better delineating the genealogical patterns of the *Z. leucophrys* subgroups. These provide an important opportunity to study different stages of the speciation process *in situ*. Finally, we show a clear need to incorporate both nuclear and mitochondrial genomes when investigating avian evolutionary histories. For example, McKay and Zink (2010) found that in 21.3% of cases of paraphyly in birds, introgression could not be reliably distinguished from incomplete lineage sorting (McKay and Zink, 2010). Mitochondrial DNA is effectively a single locus and lacks the resolution needed to discern close relationships and cannot on its own reveal signatures of hybridization. Genome-wide nDNA sequencing is more powerful, provides better resolution, and allows us to make stronger inferences about the histories of many taxa. Ultimately cases of mitonuclear discordance tell us something important about both the history of species (such as *Zonotrichia* sparrows) and about mitochondrial evolution. We must consider mtDNA as simply a ‘tool’ for phylogeography (Ballard and Whitlock, 2004).

6. Author statement

RST conducted analyses and contributed to the writing of the manuscript; ACB conducted lab work and contributed to the analysis and writing of the manuscript; NAC contributed to the writing of the manuscript; FB contributed to the writing of the manuscript

and assisted with procuring samples, RC-C conducted lab work and assisted with writing of the manuscript. SCL conceived the study, did some phylogenetic analysis, and contributed to the writing of the manuscript.

Acknowledgements

We thank John Wingfield (University of California, Davis); Dai Shizuka (University of Nebraska); the Burke Museum of Natural History and Culture; the Royal Alberta Museum; the Royal Ontario Museum; the University of Alaska Museum; Donna Maney (Emory University); Beth MacDougall-Shackleton (Western University); Paul Martin (Queen's University); and the Louisiana State University Museum of Natural History for contributing tissue samples. Brian Boyle (Genomic Analysis Platform, Laval University) and Génome Québec Innovation Centre performed the library preparation and Illumina sequencing, respectively. This research was part of two undergraduate theses (ACB and KD) at Queen's University. This work was supported by a Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery Grant to S.C.L., the Baillie Family Chair in Conservation Biology (SCL), a NSERC Undergraduate Student Research Award to ACB, a NSERC Alexander Graham Bell Canada Graduate Scholarship to NAC, and an Ontario Trillium Scholarship to RST.

Declaration of Competing Interest

The authors have no competing interests to declare

References

549 Adams, J.M., 1997. Global land environments since the last interglacial. TN, USA: Oak
550 Ridge National Laboratory, <http://www.esd.ornl.gov/ern/qen/nerc.html>.

551 Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in
552 unrelated individuals. *Genome Res.* 19, 1655-1664.

553 Arnaiz-Villena, A., Álvarez-Tejado, M., Ruíz-del-Valle, V., García-de-la-Torre, C.,
554 Varela, P., Recio, M. J., Ferre, S., Martínez-Laso, J., 1998. Phylogeny and rapid
555 northern and southern hemisphere speciation of goldfinches during the Miocene and
556 Pliocene epochs. *Cell Mol. Life Sci.* 54, 1031-1041.

557 Arnaiz-Villena, A., Álvarez-Tejado, M., Ruíz-del-Valle, V., García-de-la-Torre, C.,
558 Varela, P., Recio, M.J., Martínez-Laso, J., 1999. Rapid Radiation of Canaries (Genus
559 *Serinus*). *Mol. Biol. Evol.* 16, 2-11.

560 Bachtrog, D., Thornton, K., Clark, A., Andolfatto, P., 2006. Extensive introgression of
561 mitochondrial DNA relative to nuclear genes in the *drosophila yakuba* species group.
562 *Evolution* 60, 292–302.

563 Baker, M.C., Spitler-Nabors, K.J., Bradley, D.C., 1982. The response of female mountain
564 white-crowned sparrows to songs from their natal dialect and an alien dialect. *Behav.*
565 *Ecol. and Sociobiol.* 10, 175-179.

566 Ballard, J.W.O., Whitlock, M.C., 2004. The incomplete natural history of mitochondria. *Mol.*
567 *Ecol.* 13, 729-744.

568 Benkman, C.W., 2003. Divergent selection drives the adaptive radiation of crossbills.
569 *Evolution* 57, 1176-1181.

570 Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. *Bioinformatics*
571 26, 1372-1373.

572 Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C-H., Xie, D., Suchard, M.A.,
573 Rambaut, A., Drummond, A.J., 2014. BEAST 2: A software platform for Bayesian
574 evolutionary analysis. PLOS Comput. Biol. 10, e1003537.

575 Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., Roy-Choudhury, A., 2012.
576 Inferring species trees directly from biallelic genetic markers: bypassing gene trees in
577 a full coalescent analysis. Mol. Biol. Evol. 29, 1917-1932.

578 Campagna, L., Benites, P., Loughheed, S.C., Lijtmaer, D.A., Di Giacomo, A.S., Eaton, M.D.,
579 Tubaro, P.L., 2012. Rapid phenotypic evolution during incipient speciation in a
580 continental avian radiation. Proc. R. Soc. B: Biol. Sci. 279, 1847-1856.

581 Campagna, L., Kopuchian, C., Tubaro, P.L., Loughheed, S.C., 2014. Secondary contact
582 followed by gene flow between divergent mitochondrial lineages of a widespread
583 Neotropical songbird (*Zonotrichia capensis*). Biol. J. Linn. Soc. 111, 863-868.

584 Campagna, L., Repenning, M., Silveira, L.F., Fontana, C.S., Tubaro, P.L., Lovette, I.J., 2017.
585 Repeated divergent selection on pigmentation genes in a rapid finch radiation. Sci.
586 Adv. 3, e1602404.

587 Carson, R.J., Spicer, G. S., 2003. A phylogenetic analysis of the emberizid sparrows based on
588 three mitochondrial genes. Mol. Phylogenet. Evol. 29, 43-57.

589 Clement, M., Posada, D., Crandall, K., 2000. TCS: a computer program to estimate gene
590 genealogies. Mol. Ecol. 9, 1657–1659.

591 Cortopassi, A.J., Mewaldt, L.R., 1965. The circumannual distribution of white-crowned
592 sparrows. Bird-Banding 36, 141-169.

593 Coyne, J.A., Orr, H.A., 2004. Speciation. Sunderland, MA, USA: Sinauer Publishers.

594 del Hoyo, J., Elliot, A., Christie, D.A., 2011. Handbook of Birds of the World, Volume 16:
595 Tanagers to New World Sparrows. Lynx Edicions, Catalonia, Spain.

596 Danner, J.E., Danner, R.M., Bonier, F., Martin, P., Small, T.W., Moore, I.T., 2011. Female,
597 but not male, tropical sparrows respond more strongly to the local song dialect:
598 implications for population divergence. *Am. Nat.* 178, 53-63.

599 Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture
600 between closely related populations. *Mol. Biol. Evol.* 28, 2239-2252.

601 Edgar, R.C., 2004. MUSCLE: a multiple sequence alignment method with reduced time and
602 space complexity. *BMC Bioinformatics* 5, 113.

603 Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson,
604 W.J., Sorenson, M.D., 2005. Speciation in birds: genes, geography, and sexual
605 selection. *PNAS* 102, 6550-6557.

606 Funk, D.J., Omland, K.E., 2003. Species-level paraphyly and plophyly: frequency, causes,
607 and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol.*
608 *Evol. Syst.* 34, 397-423.

609 Good, J.M., Hird, S., Reid, N., Demboski, J.R., Stepan, S.J., Martin-Nims, T.R., Sullivan, J.,
610 2008. Ancient hybridization and mitochondrial capture between two species of
611 chipmunks. *Mol. Ecol.* 17, 1313-1327.

612 Goudet, J., 2005. HIERFSTAT, a package for R to compute and test hierarchical F-statistics.
613 *Mol. Ecol. Notes* 5, 184-186.

614 Grant, P.R., 1999. *Ecology and Evolution of Darwin's Finches*. Princeton, New Jersey:
615 Princeton University Press.

616 Haldane, J.B.S., 1922. Sex ratio and unisexual sterility in animal hybrids. *J. Genet.*, 12, 101-
617 109.

618 Handford, P., 1980. Aspects of the annual cycle in highland populations of the rufous-
619 collared sparrow, *Zonotrichia capensis*. *The Auk* 97, 400-403.

620 Handford, P., Nottebohm, F., 1976. Allozymic and morphological variation in population
621 samples of rufous-collared sparrow, *Zonotrichia capensis*, in relation to vocal
622 dialects. *Evolution* 30, 802-817.

623 Hewitt, G.M., 2004. Genetic consequences of climatic oscillations in the Quaternary. *Phil.*
624 *Trans. R. Soc. Lond. B: Biol. Sci.* 359, 183-195.

625 Hill, G.E., 2017. The mitonuclear compatibility species concept. *The Auk* 134, 393-409.

626 Houtman, A.M., Falls, J.B., 1994. Negative assortative mating in the white-throated sparrow,
627 *Zonotrichia albicollis*: the role of mate choice and intra-sexual competition. *Anim.*
628 *Behav.* 48, 377-383.

629 Huelsenbeck, J.P., Ronquist, F., 2001. MRBAYES: Bayesian inference of phylogenetic trees.
630 *Bioinformatics* 17: 754-755.

631 Irwin, D.E., Rubtsov, A.S., Panov, E.N., 2009. Mitochondrial introgression and replacement
632 between yellowhammers (*Emberiza citrinella*) and pine buntings (*Emberiza*
633 *leucocephalos*) (Aves: Passeriformes). *Biol. J. Linn. Soc.* 98, 422-438.

634 Johnson, N.K., Cicero, C., 2004. New mitochondrial DNA data affirm the importance of
635 Pleistocene speciation in North American birds. *Evolution* 58, 1122-1130.

636 Jombart, T., 2008. adegenet: a R package for the multivariate analysis of genetic
637 markers. *Bioinformatics* 24, 1403-1405.

638 Knaus, B.J., Grüwald, N.J., 2017. vcfr: a package to manipulate and visualize variant call
639 format data in R. *Mol. Ecol. Resour.*, 17, 44-53.

640 Lamb, A.M., Gan, H.M., Greening, C., Joseph, L., Lee, Y.P., Moran-Ordóñez, Sunnucks, P.,
641 Pavlova, A., 2018. Climate-driven mitochondrial selection: A test in Australian
642 songbirds. *Mol. Ecol.* 27, 898-918.

643 Landys, M.M., Wingfield, J.C., Ramenofsky, M., 2004. Plasma corticosterone increases
644 during migratory restlessness in captive white-crowned sparrow *Zonotrichia*
645 *leucophrys gambelii*. Horm. Behav. 46, 574-581.

646 Lanfear R, Frandsen P.B., Wright A.M., Senfeld T., Calcott B., 2016 PartitionFinder 2: New
647 methods for selecting partitioned models of evolution for molecular and
648 morphological phylogenetic analyses. Mol. Biol. Evol. 34, 72-773.

649 Leigh, J.W., Bryant, D., 2015. PopART: Full-feature software for haplotype network
650 construction. Methods Ecol. Evol. 6, 1110–1116

651 Linnen, C.R., Farrell, B D., 2007. Mitonuclear discordance is caused by rampant
652 mitochondrial introgression in *neodiprion* (hymenoptera: diprionidae) sawflies.
653 Evolution 61, 1417-1438.

654 Lipshutz, S.E., Overcast, I.A., Hickerson, M.J., Brumfield, R.T., Derryberry, E.P., 2017.
655 Behavioral response to song and genetic divergence in two subspecies of white-
656 crowned sparrows (*Zonotrichia leucophrys*). Mol. Ecol. 26, 3011-3027.

657 Lijtmaer, D.A., Mahler, B., Tubaro, P.L., 2003. Hybridization and postzygotic isolation
658 patterns in pigeons and doves. Evolution 57, 1411-1418.

659 Lischer H.E.L., Excoffier, L., 2012. PGDSpider: An automated data conversion tool for
660 connecting population genetics and genomics programs. Bioinformatics 28, 298-299.

661 Loughheed, S.C., Campagna, L., Dávila, J.A., Tubaro, P.L., Lijtmaer, D.A., Handford, P.,
662 2013. Continental phylogenography of an ecologically and morphologically diverse
663 Neotropical songbird, *Zonotrichia capensis*. BMC Evol. Biol. 13, 58.

664 Loughheed, S.C., Handford, P., 1992. Vocal dialects and the structure of geographic variation
665 in morphological and allozymic characters in the Rufous-collared Sparrow,
666 *Zonotrichia capensis*. Evolution 46, 1443-1456.

667 MacDougall-Shackleton, E.A., MacDougall-Shackleton, S.A., 2001. Cultural and genetic
668 evolution in mountain white-crowned sparrows: song dialects are associated with
669 population structure. *Evolution* 55, 2568-2575.

670 McCarthy, E.M., 2006. Handbook of Avian Hybrids of the World. New York, NY, USA:
671 Oxford University Press.

672 McFarlan, J.T., Bonen, A., Guglielmo, C.G., 2009. Seasonal upregulation of fatty acid
673 transporters in flight muscles of migratory white-throated sparrows (*Zonotrichia*
674 *albicollis*). *J. Exp. Biol.* 212, 2934-2940.

675 McGuire, J.A., Linkem, C.W., Koo, M.S., Hutchison, D.W., Lappin, A.K., Orange, D.I.,
676 Lemos-Espinal, J., Riddle, B.R., Jaeger, J. R., 2007. Mitochondrial introgression and
677 incomplete lineage sorting through space and time: phylogenetics of crotaphytid
678 lizards. *Evolution* 61, 2879-2897.

679 McKay, B.D., Zink, R.M., 2010. The causes of mitochondrial DNA gene tree paraphyly in
680 birds. *Mol. Phylogenet. Evol.* 54, 647-650.

681 McKereghan-Dares, K., 2008. Diversification and speciation of crowned-sparrows, genus
682 *Zonotrichia*, in the New World (Undergraduate thesis, Queen's University, Kingston,
683 Ontario). Retrieved from <http://hdl.handle.net/1974/1164>.

684 Melo-Ferreira, J., Boursot, P., Suchentrunk, F., Ferrand, N., Alves, P.C., 2005. Invasion from
685 the cold past: extensive introgression of mountain hare (*Lepus timidus*) mitochondrial
686 DNA into three other hare species in northern Iberia. *Mol. Ecol.* 14, 2459-2464.

687 Miller, A.H. 1940. A hybrid between *Zonotrichia coronata* and *Zonotrichia leucophrys*. *The*
688 *Condor* 42, 45-48.

689 Moore, I.T., Bonier, F., Winfield, J.C., 2005. Reproductive asynchrony and population
690 divergence between two tropical bird populations. *Behav. Ecol.* 16, 755-762.

691 Morton, M.L., Mewaldt, L.R., 1960. Further evidence of hybridization between *Zonotrichia*
692 *atricapilla* and *Zonotrichia leucophrys*. The Condor 62, 485-486.

693 Pickrell, J.K., Pritchard, J.K., 2012. Inference of population splits and mixtures from
694 genome-wide allele frequency data. PLOS Genet. 8, e1002967.

695 Pleszczynska, W., 1979. Interspecific and intraspecific song discrimination by *Zonotrichia*
696 *albicollis* and *Z. leucophrys*. Can. J. Zool. 58, 1433-1438.

697 Poland J.A., Brown, P.J., Sorrells, M.E., Jannink, J-L., 2012. Development of high-density
698 genetic maps for barley and wheat using a novel two-enzyme genotyping-by-
699 sequencing approach. PLOS ONE 7, e32253.

700 Price, E.R., McFarlan, J.T., Guglielmo, C.G., 2010. Preparing for migration? The effects of
701 photoperiod and exercise on muscle oxidative enzymes, lipid transporters, and
702 phospholipids in white-crowned sparrows. Physiol. Biochem. Zool. 83, 252-262.

703 Pritchard, J.K., Stephens, M., Donnelly, P., 2000. Inference of population structure using
704 multilocus genotype data. Genetics 155, 945-959.

705 Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J.,
706 Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A tool set for
707 whole-genome association and population-based linkage analyses. Am. J. Hum.
708 Genet. 81, 559-575.

709 R Core Team., 2006. R: a language and environment for statistical computing. R Foundation
710 for Statistical Computing, Vienna, Austria. Available at <https://www.R-project.org/>.

711 Rambaut A, Drummond A.J., Xie D., Baele G., Suchard M.A., 2018. Posterior
712 summarization in Bayesian phylogenetics using Tracer 1.7. Syst. Biol. 67, 901-904.

713 Rambaut, A., Drummond, A.J., 2007. TRACER v1.5. Available at
714 <http://beast.bio.ed.ac.uk/Tracer>.

715 Rheindt, F.E., Edwards, S.V., 2011. Genetic introgression: an integral but neglected
716 component of speciation in birds. *The Auk* 128, 620-632.

717 Rognes, T., Flouri, T., Nichols, B., Quince, C., Mahé, F., 2016. VSEARCH: a versatile open
718 source tool for metagenomics. *PeerJ* 4, e2584.

719 Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget,
720 B., Liu, L., Suchard, M.A. and Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient
721 Bayesian phylogenetic inference and model choice across a large model space. *Syst.*
722 *Biol.* 61: 539-542.

723 RStudio Team., 2015. RStudio: Integrated Development for R. Boston, MA: RStudio, Inc.
724 Available at <http://www.rstudio.com/>

725 Sambrook, J., Russell, D.W., 2001. *Molecular Cloning: A Laboratory Manual*. Third edition.
726 New York, USA: Cold Spring Harbor Laboratory Press.

727 Schilthuizen, M.C., Giesbers, W.G., Beukeboom, L.W., 2011. Haldane's rule in the 21st
728 century. *Heredity* 107, 95-102.

729 Shen, Y-Y., Liang, L., Sun, Y-B., Yue, B-S., Yang, X-J., Murphy, R.W., Zhang, Y-P., 2010.
730 A mitogenomic perspective on the ancient, rapid, radiation in the Galliformes with an
731 emphasis on the Phasianidae. *BMC Evol. Biol.* 10, 132.

732 Shizuka, D., 2014. Early song discrimination by nestling sparrows in the wild. *Anim. Behav.*
733 92, 19-24.

734 Sorenson, M.D., Sefc, K.M., Payne, R.B., 2003. Speciation by host switch in brood
735 parasitic indigobirds. *Nature* 424, 928-931.

736 Spinks, P.Q., Shaffer, H.B., 2007. Conservation phylogenetics of the Asian box turtles
737 (*Geoemydidae*, *Cuora*): mitochondrial introgression, numts, and inferences from
738 multiple nuclear loci. *Conserv. Genet.* 8, 641-657.

739 Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
740 large phylogenies. *Bioinformatics* 30, 1312-1313.

741 Stamatakis, A., Hoover, P., Rougemont, J.A., 2008. Rapid bootstrap algorithm for the
742 RAxML Web Servers. *Syst. Biol.* 57, 758-771.

743 Takahata, N., Slatkin, M., 1984. Mitochondrial gene flow. *PNAS* 81, 1764-1767.

744 Toews, D.P.L., 2017. From song dialects to speciation in white-crowned sparrows. *Mol.*
745 *Ecol.* 26, 2842-2844.

746 Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear
747 discordance in animals. *Mol. Ecol.* 21, 3907-3930.

748 Toews, D.P.L. Mandic, M., Richards, J.G., Irwin, D.E., 2014. Migration, mitochondria, and
749 the yellow-rumped warbler. *Evolution* 68, 241-155.

750 Tomback, D.F., Baker, M.C., 1984. Assortative mating by white-crowned sparrows at song
751 dialect boundaries. *Anim. Behav.* 32, 465-469.

752 Turelli, M., Orr, H.A., 1995. The dominance theory of Haldane's rule. *Genetics* 140, 389-
753 402.

754 Weckstein, J.D., Zink, R.M., Blackwell-Rago, R.C., Nelson, D.A., 2001. Anomalous
755 variation in mitochondrial genomes of white-crowned (*Zonotrichia leucophrys*) and
756 golden-crowned (*Z. atricapilla*) sparrows: pseudogenes, hybridization, or incomplete
757 lineage sorting? *The Auk* 118, 231-236.

758 Weir, B.S., Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population
759 structure. *Evolution* 38, 1358-1370.

760 Weir, J.T. Price, T., 2011. Limits to speciation inferred from times to secondary sympatry
761 along a latitudinal gradient. *Am. Nat.* 177, 462-469.

- Wilson, C.C., Bernatchez, L., 1998. The ghost of hybrids past: fixation of arctic charr
(*Salvelinus alpinus*) mitochondrial DNA in an introgressed population of lake trout (*S.*
namaycush). Mol. Ecol. 7, 127-132.
- Zink, R.M., 1982. Patterns of genic and morphologic variation among sparrows in the genera
Zonotrichia, *Melospiza*, *Junco*, and *Passerella*. The Auk 99, 632-649.
- Zink, R.M., Dittmann, D.L., Rootes, W.L., 1991. Mitochondrial DNA variation and the
phylogeny of *Zonotrichia*. The Auk 108, 578-584

Table 1

Species and subspecies designations, sampling locales, collection, tissue source, and datasets for specimens used in this study (ddRAD and mtDNA).

Species	Subspecies	Sampling Location	Collection ¹ / Specimen ID	Tissue Source	ddRAD	mtDNA
<i>Z. atricapilla</i>	na	Alaska: North Slope Borough; Barrow	UWBM 80383	tissue	yes	yes
	na	Bodega Bay, CA, USA	UWWL WL004	blood	yes	yes
	na	California: Modoc County; Modoc National Wildlife Refuge	UWBM 83465	tissue	yes	yes
	na	California; Marin County; Stinson Beach	UWBM 80072	tissue	no	yes
	na	Mile 26 Taylor Highway, AK, USA	UAM RWD24801	tissue	no	yes
	na	Mother Goose Lake, AK, USA	UAM RWD24034	tissue	failed	no
	na	Kanuti National Wildlife Refuge, AK, USA	UAM RWD24064	tissue	failed	no
	na	Washington: Clallam County; Sequim	UWBM 81808	tissue	failed	yes
	na	Washington: Grant County; Moses Lake	UWBM 83464	tissue	yes	yes
	na	Washington: King County; Bellevue	UWBM 80425	tissue	yes	yes
	na	Washington: King County; Bellevue	UWBM 80465	tissue	no	yes
	na	Washington: King County; Renton	UWBM 69549	tissue	yes	yes
	na	Washington: King County; Renton	UWBM 69491	tissue	no	yes
	na	Washington: King County; Renton	UWBM 69582	tissue	no	yes
	na	Washington: King County; Seattle	UWBM 81767	tissue	yes	yes
	na	Washington: King County; Seattle	UWBM 79613	tissue	no	yes
	na	Washington: Kittitas County; Manastash Canyon	UWBM 79560	tissue	yes	yes

	na	Washington: Pacific County; Enumclaw	UWBM 79419	tissue	yes	yes
	na	Washington: Pacific County; Enumclaw	UWBM 79434	tissue	yes	yes
	na	Washington: Pacific County; Enumclaw	UWBM 74042	tissue	yes	yes
	na	Washington: Pacific County; Enumclaw	UWBM 79416	tissue	yes	yes
	na	Washington: Pacific County; Enumclaw	UWBM 79432	tissue	yes	yes
	na	Washington: Pacific County; Grayland	UWBM 61741	tissue	yes	yes
	na	Washington: Pacific County; Grayland	UWBM 61689	tissue	no	yes
	na	Washington: Yakima County; Ellensburg	UWBM 70566	tissue	yes	yes
	na	Washington: Yakima County; Ellensburg	UWBM 70565	tissue	no	yes
	na	Washington: Yakima County; Ellensburg	UWBM 70567	tissue	no	yes
	na	Hudson Bay Mountains, Smithers, B.C.	UNL 264101006	tissue	yes	no
	na	Hudson Bay Mountains, Smithers, B.C.	UNL 'OP'	tissue	yes	no
	na	Hatcher Pass, AK, USA	UNL 264101038	tissue	yes	no
<i>Z. leucophrys</i>	<i>nuttallii</i>	Big River State Beach, CA, USA	UWWL WL087	blood	yes	yes
	<i>nuttallii</i>	Big River State Beach, CA, USA	UWWL WL086	blood	no	yes
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL009	blood	yes	no
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL011	blood	yes	no
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL014	blood	yes	yes
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL015	blood	yes	yes
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL016	blood	yes	yes
	<i>nuttallii</i>	Bodega Bay, CA, USA	UWWL WL017	blood	yes	yes

	<i>nutallii</i>	Bodega Bay, CA, USA	UWWL WL012	blood	no	yes
	<i>nutallii</i>	Little River State Beach, CA, USA	UWWL WL084	blood	yes	no
	<i>nutallii</i>	Little River State Beach, CA, USA	UWWL WL085	blood	yes	yes
	<i>nutallii</i>	Pismo Beach/Oceano Dunes State Beach, CA, USA	UWWL WL072	blood	yes	yes
	<i>nutallii</i>	Pismo Beach/Oceano Dunes State Beach, CA, USA	UWWL WL073	blood	yes	no
	<i>nutallii</i>	Pismo Beach/Oceano Dunes State Beach, CA, USA	UWWL WL077	blood	yes	no
	<i>nutallii</i>	Roosevelt Beach State Park, CA, USA	UWWL WL100	unknow n ²	yes	yes
	<i>nutallii</i>	Roosevelt Beach State Park, CA, USA	UWWL WL348	unknow n	yes	yes
	<i>nutallii</i>	Roosevelt Beach State Park, CA, USA	UWWL WL364	unknow n	yes	yes
	<i>nutallii</i>	Venice Beach State Park, CA, USA	UWWL WL092	unknow n	yes	yes
	<i>nutallii</i>	Venice Beach State Park, CA, USA	UWWL WL093	unknow n	failed	no
	<i>nutallii</i>	Venice Beach State Park, CA, USA	UWWL WL095	unknow n	yes	yes
	<i>nutallii</i>	Venice Beach State Park, CA, USA	UWWL WL099	unknow n	yes	yes
<i>Z. leucophrys</i>	<i>pugetensis</i>	Stevens, WA, USA	UWWL WL061	blood	yes	yes
	<i>pugetensis</i>	Stevens, WA, USA	UWWL WL400	blood	no	yes
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL UWS10	blood	yes	no
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL UWS1	blood	yes	no
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL UWS8	blood	yes	no
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL UWS9	blood	yes	no

<i>Z. leucophrys</i>	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL 196144918	blood	no	yes
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL 196144948	blood	no	yes
	<i>pugetensis</i>	University of Washington, Seattle, WA, USA	UWWL 196144993	blood	no	yes
	<i>pugetensis</i>	Washington: Pacific County; Frances	UWBM 73051	tissue	yes	yes
	<i>pugetensis</i>	Washington: Pacific County; Frances	UWBM 73028	tissue	yes	yes
	<i>pugetensis</i>	Washington: Pacific County; Frances	UWBM 73050	tissue	yes	yes
	<i>pugetensis</i>	Washington: Snohomish County; Gold Bar	UWWL WL398	blood	yes	yes
	<i>pugetensis</i>	Washington: Snohomish County; Gold Bar	UWWL WL399	blood	no	yes
	<i>pugetensis</i>	Westport, Washington, USA	UWWL 196144985	blood	no	yes
	<i>pugetensis</i>	Westport, Washington, USA	UWWL 196144986	blood	no	yes
	<i>gambelii</i>	North Slope Borough, AK, USA	UWBM 78649	tissue	yes	yes
	<i>gambelii</i>	Churchill, MB, CAN	ROM ROM 3532	tissue	yes	yes
	<i>gambelii</i>	Fairbanks, AK, USA	UWBM 54024	tissue	yes	yes
	<i>gambelii</i>	Toolik Lake, North Slope, AK, USA	UWWL TK264714	tissue	yes	yes
	<i>gambelii</i>	Toolik Lake, North Slope, AK, USA	UWWL TK264762	tissue	yes	yes
	<i>gambelii</i>	Toolik Lake, North Slope, AK, USA	UWWL TK2651117	tissue	yes	yes
	<i>gambelii</i>	Toolik Lake, North Slope, AK, USA	UWWL TK26520	tissue	yes	yes
	<i>gambelii</i>	Alaska (south of Toolik Field Station), AK, USA	UWBM 78642	tissue	no	yes
	<i>gambelii</i>	Alaska (south of Toolik Field Station), AK, USA	UWBM 78643	tissue	no	yes

	<i>gambelii</i>	Alaska (south of Toolik Field Station), AK, USA	UWBM 78644	tissue	no	yes
<i>Z. leucophrys</i>	<i>oriantha</i>	3 miles north-west of Beaver Mines Lake, AB, CAN	RAM Z961814	tissue	yes	yes
	<i>oriantha</i>	Colorado: Gunnison County	UWBM 53405	tissue	yes	yes
	<i>oriantha</i>	Colorado: Summit County; Frisco	UWBM 70309	tissue	yes	yes
	<i>oriantha</i>	Skyline Trail, Porcupine Hills, AB, CAN	RAM Z12217	tissue	yes	yes
<i>Z. leucophrys</i>	<i>leucophrys</i>	James Bay, QC, CAN	ROM ROM3052	blood	failed	no
	<i>leucophrys</i>	James Bay, QC, CAN	ROM ROM3053	blood	failed	no
	<i>leucophrys</i>	James Bay, QC, CAN	ROM ROM3057	blood	failed	no
	<i>leucophrys</i>	Queen's University Biological Station, Kingston, ON, CAN	SL 2015-2	tissue	yes	yes
	<i>leucophrys</i>	Toronto, ON, CAN	ROM ROM329	tissue	yes	yes
<i>Z. albicollis</i>	na	Conner Creek Grazing Reserve, AB, CAN	RAM Z0632	tissue	yes	yes
	na	House Mountain, Swan Hills, AB, CAN	RAM Z991431	tissue	yes	yes
<i>Z. querula</i>	na	Nebraska: Lancaster County; Denton	UWBM 69719	tissue	yes	yes
	na	Unknown	ROM ROM4409	tissue	yes	yes
<i>Z. capensis</i>	<i>huancamba</i>	Cushi, Peru	LSUM B8084	liver	yes	yes
<i>Z. capensis</i>	<i>matutina</i>	Catarata Arco Iris, Bolivia	LSUM B14823	liver	yes	yes
	<i>matutina</i>	Charagua, Bolivia	LSUM B18743	liver	yes	yes
<i>Z. capensis</i>	<i>australis</i>	Ushuaia, Tierra del Fuego, Argentina. 54°48' S, 68°18' W	SL SCL106	blood	yes	no

¹ Acronyms refer to the museum or university collection that the specimen(s) were obtained from: LSUM=Louisiana State University Museum; RAM=Royal Alberta Museum; ROM=Royal Ontario Museum; SL=Lougheed Lab; Queen's University; UAM=University of Alaska Museum (of the North); UNL=University of Nebraska-Lincoln (Daizaburo

814 Shizuka); UWBM=Burke Museum of Natural History and Culture, University of
 815 Washington; UWWL=Wingfield Lab, University of Washington

816
 817 ² DNA aliquot

818
 819
 820
 821

Table 2

Pairwise F_{ST} values from analysis of nuclear SNP data for *Z. atricapilla* and all five *Z. leucophrys* subspecies.

	<i>Z. atricapilla</i>	<i>Z. l. nuttallii</i>	<i>Z. l. pugetensis</i>	<i>Z. l. gambelii</i>	<i>Z. l. oriantha</i>
<i>Z. l. nuttallii</i>	0.31				
<i>Z. l. pugetensis</i>	0.31	0.04			
<i>Z. l. gambelii</i>	0.27	0.13	0.13		
<i>Z. l. oriantha</i>	0.28	0.13	0.13	0.01	
<i>Z. l. leucophrys</i>	0.28	0.13	0.14	0.02	0.02

Table 3

Pairwise F_{ST} values from analysis of nuclear SNP data between the three genetic clusters uncovered in our analyses (*Z. atricapilla*, NP and GOL) – see Figure 4. NP = cluster comprising *Z. l. nutalli* and *Z. l. pugatensis*. GOL = cluster comprising *Z. l. gambelli*, *Z. l. oriantha*, and *Z. l. leucophrys*.

	<i>Z. atricapilla</i>	NP
NP	0.30	
GOL	0.27	0.12

Fig. 1. The locations of sampling sites of *Z. atricapilla* and *Z. leucophrys* individuals. The background colors represent the breeding ranges, pink for *Z. atricapilla* and purple for *Z. leucophrys* with deep pink representing range overlap. Circles indicate samples used for both ddRADseq and mitochondrial DNA sequencing, triangles represent locations where at least one sample has mitochondrial DNA only, squares represent locations where at least one sample has ddRAD data only. More sampling details are provided in Table 1.

Fig. 2. Statistical parsimony network of 1211 bp of the mitochondrial control region and COI concatenated for *Z. atricapilla* and all *Z. leucophrys* subspecies. Black circles represent unsampled or now extinct haplotypes and the hatch marks indicate mutations. Circle size indicates the number of samples of each haplotype.

Fig. 3. Bayesian mitochondrial DNA phylogenetic analyses using individual haplotypes from all five *Zonotrichia* subspecies. The legend shows which haplotypes are present in *Z. atricapilla* and the *Z. leucophrys* subspecies.

Fig. 4. Maximum-likelihood trees produced in RAxML with all outgroups (A) and with only one outgroup taxa to show relationships within *Z. leucophrys* more clearly (B). Node labels indicate bootstrap support values from 1000 bootstrap replicates; bootstrap values above 90% are shown. Name labels are coloured according to species and subspecies (red=*Z. atricapilla*; orange=*Z. l. nuttalli*; yellow=*Z. l. pugetensis*; green=*Z. l. gambelii*; purple=*Z. l. leucophrys*; blue=*Z. l. oriantha*; black=outgroup taxa).

Fig. 5. Maximum likelihood tree produced in Treemix (A), when inferring one migration event as suggested by OptM (B), when inferring two migration events (B) and three migration events (D). The bottom axis shows the drift parameter. Name labels are coloured according to species and subspecies (red=*Z. atricapilla*; orange=*Z. l. nuttallii*; yellow=*Z. l. pugetensis*; green=*Z. l. gambelii*; purple=*Z. l. leucophrys*; blue=*Z. l. oriantha*).

Fig. 6. Molecular assignments of *Z. atricapilla* and *Z. leucophrys* individuals using ADMIXTURE. (A) The plot for K= 2 separating *Z. atricapilla* from *Z. leucophrys*, (B) The plot for K = 3 separating *Z. atricapilla* as one cluster, *Z. l. nuttalli* and *Z. l. pugetensis* (NP) as another, and *Z. l. gambelii*, *Z. l. oriantha*, and *Z. l. leucophrys* (GOL) as the third, and (C) The plot for K = 4 and (D) K = 5 showing some structuring within the NP and GOL clades.

Fig. 7. Results of a principal component analysis of *Z. atricapilla* and the five *Z. leucophrys* subspecies (A), of the NP clade (B) and the GOL clade (C). Ellipses and colours are encoded as follows: red=*Z. atricapilla*; orange=*Z. l. nuttallii*; yellow=*Z. l. pugetensis*; green=*Z. l. gambelii*; purple=*Z. l. leucophrys*; blue=*Z. l. oriantha*.