Molecular Ecology

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE

Thrushes in Love: Extensive Gene Flow, With Differential Resistance and Selection, Obscures and Reveals the Evolutionary History of a Songbird Clade

Kira E. Delmore[1] | Jeffrey M. DaCosta[2] | Kevin Winker[3]

[1]Department of Biology, Texas A&M University, College Station, Texas, USA | [2]Biology Department, Boston College, Chestnut Hill, Massachusetts, USA | [3]Department of Biology and Wildlife, University of Alaska Museum, Fairbanks, Alaska, USA

Correspondence: Kira E. Delmore (ked2195@columbia.edu)

Received: 18 May 2024 | **Revised:** 6 December 2024 | **Accepted:** 16 December 2024

Handling Editor: Yanhua Qu

Funding: This work was supported by the Division of Integrative Organismal Systems (Grant IOS-2143004), and the National Institute of General Medical Sciences (Grant 1R35GM151012).

Keywords: adaptive introgression | phylogenomics | recombination | seasonal migration | speciation with gene flow | Z chromosome

ABSTRACT

The application of high-throughput sequencing to phylogenetic analyses is allowing authors to reconstruct the true evolutionary history of species. This work can illuminate specific mechanisms underlying divergence when combined with analyses of gene flow, recombination and selection. We conducted a phylogenomic analysis of *Catharus*, a songbird genus with considerable potential for gene flow, variation in migratory behaviour and genomic resources. We documented discordance among trees constructed for mitochondrial, autosomal and sex (Z) chromosome partitions. Two trees were recovered on the Z. Both trees differed from the autosomes, one matched the mitochondria, and the other was unique to the Z. Gene flow with one species likely generated much of this discordance; substantial admixture between *ustulatus* and the remaining species was documented and linked to at least two historic events. The tree unique to the Z likely reflects the true history of *Catharus*; local genomic analyses recovered the same tree in autosomal regions with reduced admixture and recombination. Genes previously connected to migration were enriched in these regions suggesting transitions between migratory and non-migratory states helped generate divergence. Migratory (vs. nonmigratory) *Catharus* formed a monophyletic clade in a subset of genomic regions. Gene flow was elevated in some of these regions suggesting adaptive introgression may have occurred, but the dominant pattern was of balancing selection maintaining ancestral polymorphisms important for olfaction and perhaps, by extension, adaptation to temperate climates. This work illuminates the evolutionary history of an important model in speciation and demonstrates how differential resistance to gene flow can affect local genomic patterns.

1 | Introduction

Phylogenetic inference is important for reconstructing the evolutionary history of related lineages but has been a difficult task. For example, discordant phylogenies are often recovered when using different sets or partitions of genetic data, making it difficult to identify the true species tree among radiating lineages (Degnan and Rosenberg 2009; Nichols 2001). Whole-genome approaches are helping overcome this problem, suggesting that processes with varying or local effects on the genome generate much of the phylogenetic discordance documented. These whole-genome approaches are allowing researchers to

resolve the true evolutionary history of closely related species and gain insight into long-standing questions in evolutionary biology. For example, gene flow is one process with varying effects on the genome; some genomic regions remain resistant to gene flow while others introgress at neutral or even adaptive rates (Barton and Gale 1993; Gillespie et al. 2020; Harrison and Larson 2014; Wu 2001). Genomic regions that remain resistant to gene flow often reflect the true species tree, and genes in these regions likely help maintain reproductive isolation between species (i.e., serve as barrier or speciation genes; Ting, Tsaur, and Wu 2000; Nosil and Schluter 2011; Cutter and Payseur 2013; Bravo et al. 2019). Identifying speciation genes is a long-standing goal in evolutionary biology, with the potential to provide insight into the ecological, evolutionary and molecular mechanisms underlying speciation (Orr 2005; Rieseberg and Blackman 2010; Ravinet et al. 2017; Nosil and Schluter 2011).

Gene flow is not the only process with varying effects on the genome, and it likely acts in concert with additional factors. For example, lower recombination rates likely help counter the homogenising effects of gene flow and maintain the true species tree in some genomic regions. This may be especially true on the sex chromosomes, where speciation genes are also known to cluster (Fontaine et al. 2015; Edelman et al. 2019; Li et al. 2019). Similar to speciation genes, genomic regions where adaptive introgression is occurring are relevant to evolutionary biology (Abbott et al. 2013; Mallet 2005). Adaptive introgression involves the interplay between gene flow and positive selection, with alleles that are favoured in more than one group introgressing across groups. Alleles that allow populations to expand their geographic ranges, explore novel niches and avoid extinction often exhibit signatures of adaptive introgression (Pfennig, Kelly, and Pierce 2016; Oziolor et al. 2019; Jones et al. 2020). More generally, adaptive introgression speaks to the creative role gene flow can play in evolution. Especially in animals, gene flow was traditionally viewed as a destructive force in evolution, but the application of whole-genome approaches to natural populations has turned this idea on its head, highlighting many ways gene flow can promote adaptation and speciation (Morjan and Rieseberg 2004; Dasmahapatra et al. 2012; Seehausen et al. 2014; Abbott, Barton, and Good 2016; Taylor and Larson 2019).

Much of the work on local phylogenomic patterns has focused on shallow evolutionary scales (e.g., pairs of closely related populations or recent adaptive radiations) and phenotypic traits that are both easy to observe and have a simple genetic basis (e.g., colour in butterflies, hares, and cichlids Irisarri et al. 2018; Edelman et al. 2019; Ferreira et al. 2021). Gene flow, recombination and selection likely affect deeper evolutionary relationships as well, and to gain a complete understanding of how these processes affect local genomic patterns, we must expand beyond simple phenotypic traits. Here, we extend phylogenomic research to deeper evolutionary timescales and more complex traits using *Catharus*, a genus of songbirds that has considerable potential for gene flow and whose members differ primarily in their seasonal migratory behaviour.

Seasonal migration could have varying effects on phylogenetic relationships exhibited by different portions of the genome. For example, migration could facilitate gene flow (Montgomery 1896; Rockwell and Barrowclough 1987; Arguedas and Parker 2000; Belliure, and Møller, and Clobert. 2000). If this gene flow is adaptive, we might expect to find monophyletic relationships among a subset of lineages in genomic regions that harbour universally favoured genes (e.g., migrants may be monophyletic in regions with genes that encode migratory traits and/or genes that are important for adaptation to temperate breeding grounds). On the other hand, migration may also help maintain species boundaries by contributing to geographical or temporal isolation. Historical and ongoing differentiation is common among migratory lineages and could be generated by divergent selection related to differences in the location and timing of resources (Winker 2010; Rohwer and Irwin 2011; Turbek, Scordato, and Safran 2018). This form of differential selection could promote speciation through either assortative mating (e.g., if populations differ in when and where they breed) and/or reductions in the fitness of immigrants and hybrids (e.g., if hybrids exhibit intermediate traits). If migration helps maintain species boundaries, we might expect to find genes underlying migration in genomic regions that contain the true species tree (Cutter and Payseur 2013).

The songbird genus *Catharus* includes 12 species: seven non-migrants in Mexico, Central America and South America, and five long-distance migrants that breed from Siberia to the northern United States and winter across parts of Mexico, Central and South America (Figure 1). Several species of *Catharus* have parapatric distributions, and many of the migratory species pass through one another's ranges during migration. Gene flow has already been implicated in discordant phylogenetic patterns in *Catharus*, with the placement of one species (*ustulatus*, the migratory Swainson's thrush) being the most variable (Outlaw et al. 2003; Klicka, Voelker, and Spellman 2005; Winker and Pruett 2006; Voelker, Bowie, and Klicka 2013). A recent study using ultraconserved elements (UCEs) and data for 10/12 *Catharus* species documented extensive gene flow across the genome, even among non-sister species (Everson et al. 2019). We expanded on the former UCE study here, using whole-genome resequencing data, a reference genome for one of the species (*ustulatus*), and estimates of recombination from this reference to examine the evolutionary history of *Catharus*. We also examined the role seasonal migration could have on the phylogenomic patterns we documented. Beyond differences in the propensity to migrate, additional differences exist among the migratory species, including the routes they take (e.g., subspecies of *ustulatus* and *guttatus* migrate along the western flyway in North America while most other migrants use the eastern flyway; Delmore, Fox, and Irwin 2012), where they winter (e.g., *fuscescens* winters further south than most other migrants) and when they migrate (e.g., the leapfrog migration occurring between *minimus* and *bicknelli*; Winker 2010). One species (*ustulatus*) has been the focus of considerable work on the genetics of migration, including a recent genome-wide association analysis identifying loci underlying this behaviour (Justen, Easton, and Delmore 2024).

We started our analysis by constructing separate phylogenetic trees for mitochondrial, autosomal and sex (Z) chromosome partitions. We uncovered substantial phylogenetic discordance at this level. Accordingly, we continued with a series of analyses designed to examine the role gene flow, recombination and

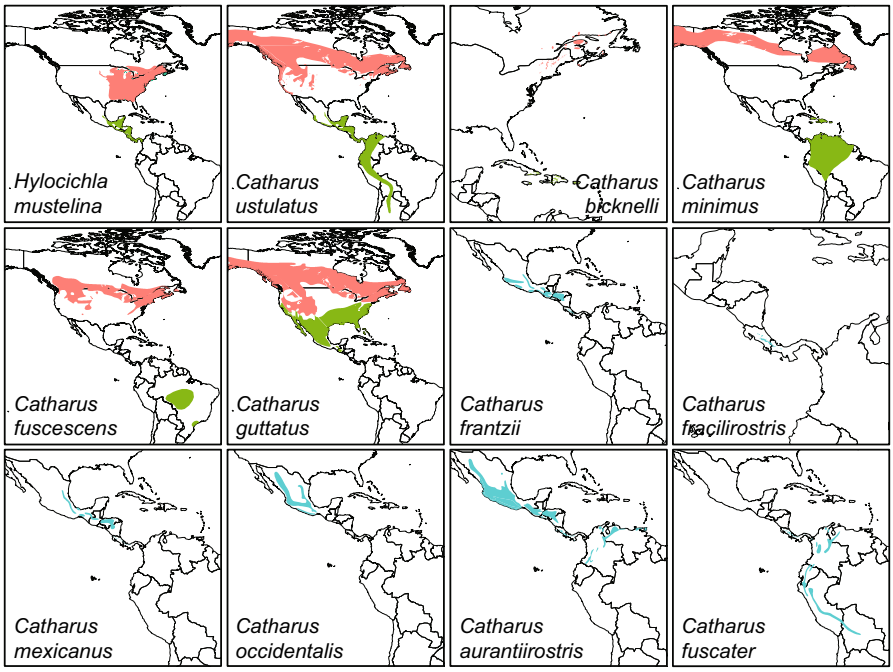

FIGURE 1 | Geographic range of species included in this study. Breeding (pink) and wintering (green) distributions shown for migratory species; blue shows range of resident species. Maps obtained from the BirdLife International and Handbook of the Birds of the World (http://datazone.birdlife.org/species/requestdis).

selection play in generating this discordance. These analyses included tests for gene flow at both the whole-genome and local genomic scales, in windows that likely reflect the putatively true species tree and in windows where migrants were monophyletic. We also examined recombination rates, gene content and sequence divergence in these sets of genomic windows. Our results suggest that variation in both historical gene flow and recombination rates generate much of the phylogenomic discordance we documented. Divergent selection at loci underlying migration and balancing selection may also contribute to discordant phylogenomic patterns.

2 | Methods

2.1 | Sampling, Sequencing and Genotyping

We extracted DNA from high-quality voucher specimens or blood samples from 11 species of *Catharus* (dryas is the only species from *Catharus* that was not included here; Table S1). Whole-genome resequencing libraries were constructed using Nextera (Illumina) DNA Flex Library Prep kits, and libraries were sequenced on a NovaSeq S4 (Illumina, San Diego, CA) using paired-end 150 bp reads. We aligned reads to the *ustulatus* genome (inland subspecies, Feng et al. 2020) using bwa (mem algorithm with default settings, Li and Durbin 2009). The resulting .sam files were converted to bam format with samtools (Li et al. 2009), and picardtools was used to clean, sort and add read groups to .bam files (http://broadinstitute.github.io/picard/). Between 97% and 98% of reads mapped to the reference for an average read depth of 23× (range 15.95–29.84).

We called SNPs using GATK Best Practices (McKenna et al. 2010). Briefly, the HaplotypeCaller algorithm in GVCF mode was run on each scaffold separately (--sample-ploidy 2 unless the bird was female and scaffold was on the Z chromosome [female birds only have one copy of the Z]). We gathered data for each individual using GatherVcfs and created a database using GenomicsDBImport. We called genotypes across all samples using GenotypeGVCFs. The resulting .vcf file was hard filtered using VariantFiltration and SelectVariants to select biallelic SNPs that passed quality thresholds suggested in GATK documentation (QD ≥ 2, QUAL ≥ 30, SOR ≤ 3, MQ ≥ 40, MQRankSum ≥ -12.5, ReadPosRankSum ≥ -8), and further filtering was applied in VCFtools v0.1.17 (Danecek et al. 2011) to select variants with no missing genotypes, minor allele count ≥ 2 and minimum mean depth ≥ 4.

2.2 | Phylogenomic Analyses

Whole mitogenome data were used to estimate the mitochondrial gene tree. For each sample, reads that mapped to the mitochondrial scaffold of the *ustulatus* reference genome were extracted using samtools. We then used samtools mpileup and bcftools v1 (Li 2011) to call a consensus mitochondrial sequence for each sample. Sequences were aligned using MAFFT v7 (Katoh and Standley 2013), and the location of each protein coding position was determined with the guidance of the reference mitogenome annotation. The sequences for the gene ND6 were reverse complemented. Partitionfinder v2 (Lanfear et al. 2017) was run to determine partitions and their appropriate models of sequence evolution. MrBayes v3 (Ronquist and Huelsenbeck 2003) was then used to estimate a Bayesian gene tree, using two runs each with four chains, 10M generations, and sampling every 1000 generations. At the end of the runs, all parameters had a potential scale reduction factor of ~1 and effective sample sizes > 200. The two runs

were combined into a single posterior with a burnin of 25% for each run, and the posterior was summarised in a clade credibility tree.

Relationships among species were estimated separately using data from scaffolds assigned to autosomes and the sex-linked Z chromosome. For autosomes, all SNPs that passed filtering ($N = 24{,}104{,}438$) were joined into a single data matrix for a concatenated analysis. The Python script vcf2phylip.py (https://github.com/edgardomortiz/vcf2phylip) was used to convert diploid genotypes to a single sequence for each sample, with ambiguity codes for heterozygous positions. RAxML-NG v1 (Kozlov et al. 2019) was used to estimate a maximum likelihood topology with the GTR+G model of sequence evolution, 20 searches (10 with random starting trees and 10 with maximum parsimony starting trees) and 100 bootstraps. A species tree analysis was also performed using ASTRAL-III (Zhang et al. 2018), which finds a species tree with the maximum number of induced quartet trees from a set of gene trees. For input gene trees, we used VCFtools to parse SNPs for non-overlapping 100 kb windows across each scaffold and estimated a gene tree for each window using RAxML-NG (same settings as concatenated analysis). Nodal support was estimated by calculating the percentage of quartets in the gene trees that agree with the node. We also used SVDquartets (Chifman and Kubatko 2014) as implemented in PAUP* v4 (Swofford 2001) to estimate a species tree. This method uses SNP data and the multispecies coalescent to estimate a species tree using quartet puzzling. To avoid using closely linked SNPs, we 'thinned' the collection of SNPs on each scaffold by randomly selecting sites with a minimum distance of 10 kb between each site (https://github.com/jeffdacosta/Bioinformatics-scripts/blob/main/vcf_thin_variants.py). This random selection process was repeated 10 times, and each data set was analysed with SVDquartets with 100 bootstrap replicates. All bootstrap trees were saved, and a majority-rule consensus tree was estimated in PAUP*.

The same methodologies (i.e., concatenated maximum likelihood, gene tree quartet puzzling, SNP quartet puzzling) were used to estimate topologies from SNPs on the Z chromosome (super scaffold 7 of the reference genome; $N = 1{,}872{,}822$ SNPs). Since preliminary phylogenetic analyses using SNPs from the Z chromosome produced a topology with a different first split within *Catharus* compared to autosomal analyses, we took further steps to explore these data. For the analysis using concatenated data, we also conducted an analysis using only the SNPs ($N = 938{,}921$) found in a region of the scaffold that was colinear with a reference genome assembled for the coastal subspecies of Swainson's thrush (several structural variants distinguish inland and coastal subspecies of thrush on this chromosome and could affect recombination rates if they are ancestral in *Catharus*; Justen, Easton, and Delmore 2024). Also, for the SVDquartets analysis we increased the number of random draws of SNPs from 10 to 100.

2.3 | Gene Flow Analyses

Gene flow analyses were done using ABBA-BABA tests (the *D* statistic of Green et al. 2010; Durand et al. 2011) and the 1 Mb block jackknifing approach of Martin et al. (2013), generating a *Z* score used to determine whether *D* differs significantly from zero. Because we had two populations of *ustulatus* represented, we were also able to estimate the proportion of admixture, *f*, between select lineages and perform a similar block jackknife to estimate the mean and 95% CI of *f* (Martin et al. 2013). For both sets of analyses we used the code developed and available from Martin et al. (2013) and https://evomics.org/learning/population-and-speciation-genomics/2018-population-and-speciation-genomics/abba-baba-statistics/. As required for this approach, we reordered samples within VCF files so that the outgroup was the last sample. And because our datasets included some females, we diploidised the Z chromosome of females by forcing homozygous genotypes. Koppetsch, Malinsky, and Matschiner (2024) recently showed that high mutation rate variation among lineages could produce a false inference of gene flow when performing ABBA-BABA tests contrasting old lineages. Their models, however, used rate variations that are high relative to what is found between avian *families* (Lanfear et al. 2010), and, at their minimum, systems that are about five times older than the *Catharus-Hylocichla* clade. We do not expect substantial rate variation in this clade, given that these are endothermic animals of similar size with similar generation lengths and similar ecologies.

Our initial gene flow analyses used data from the entire genome (although separating data from the autosomes and Z chromosome). We also performed ABBA-BABA tests on subsets of the data. First, a subset of 100 kb autosomal windows supported *ustulatus* as sister to all other *Catharus* and may represent the true species tree (see Results). We compared admixture values of these windows to admixture values derived from 10 replicated datasets of random draws of equivalent autosomal windows (excluding those with *ustulatus* as sister to all *Catharus*). Differences were considered significant if the 95% CIs did not overlap or if $p < 0.01$ in a one-sample *t*-test (Sokal and Rohlf 1995). A subset of Z chromosome windows also supported *ustulatus* as sister to all other *Catharus*. We compared the admixture values of these windows with Z chromosome windows that did not place *ustulatus* as sister to all other *Catharus* (260 windows). Finally, a subset of 100 kb autosomal windows we examined supported a monophyletic clade of the five migratory species. We compared admixture values in these windows to admixture values derived from 100 replicated datasets of random draws of equivalent autosomal windows (excluding those where migrants were monophyletic).

2.4 | Estimating Recombination Rates

Recombination rates were estimated in a previous study (Justen, Easton, and Delmore 2024). Briefly, high-coverage whole-genome resequencing data were obtained from 15 *ustulatus* individuals (inland subspecies, matching subspecies used to assemble reference genome) and estimated recombination using LDhat (McVean and Auton 2007). This program uses a population genetic approach, using a Bayesian reversible-jump Markov chain Monte Carlo to fit a model of recombination rate variation across the genome. Justen, Easton, and Delmore (2024) used 'interval' to generate the likelihood file and 'rhomap' to estimate

recombination rates and summarised these rates into windows of 100 kb.

2.5 | Testing for Enrichment of Migration Genes and Specific Gene Ontologies (GO)

We tested for enrichment of migration-associated genes in windows that (1) likely match the true species tree and (2) where migrants are monophyletic. We used two lists of genes for this analysis. First, a recent GWAS used two subspecies of *ustulatus* that differ in several features of their migratory behaviour (the timing, distance, and orientation and features of wing morphology important for migration) to identify 192 genes linked to migration (Justen, Easton, and Delmore 2024). Second, outside Swainson's thrushes (*ustulatus*), four other studies have used genomic data to identify genes linked to migration in songbirds (European blackcaps, *Sylvia atricapilla*, Delmore et al. 2020; *Vermivora* warblers, Toews et al. 2019; common quail, *Coturnix coturnix*, Sanchez-Donoso et al. 2022; and willow warblers, *Phylloscopus trochilus*, Lundberg et al. 2023). Combined, 1784 unique genes were identified in these studies.

GO analyses were run in go:profiler (Raudvere et al. 2019) using ontologies for *C. ustulatus* and the rest of the genome as background. *p*-values were corrected for multiple testing using the Benjamini-Hochberg FDR method (Benjamini and Hochberg 1995).

2.6 | Estimating Dxy

We used pixy (Korunes and Samuk 2021) to estimate dxy between all pairs of migratory species. Estimates of dxy require data from all sites (not only invariant sites). Accordingly, we reran GATK's GenotypeGVCFs using the -all-sites flag. We ran a series of filters using both GATK (--max-alternate-alleles 4 --standard-min-confidence-threshold-for-calling 30) and VCFtools (--max-meanDP 100, --min-meanDP 7, --max-missing 0.75).

2.7 | Estimating Chronogram and Ancestral Reconstruction of Migratory Behaviour

Based on our analyses of gene flow (see RESULTS), the topology inferred using data from the Z chromosome is our best estimate of a bifurcating history of *Catharus* evolution. To estimate an approximate time scale of speciation events we used two methods: BEAST v2.6 (Bouckaert et al. 2019) to infer a chronogram of this topology. All variants from the Z chromosome were concatenated and analysed with a GTR+G4 site model, an optimised relaxed clock model, and 10 M generations (sampling every 10 k generations). Monophyletic constraints were enforced for *Catharus*, *ustulatus*, *aurantirostris+mexicanus+fuscater*, *fuscescens+bicknelli+minimus*, and *guttatus+occidentalis+frantzii+gracilirostris*. Due to a paucity of putative fossils for the genus, we calibrated the phylogeny with a uniform prior of 6–8 mya for the root node based on previous work by Voelker, Bowie, and Klicka (2013). Two runs with different seeds were completed. Convergence between runs was evaluated in Tracer v1.7 (Rambaut et al. 2018), and after removing 10% of samples as burnin the posteriors of the two runs were combined and a maximum credibility tree was estimated. We also estimated speciation times using the A00 method in BPP v4.7 (Flouri et al. 2018), in which the multispecies coalescent is used to infer divergence times on a fixed topology. For this analysis we thinned our set of SNPs on the Z chromosome such that there was a minimum of 10 kb between sites in order to better assure free recombination between loci. The analysis was run with inverse-gamma theta (alpha = 3, beta = 0.08) and tau (alpha = 3, beta = 0.03) settings, a burnin of 1000 generations, and a posterior of 10,000 generations (sampling every 10). The BPP tree was scaled so that the age of the root node is 7 mya (Voelker, Bowie, and Klicka 2013). We used both chronograms for ancestral state reconstruction of migratory behaviour in Mesquite v3.8 (Maddison and Maddison 2023) using both maximum parsimony and maximum likelihood methods.

Computational notes outlining all steps in the analyses can be found in the Supporting Information.

3 | Results

3.1 | Resolving the *Catharus* Phylogeny

Our final filtered dataset yielded 23,885,232 SNPs spread across 39 autosomal scaffolds, 1,656,772 SNPs on the Z chromosome and complete mitochondrial genomes for each sample. This dataset was thus several orders of magnitude larger than the most recent *Catharus* phylogenomics study that used 2111 UCE loci (Everson et al. 2019).

Starting with the mitochondrial gene tree, the topology we recovered matched previous mitochondrial trees constructed for *Catharus* (Everson et al. 2019; Voelker, Bowie, and Klicka 2013), with the inclusion of *fuscater* as sister to *aurantiirostris* (Figure 2a). Moving on to trees constructed using autosomal data (excluding the Z chromosome), all methods produced the same topology. This topology was different from the mitochondrial gene tree but matched results based on UCEs (Everson et al. 2019), with the inclusion of *fuscater* as sister to *aurantiirostris* and *mexicanus* (Figure 2b). The species *ustulatus* is sister to a clade including *fuscater*, *minimus* and *bicknelli* in this autosomal tree.

We documented less agreement among methods when constructing trees for the Z chromosome. The SVDquartets species tree based on thinned SNPs produced the same topology as the mitochondrial tree, with *ustulatus* as sister to all taxa excluding a clade that comprised *aurantirostris*, *mexicanus* and *fuscater* (Figure 2c). Other methods (RAxML concatenated tree of SNPs and ASTRAL-II species tree based on gene trees from 100 kb windows [*n* = 10,135 windows]) moved *ustulatus* to a basal position in the tree, sister to all other *Catharus* (Figure 2d). This basal position of *ustulatus* has only been documented once before, using sequence data from eight nuclear introns (Voelker, Bowie, and Klicka 2013); it was not recovered in the more recent UCE analysis (Everson et al. 2019). The distribution of these

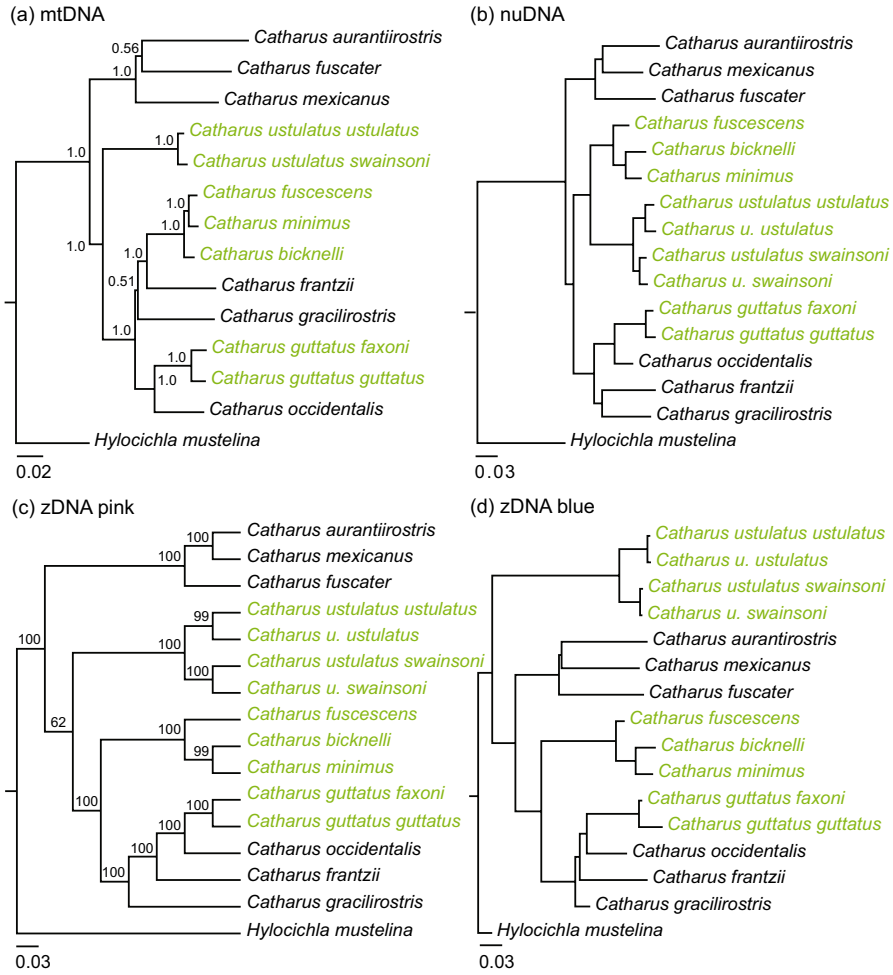

FIGURE 2 | Estimated topologies for separate genomic partitions. (a) Constructed using all protein coding regions in the mitochondrial genome, aligned using MAFFT ($N = 15{,}484$ bp). Model partitioning was determined using PartitionFinder, and a topology was estimated using MrBayes. (b) Dominant topology from autosomal data. The same topology was estimated using maximum likelihood and concatenated SNPs ($N = 24{,}104{,}438$), species tree methodology using topologies from 100,000 bp sliding windows ($N = 10{,}135$ windows), and quartet puzzling using SNPs with a maximum distance of 10,000 bp between each sampled position ($N = 94{,}138$). In each analysis all nodes were recovered with 100% bootstrap or 1.0 posterior probability support. Branch lengths shown are from the concatenated maximum likelihood analysis. (c, d) Topologies based on SNPs from the Z chromosome and matching either pink or blue topologies from Figure 3. (c) Matches mitochondrial tree and pink topologies in Figure 3. Based on quartet puzzling using SNPs with a maximum distance of 10,000 bp between each sampled position ($N \approx 6150$). This is the consensus topology based on 100 datasets, each with a random draw of SNPs and 100 bootstraps. (d) Unique topology based on a maximum likelihood estimate from all concatenated SNPs ($N = 1{,}872{,}822$), a maximum likelihood estimate based on concatenated SNPs from colinear regions of the scaffold ($N = 938{,}932$), and a species likelihood estimate based on trees from 100,000 bp windows ($N = 654$). Across all three analyses each node had a bootstrap percentage of 100%. Topology represented in blue in Figure 3. Numbers on nodes show posterior probabilities (a) or bootstrap percentages (c). Migratory species shown in green. Bars and 'P' annotations at right indicate the structure of two series of ABBA-BBA tests, both with *C. ustulatus* as P3 and using P2 and P1 clades as indicated: series1, series2 ('--' means not used in that series).

alternative placements for *ustulatus* across the genome is summarised in Figure 3.

3.1.1 | Gene Flow at the Whole-Genome Scale

We began our tests for gene flow around the movement of the origins of *ustulatus* among the three different topologies (Figure 2), hypothesising that gene flow involving this lineage in particular was probably important in the history of the group, given prior, more general evidence of gene flow in *Catharus* (Everson et al. 2019). A series of six ABBA-BABA tests examining evidence for gene flow between *ustulatus* (P3) and the *guttatus*-to-*gracilirostris* clade (P2) or the *aurantiirostris-mexicanus-fuscater* clade (P1; Figure 2d) showed strong evidence for gene flow with all of the P2 clade's members, suggesting that it occurred ancestrally to that clade (Table S2; Figure S1). Admixture proportions between *ustulatus* and the *guttatus*-to-*gracilirostris* clade (P2) were substantial, ranging from 11.2% to 13.9% (average 12.7%; Table S2). Their relative evenness also might indicate that this is the result of gene flow that primarily occurred ancestrally to the clade.

Our second series of tests for gene flow involved six contrasts between *ustulatus* (P3) and the *fuscescens-minimus-bicknelli* clade (P2) and the *guttatus*-to-*gracilirostris* clade (P1; Figure 2d). Here

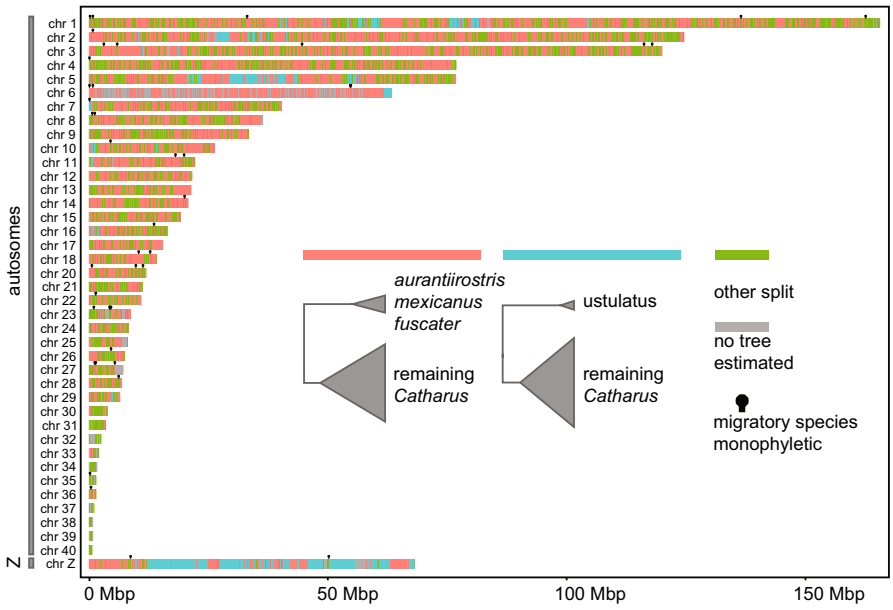

FIGURE 3 | Variation in the dominant phylogenetic signal recovered in windows across the genome. For 100,000 bp sliding windows across each scaffold, the different first splits within the *Catharus* clade are colour coded. This highlights the difference in the dominant signal across the autosomal (*aurantiirostris-mexicanus-fuscater* sister to all other species, pink) and Z (*ustulatus* sister to all other species, blue) scaffolds. No topologies were estimated for windows with fewer than six SNPs (shown in grey). Windows for which the migratory species form a clade are marked with a pin.

again there is strong evidence for gene flow, occurring between *ustulatus* and the *fuscescens-minimus-bicknelli* clade (P2) with remarkably high admixture proportions, ranging from 24.4% to 27.7% (average 26.3%; Table S2, Figure S1). Again, the ubiquity and relative magnitudes of gene flow and admixture proportions suggest that much if not all of this gene flow occurred ancestrally to the *fuscescens-minimus-bicknelli* clade.

The same series of 12 tests and estimates of admixture proportions were next done on the Z chromosome data. Although the same patterns of gene flow were present, admixture proportions were lower than those estimated for the autosomes throughout, ranging from 3.6% to 5.9% (avg. 5.0%; Table S2). The evenness of these admixture values for gene flow on the Z chromosome across the two sets of six tests is noteworthy (Table S2; Figure 4a).

Together these initial analyses provide compelling evidence that gene flow has had a major effect on the position of *ustulatus* in the autosomal phylogeny (Figure S1). Our final series of whole-genome ABBA-BABA tests replicated the series of tests done by Everson et al. (2019) using UCEs. These results demonstrated rampant gene flow throughout the genus (i.e., not only between *ustulatus* and other *Catharus*; only 6 of 61 tests were not significant) and also revealed that the amount of data UCEs bring to this question is inadequate to show the true extent of this gene flow (Table S3).

3.1.2 | Gene Flow in Windows That Reflect the Likely History of *Catharus*

We next switched to focus on more local genomic analyses, constructing topologies in windows along the genome (10,135,100 kb windows). Two main topologies were recovered: (1) trees where *aurantirostris*, *mexicanus*, and *fuscater* were sister to all other taxa (the mitochondrial topology [Figure 2a] and one topology recovered from the Z; Figure 2c), and (2) trees where *ustulatus* was sister to all other taxa (the other topology recovered from the Z; Figure 2d).

Windows on the Z chromosome primarily supported the second set of topologies, with *ustulatus* as sister to the rest of *Catharus* (392/652 windows; Figure 3). This same tree was also found on a subset of the macrochromosomes (323 windows; Figure 3), and results from ABBA/BABA tests suggest there is less gene flow in these windows. We still documented significant gene flow in all but one of the contrasts we ran (*D* values were positive and *Z* scores were > 3 or 4), but the admixture proportions we documented in the 392 windows that reflect the likely true species topology were lower than the 260 remaining Z windows with different topologies (Figure 4b; Table S4; all of the admixture estimates from the 392 windows have a 95% CI whose maximum value falls below that comparison's lower 95%CI for the remaining 260 windows on the *Z*, thus being significantly lower under a one-tailed framework). We documented similar results when running the same analysis using the 323 autosomal windows that placed *ustulatus* as basal to all other *Catharus*. Specifically, we generated 100 random datasets of 323 autosomal windows. All contrasts exhibit significant gene flow but lower admixture proportions than these random subsets (Figure 4c; Table S5).

In sum, the topology of the Z chromosome, with its substantially lower levels of gene flow (Figure 4), likely reflects the true evolutionary history of the group. Additional analyses contrasting admixture rates among subsets of autosomal and Z chromosome windows supporting conflicting topologies (*C. ustulatus* sister to rest vs. not) bear out this conclusion: the Z topology, with *C. ustulatus* sister to the rest of *Catharus*, is consistently associated with lower levels of gene flow (Tables S4 and S5).

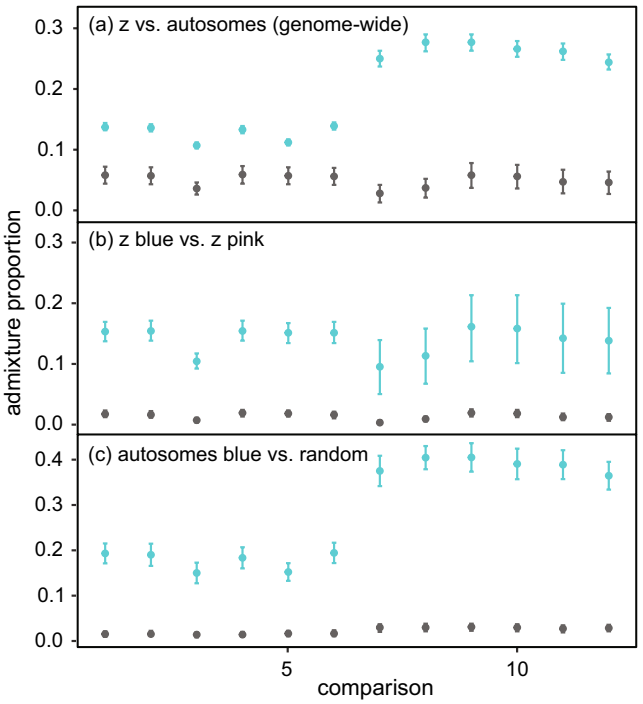

FIGURE 4 | Admixture proportions from ABBA/BABA tests. (a) Compares estimates between the entire Z chromosome (grey) to all autosomes (blue). (b) Compares windows on the Z where the topology either placed *ustulatus* as basal to the rest of the clade (grey) or internal to a clade with *aurantiirostris-mexicanus-fuscater* (blue). (c) Compares autosomal windows that placed *ustulatus* as basal to the rest of the clade (grey) and mean estimates from X random sets of autosomal windows (blue). 95% confidence intervals are shown. 'Z blue', 'Z pink', and 'autosomes blue' in panels b and c refer to colours from Figure 3.

3.2 | Examining Genomic Windows Reflecting the Likely History of *Catharus*

The topology with *ustulatus* basal to the rest of *Catharus* seems likely to be the true species tree, with gene flow being lower in windows having this topology, as above. Recombination rates in windows with *ustulatus* being basal to the rest of *Catharus* were also lower than in the rest of the genome (ANOVA $F_{4,9789} = 222.7$, $p < 0.0001$; ANOVA without Z chromosome, $F_{4,9137} = 109.4$, $p < 0.0001$).

With both gene flow and recombination being lower in these genomic windows, genes within them could be important for maintaining species boundaries. We did not document any enrichment of GO categories in these windows, but genes associated with migratory behaviour were enriched in these windows. Previous work on the genetics of migration in songbirds identified 1784 genes connected with migration (see Methods). Of these genes, 79 occurred in windows where *ustulatus* is sister to the rest of the genus. This number is more than expected by chance (74 expected by chance, 812/19,631 genes in windows and there are 1784 genes from the search [or 'draws']). Previous work on the genetics of migration in *Catharus* specifically identified 192 genes linked to migration. Of these genes, 81 occurred in windows where *ustulatus* is basal to the rest of the genus. This number is far more than expected by chance (8 expected by chance, 812/19,631 genes in windows, 192 draws).

3.3 | Examining Genomic Windows Where Migratory *Catharus* Are Monophyletic

Migratory *Catharus* species were monophyletic in 41/10,135,100 kb windows (~0.4% of the genome; Figure 3). This number of windows is more than expected by chance. Specifically, we simulated 100 sets of 10,135 random topologies and counted the number of topologies where migrants were monophyletic. The number ranged from just zero to four ($p < 0.0001$). These 41 windows include 215 genes. This number of genes is also more than expected by chance. We randomly sampled 41 windows 1000 times and counted the number of genes each time. The mean gene number in these random samples was 91 ($p < 0.0001$).

We did not find any overlap between these 215 genes and genes formally associated with migratory behaviour (in other songbirds or *ustulatus*). A GO analysis did identify enrichment of three terms: olfactory receptor activity (adjusted $p < 0.0001$, 9/36 genes with this ontology in the genome were found in these 41 windows), G protein-coupled receptor activity (adjusted $p = 0.033$, 15/411 genes) and serine-type endopeptidase activity (adjusted $p = 0.0018$, 10/138 genes).

If adaptive introgression is generating monophyly among migrants, we would expect to find evidence for gene flow in these windows. In support of this suggestion, *f* statistics from admixture analyses were higher than mean *f* statistics estimated from 100 random datasets in 10 of 12 different configurations (varying P1 and P2 but keeping *ustulatus* as P3 [ancestral]; Table S6). Admixture proportions between *ustulatus* and the *fuscescens-minimus-bicknelli* clade in most comparisons were remarkably high (40.2%–53.4%; Table S6). Deeper in the phylogeny, *ustulatus* shows lower admixture proportions in these 41 windows relative to members of the *guttatus*-to-*gracilirostris* clade, although they are still high relative to the autosomal averages, ranging from 20.6% to 41.8%, with an average delta *f* above random of 19.2% (Table S6).

Variation in admixture proportions among these contrasts is noteworthy in several respects. First, while *f* values are particularly elevated between migratory lineages, they are also elevated, though to a lower degree, among nonmigrants compared to the average autosomal levels (e.g., 20.6% & 24.8% in Table S6 vs. 11.2% and 13.3% in Table S2). When all three lineages were migrants (*guttatus* P1, *fuscescens* P2 and *ustulatus* P3) admixture was not different from random, indicating a lower relative difference between the P1 and P2 migratory lineages (Table S6). The only other nonsignificant contrast occurred in relation to *ustulatus* (P3) and nonmigrants *mexicanus* (P1) and *gracilirostris* (P2), again indicating a lower relative difference between the two and suggesting that the stepwise elevation of admixture proportion noted above seems to include a baseline of zero. Finally, while recombination would lead us to expect an initially elevated admixture proportion in these 41 windows to decline with time such that the deeper clade comparisons would be closer to the autosomal average (here represented by the random datasets), the opposite is the case, with the more recent clade's $f - f_{rand}$ values averaging just 0.156 versus the deeper clade's 0.193 (Table S6; these comparisons are not independent within groups).

We used estimates of sequence divergence (dxy) to gain further insight into the evolutionary forces affecting the 41 windows

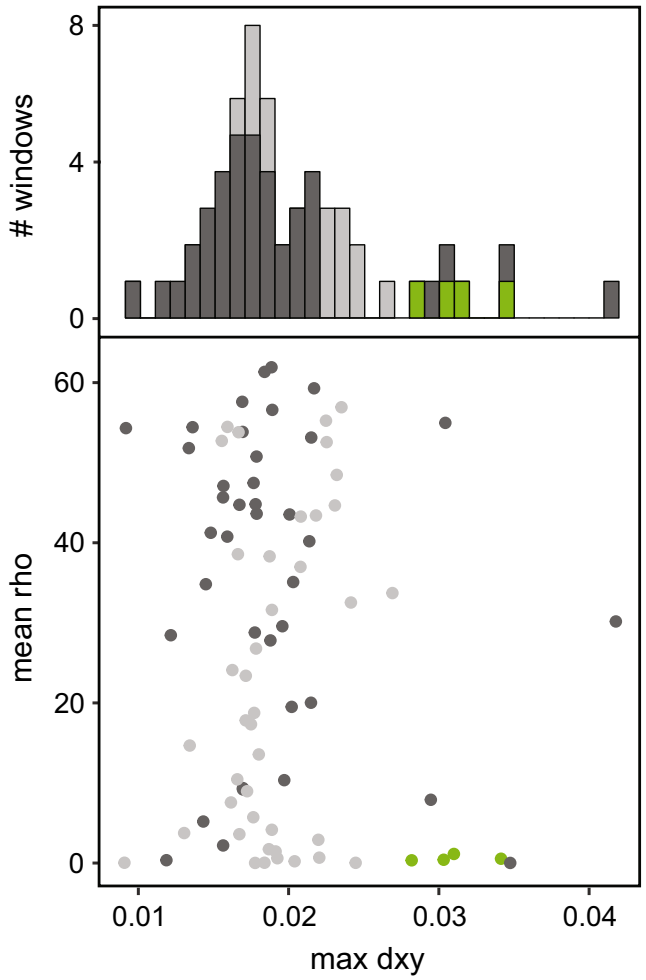

FIGURE 5 | Estimates of dxy from 41 windows where migrants formed a monophyletic clade. (a) Histogram showing maximum estimates from comparisons between all possible pairs of migratory *Catharus* in 41 windows matching this pattern. (b) Relationship between maximum estimates of dxy and recombination rates (rho). Estimates from windows including olfactory receptor genes are shown in green.

where migrants are monophyletic. Specifically, we estimated dxy between all possible pairs of migratory species in these windows, extracted the maximum value in each window, and compared these values to those estimated from a random set of 41 windows (Figure 5, random windows shown in light grey). A subset of the true windows ($N=8$) had higher values of dxy than the random windows. Higher estimates of sequence divergence can reflect strong balancing selection in specific genomic regions. Lower recombination rates can extend the effects of selection. In support of this suggestion, most of these eight windows exhibited lower rates of recombination than the random windows. Note, four of these eight windows are those that include the olfactory genes enriched in the monophyletic windows (Figure 5 shown in green).

3.4 | *Catharus* Chronogram and the Evolution of Migratory Behaviour

Using 1,656,748 biallelic SNPs from the Z chromosome we estimated a chronogram of *Catharus* evolution using concatenation in BEAST (Figure S2). Likely due to having only a single calibration point at the root node, the 95% highest posterior distribution intervals were wide across all nodes. The chronogram estimated with BPP (multispecies coalescent; 6151 biallelic SNPs) had more narrow error bars around node ages (Figure 2). This precision suggests more recent ages for presumed recent divergences (e.g., *ustulatus* subspecies, *fuscescens-bicknelli-minimus*), but in general older divergence times across most nodes. Ancestral reconstructions of migratory behaviour were estimated using these (Z chromosome) chronograms with maximum parsimony and maximum likelihood methods (Figure S2). Maximum parsimony does not consider branch lengths, and thus the results were the same on the BEAST and BPP trees. This method predicted the ancestral state of *Catharus* as migratory, with ambiguous results for the subsequent two nodes. Thus, it is equally parsimonious that there was an early switch to sedentary behaviour followed by two independent gains of migration (*fuscescens-bicknelli-minimus* and *guttatus*), or a longer retention of migratory behaviour followed by two switches to sedentary (*aurantirostris-mexicanus-fuscater*, and *guttatus* clade) and one gain of migration (*guttatus*). Maximum likelihood reconstructions take branch lengths into account, and the results were different using the BEAST and BPP trees. Using the BEAST tree, the ancestral state for *Catharus* is (marginally) migratory, with an early loss followed by two independent gains of migration (first maximum likelihood scenario described above). In the BPP tree some terminal branches are comparatively much longer and basal nodes are older. Maximum likelihood reconstruction on this tree favours migration as ancestral, a longer retention of ancestral migration, and several independent losses of migration. Future work with additional genomic sampling from closely related species that also differ in their migratory behaviour is needed to resolve both the timing of speciation events and the evolution of migratory behaviour in this group.

4 | Discussion

We used whole-genome sequences to examine the evolutionary history of *Catharus*, a genus of songbirds with considerable potential for gene flow and stark differences in seasonal migratory behaviour. We documented discordance between phylogenetic trees constructed using data from the mitochondria, autosomal and sex chromosomes. ABBA/BABA tests and local genomic analyses indicate that extensive gene flow (especially between *ustulatus* and the rest of the genus), recombination and selection generated this discordance. We interpret these results below, reconstructing the likely evolutionary history of the genus and explaining the probable origins of phylogenetic discordance. We conclude by discussing how our results relate to long-standing questions in evolution, including the genetics of adaptation and speciation, and highlighting avenues for future work.

4.1 | Evolutionary History of *Catharus*

We recovered three main trees in our analysis; these trees differ primarily in their placement of *ustulatus*. The same topologies for the mitochondrial and autosomal trees were recovered in a previous analysis using UCEs (Everson et al. 2019), but the additional resolution provided by whole-genome sequencing and new genetic tools for *Catharus* allow us to better elucidate,

understand and interpret these discordances. Our Z-based tree placed *ustulatus* as sister to the rest of the members of the genus (Figure 2d). This tree was found in a subset of the windows on the Z and autosomal chromosomes and likely reflects the true evolutionary history of the genus. Rates of both gene flow and recombination were lower in these windows, and genes that may help maintain species boundaries in *Catharus* were enriched in these windows (i.e., speciation genes, see below). Similar associations between gene flow, recombination rates and speciation genes have been documented in other systems (e.g., Fontaine et al. 2015; Edelman et al. 2019; Li et al. 2019). Lower recombination rates and speciation genes likely counter the homogenising effects of gene flow, helping maintain the true species tree in these windows.

The mtDNA and autosomal trees likely reflect the actions of gene flow moving *ustulatus* further into the tree (Figure S1). ABBA-BABA and admixture tests uncovered remarkably high levels of gene flow, especially between *ustulatus* and the rest of the genus (Table S2). The first putative gene-flow-mediated step resulted in the tree that moved *ustulatus* internal to the *aurantiirostris-mexicanus-fuscater* clade but basal to the rest of the species (as reflected in the mtDNA topology). This likely reflects a mitochondrial capture event, with the mitochondrial lineage present today in *ustulatus* being a replacement lineage from an ancestor deep in the tree after the split of the *aurantiirostris-mexicanus-fuscater* clade (Everson et al. 2019). Mitochondrial capture events and discordance between mitochondrial and autosomal trees have been documented in many other groups (e.g., Ferreira et al. 2018; Harris et al. 2018; Joseph 2021). The second putative gene-flow-mediated step resulted in the tree that places *ustulatus* as sister to the *fuscescens-bicknelli-minimus* clade. This tree was found on the autosomal partition and is likely the result of gene flow continuing to, in effect, move *ustulatus* up into a more recent portion of the phylogeny (Figure S1). In support of this hypothesis, we documented higher levels of ancestral gene flow between *ustulatus* and its closest relatives in this tree (i.e., with the *fuscescens-bicknelli-minimus* clade vs. the *aurantiirostris-mexicanus-fuscater* clade; Table S2). We refer to these topological shifts as steps, but they could prove to be the result of long periods of ongoing gene flow, historic and perhaps even contemporary (e.g., between *ustulatus* and members of the *fuscescens-minimus-bicknelli* clade). Ultimately, these results support the prevailing view that gene flow can have a large impact on phylogenetic inference, removing information about divergence and the true phylogenetic history of groups (Fontaine et al. 2015; Pinho and Hey 2010; Zhang et al. 2021). Our work was conducted at deeper evolutionary scales than much of the existing work on this topic, but it seems likely that future studies will find these characteristics in older clades.

4.2 | Genetics of Adaptation and Speciation

Genes previously connected to the migratory behaviour of songbirds were enriched in genomic regions that likely reflect the true species tree in *Catharus* (i.e., where *ustulatus* was basal to the rest of the clade). This finding suggests that variation in migration helps create and maintain reproductive isolation during divergence and speciation in the genus, fitting heteropatric speciation theory (Winker 2010). Identifying speciation genes is a major goal in evolutionary biology. Migration is often thought to promote gene flow, preventing divergence and blurring species boundaries; our results counter this suggestion and fall in line with work conducted on one species of *Catharus* in particular—*ustulatus*. *Ustulatus* has two subspecies that differ in their migratory routes. Hybrids take intermediate and ecologically inferior routes on migration, helping maintain reproductive isolation between the subspecies (Delmore and Irwin 2014; Justen, Lee-Yaw, and Delmore 2021; Blain et al. 2024). There are several additional ways migration could contribute to speciation (beyond reducing hybrid fitness). For example, species could breed at different times and/or in different locations. This is certainly the case for migratory and non-migratory species of *Catharus* and likely applies to several migrants as well (e.g., *minimus* and *bicknelli* have leapfrog migration and temporal separation; Winker 2010; and *fuscescens* and *minimus* likely migrate along similar routes but exhibit very little overlap on their breeding grounds). We look forward to future work interrogating the relationship between speciation and migration in *Catharus*. This work should involve additional sampling within each species and could, for example, focus on comparisons between *guttatus* and the remaining migratory *Catharus* species. *C. guttatus* appears to have evolved migratory behaviour independent of the remaining migrants and engages in less gene flow with *ustulatus*. Accordingly, these comparisons could be used to test for parallel patterns in the evolution of migratory behaviour in the genus.

Migrants were monophyletic in 41 windows. We documented evidence for increased admixture in these windows in most comparisons (Table S6), suggesting adaptive introgression may have generated these monophyletic relationships. But elevated admixture was not consistent across all contrasts, and indeed showed a somewhat stepwise relationship, with migrant contrasts being highest (~46% admixed), others involving residents being mid-level (~34%), and two contrasts (migrant-resident and migrant-migrant) being zero (Table S6). Thus, it seems that adaptive introgression at these loci was highest among migrants but also had value among some nonmigrants, too. Estimates of sequence divergence suggest additional/alternate evolutionary processes may also be acting. Specifically, estimates of sequence divergence between migrants were elevated in a subset of these windows, suggesting balancing selection (the maintenance of genetic variation) may be favouring the retention of ancestral polymorphisms here. Recombination rates were also lower in these regions and could indicate that selection in these windows is linked in nature, affecting nearby neutral sites. Interestingly, genes encoding olfactory receptors were enriched in these windows. Olfaction is an underappreciated sensory modality in birds. This modality is likely used in a wide range of contexts, including adaptation to new environments. Migrants have likely had to adapt to a changing array of habitats (e.g., higher-latitude breeding grounds, migratory stopover sites) repeatedly throughout their history, especially following particularly severe glacial cycles. This would involve adaptation to not only new plant assemblages and habitats but also an expansion of traits and/or habitat-specific cues in which olfaction can be important, such as food, predator avoidance, mate choice and species recognition (Amo et al. 2008, 2012; Grieves et al. 2022; Hagelin 2007; Krause et al. 2014; Soini et al. 2013). Balancing selection on olfactory genes has been documented in many non-avian systems. Perhaps selection has favoured the retention of genetic diversity

at these genes in *Catharus* as well. Note, lower rates of recombination in these regions suggest that structural variants like inversions may be helping maintain this variation. We look forward to future work using long-read data to test this suggestion.

A final note about results from our ABBA/BABA analyses. In general, selection against hybrids is expected to decrease introgressed genomic components through time (Moran et al. 2021; Orr 1995; Svedin et al. 2008). We see this expected relationship in the decreased admixture proportions that occur across the genome in *Catharus* thrushes between the older period of hybridisation tested and the younger (~13% vs. ~26%; Figure 4a, Table S2). However, we see two departures from this expectation in subsets of the genome. First, the Z chromosome shows lower and remarkably even levels of admixture over this same timeframe (Figure 4a, Table S2), suggesting narrower opportunities for adaptive introgression and likely heightened and long-standing hybrid disadvantages for effectively all members of the genus (i.e., 'speciation genes for all'). Second, this expectation is not met in the 41 genomic windows that unite migrants. Here, the older hybridisation period suggests a slightly higher admixture proportion retained than the younger period when compared to the average expectation for each contrast (~19% vs. ~16%; Table S2). This pattern suggests that these portions of the genome have attributes important for adaptation and speciation that vary by lineage and subgroup. While such variations are expected through mosaic evolution of the genome, identifying regions that depart from expected patterns (as we have done here in several ways) helps illuminate how adaptation and speciation have occurred in this clade in the presence of substantial levels of gene flow.

We have highlighted several areas for future work above, including the use of long reads to examine the role structural variants have played in *Catharus*. Long reads would also facilitate a reference-free analysis and could eliminate some potential bias we introduced in the present study. Specifically, we aligned our reads to the *ustulatus* reference genome. We made this decision to ensure we did not miss any *Catharus*-specific patterns but it may have biased our results to *ustulatus* (e.g., it is possible candidate migration genes are not in the same order in all Catharus and/or that recombination rates vary across the genus). Future work could use long reads to generate a pangenome for the species, eliminating this potential bias (Gong et al. 2023).

4.3 | Conclusion

We documented pervasive gene flow between nearly all lineages of *Catharus* here, even those that are more distantly related. Given these findings, it seems likely that previous demonstrations of avian speciation with gene flow using subsamples of the genome (e.g., Everson et al. 2019; Singhal et al. 2021) have revealed only the tip of the iceberg, and that, as here, evidence for its extent and magnitude will increase dramatically with whole-genome sequencing data. Considering that ABBA-BABA tests cannot detect gene flow between sister lineages and that traditional bifurcating trees can be poor representations of the diversification process when gene flow is occurring, results like these emphasise the importance of transitioning our thinking to complex, multidimensional networked evolutionary relationships, even in groups traditionally thought to predominantly undergo allopatric speciation (i.e., without gene flow; Mayr 1963). Our analyses of seasonal migration contribute to a growing sentiment in the literature that differences in this behaviour play an important role in speciation (Blain et al. 2024; Turbek, Scordato, and Safran 2018). Migration is often thought to promote gene flow, breaking down species boundaries. Our results counter this idea; not only were genes underlying this behaviour enriched in genomic regions that reflect the true species tree, but we also documented evidence for balancing selection at genes that may help species colonise new habitats. It was likely the evolution of migratory behaviour that allowed these species to colonise new habitats in the first place, initiating the process of adaptation and eventual speciation.

Author Contributions

K.E.D. and K.W. designed the study and wrote the paper with input from J.M.D.C. All authors helped perform research, and analyse the data.

This was was supported by grants to KED from NIH (1R35GM151012) and NSF (IOS-2143004).

Acknowledgements

The authors thank Matt Miller for suggestions early in our study design and advice along the way. Thanks to personnel at the Louisiana State University Museum of Zoology, the Museum of Southwestern Biology and the University of Alaska Museum for lending vouchered tissue samples for this study. This study was supported by an NSF CAREER awarded to KED (IOS-2143004) and an NIH MIRA awarded to KED (1R35GM151012).

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

Raw sequence reads and associated metadata have been uploaded to the SRA under (BioProject PRJNA979932 [https://dataview.ncbi.nlm.nih.gov/object/PRJNA979932?reviewer=3gftkemigkqb5q1ellp9g7evao] and PRJNA1112856). Notes on all computational analyses conducted can be found in the Supporting Information.

Benefit Sharing

Benefits from this research accrue from the sharing of our data and results on public databases as described above.

References

Abbott, R., D. Albach, S. Ansell, et al. 2013. "Hybridization and Speciation." *Journal of Evolutionary Biology* 26, no. 2: 229–246. https://doi.org/10.1111/j.1420-9101.2012.02599.x.

Abbott, R. J., N. H. Barton, and J. M. Good. 2016. "Genomics of Hybridization and Its Evolutionary Consequences." *Molecular Ecology* 25, no. 11: 2325–2332. https://doi.org/10.1111/mec.13685.

Amo, L., J. M. Avilés, D. Parejo, A. Peña, J. Rodríguez, and G. Tomás. 2012. "Sex Recognition by Odour and Variation in the Uropygial Gland Secretion in Starlings." *Journal of Animal Ecology* 81, no. 3: 605–613. https://doi.org/10.1111/j.1365-2656.2011.01940.x.

Amo, L., I. Galván, G. Tomás, and J. J. Sanz. 2008. "Predator Odour Recognition and Avoidance in a Songbird." *Functional Ecology* 22, no. 2: 289–293. https://doi.org/10.1111/j.1365-2435.2007.01361.x.

Arguedas, N., and P. G. Parker. 2000. "Seasonal Migration and Genetic Population Structure in House Wrens." *Condor* 102, no. 3: 517–528.

Barton, N. H., and K. S. Gale. 1993. "Genetic Analysis of Hybrid Zones." In *Hybrid Zones and the Evolutionary Process*, edited by R. G. Harrison, 13–45. Oxford: Oxford University Press.

Belliure, J., G. Sorci, A. P. Møller, and J. Clobert 2000. "Dispersal Distances Predict Subspecies Richness in Birds." *Journal of Evolutionary Biology* 13, no. 3: 480–487.

Benjamini, Y., and Y. Hochberg. 1995. "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 57, no. 1: 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Blain, S. A., H. C. Justen, W. Easton, and K. E. Delmore. 2024. "Reduced Hybrid Survival in a Migratory Divide Between Songbirds." *Ecology Letters* 27, no. 4: e14420. https://doi.org/10.1111/ele.14420.

Bouckaert, R., T. Vaughan, J. Barido-Sottani, et al. 2019. "BEAST 2.5: An Advanced Software Platform for Bayesian Evolutionary Analysis." *PLoS Computational Biology* 15, no. 4: e1006650.

Bravo, G. A., A. Antonelli, C. D. Bacon, et al. 2019. "Embracing Heterogeneity: Coalescing the Tree of Life and the Future of Phylogenomics." *PeerJ* 7: e6399.

Chifman, J., and L. Kubatko. 2014. "Quartet Inference From SNP Data Under the Coalescent Model." *Bioinformatics* 30, no. 23: 3317–3324.

Cutter, A. D., and B. A. Payseur. 2013. "Genomic Signatures of Selection at Linked Sites: Unifying the Disparity Among Species." *Nature Reviews Genetics* 14, no. 4: 262–274. https://doi.org/10.1038/nrg3425.

Danecek, P., A. Auton, G. Abecasis, et al. 2011. "The Variant Call Format and VCFtools." *Bioinformatics* 27, no. 15: 2156–2158.

Dasmahapatra, K. K., J. R. Walters, A. D. Briscoe, et al. 2012. "Butterfly Genome Reveals Promiscuous Exchange of Mimicry Adaptations Among Species." *Nature* 487, no. 7405: 94–98. https://doi.org/10.1038/nature11041.

Degnan, J. H., and N. A. Rosenberg. 2009. "Gene Tree Discordance, Phylogenetic Inference and the Multispecies Coalescent." *Trends in Ecology & Evolution* 24, no. 6: 332–340. https://doi.org/10.1016/j.tree.2009.01.009.

Delmore, K. E., J. W. Fox, and D. E. Irwin. 2012. "Dramatic Intraspecific Differences in Migratory Routes, Stopover Sites and Wintering Areas, Revealed Using Light-Level Geolocators." *Proceedings of the Royal Society B: Biological Sciences* 279, no. 1747: 4582–4589. https://doi.org/10.1098/rspb.2012.1229.

Delmore, K. E., and D. E. Irwin. 2014. "Hybrid Songbirds Employ Intermediate Routes in a Migratory Divide." *Ecology Letters* 17, no. 10: 1211–1218.

Delmore, K. E., J. C. Illera, J. Pérez-Tris, et al. 2020. "The Evolutionary History and Genomics of European Blackcap Migration." *eLife* 9: e54462. https://doi.org/10.7554/eLife.54462.

Durand, E. Y., N. Patterson, D. Reich, and M. Slatkin. 2011. "Testing for Ancient Admixture between Closely Related Populations." *Molecular Biology and Evolution* 28, no. 8: 2239–2252. https://doi.org/10.1093/molbev/msr048.

Edelman, N. B., P. B. Frandsen, M. Miyagi, et al. 2019. "Genomic Architecture and Introgression Shape a Butterfly Radiation." *Science* 366, no. 6465: 594–599. https://doi.org/10.1126/science.aaw2090.

Everson, K. M., J. F. McLaughlin, I. A. Cato, et al. 2019. "Speciation, Gene Flow, and Seasonal Migration in Catharus Thrushes (Aves: Turdidae)." *Molecular Phylogenetics and Evolution* 139: 106564.

Feng, S., J. Stiller, Y. Deng, et al. 2020. "Dense Sampling of Bird Diversity Increases Power of Comparative Genomics." *Nature* 587, no. 7833: 252–257.

Ferreira, M., A. M. Fernandes, A. Aleixo, et al. 2018. "Evidence for mtDNA Capture in the Jacamar *Galbula leucogastra*/Chalcothorax Species-Complex and Insights on the Evolution of White-Sand Ecosystems in the Amazon Basin." *Molecular Phylogenetics and Evolution* 129: 149–157.

Ferreira, M. S., M. R. Jones, C. M. Callahan, et al. 2021. "The Legacy of Recurrent Introgression During the Radiation of Hares." *Systematic Biology* 70, no. 3: 593–607. https://doi.org/10.1093/sysbio/syaa088.

Flouri, T., X. Jiao, B. Rannala, and Z. Yang. 2018. "Species Tree Inference With BPP Using Genomic Sequences and the Multispecies Coalescent." *Molecular Biology and Evolution* 35, no. 10: 2585–2593.

Fontaine, M. C., J. B. Pease, A. Steele, et al. 2015. "Extensive Introgression in a Malaria Vector Species Complex Revealed by Phylogenomics." *Science* 347, no. 6217: 1258524. https://doi.org/10.1126/science.1258524.

Gong, Y., Y. Li, X. Liu, Y. Ma, and L. Jiang. 2023. "A Review of the Pangenome: How It Affects Our Understanding of Genomic Variation, Selection and Breeding in Domestic Animals?" *Journal of Animal Science and Biotechnology* 14, no. 1: 73.

Gillespie, R. G., G. M. Bennett, L. De Meester, et al. 2020. "Comparing Adaptive Radiations Across Space, Time, and Taxa." *Journal of Heredity* 111, no. 1: 1–20.

Green, R. E., J. Krause, A. W. Briggs, et al. 2010. "A Draft Sequence of the Neandertal Genome." *Science* 328, no. 5979: 710–722. https://doi.org/10.1126/science.1188021.

Grieves, L. A., M. Gilles, I. C. Cuthill, T. Székely, E. A. MacDougall-Shackleton, and B. A. Caspers. 2022. "Olfactory Camouflage and Communication in Birds." *Biological Reviews* 97, no. 3: 1193–1209. https://doi.org/10.1111/brv.12837.

Hagelin, J. C. 2007. "The Citrus-Like Scent of Crested Auklets: Reviewing the Evidence for an Avian Olfactory Ornament." *Journal of Ornithology* 148, no. S2: 195–201.

Harris, R. B., P. Alström, A. Ödeen, and A. D. Leaché. 2018. "Discordance Between Genomic Divergence and Phenotypic Variation in a Rapidly Evolving Avian Genus (Motacilla)." *Molecular Phylogenetics and Evolution* 120: 183–195.

Harrison, R. G., and E. L. Larson. 2014. "Hybridization, Introgression, and the Nature of Species Boundaries." *Journal of Heredity* 105, no. S1: 795–809.

Irisarri, I., P. Singh, S. Koblmüller, et al. 2018. "Phylogenomics Uncovers Early Hybridization and Adaptive Loci Shaping the Radiation of Lake Tanganyika Cichlid Fishes." *Nature Communications* 9, no. 1: 3159. https://doi.org/10.1038/s41467-018-05479-9.

Jones, M. R., L. S. Mills, J. D. Jensen, and J. M. Good. 2020. "The Origin and Spread of Locally Adaptive Seasonal Camouflage in Snowshoe Hares." *American Naturalist* 196, no. 3: 316–332. https://doi.org/10.1086/710022.

Joseph, L. 2021. "Species Limits in Birds: Australian Perspectives on Interrelated Challenges of Allopatry, Introgression of Mitochondrial DNA, Recent Speciation, and Selection." *Auk* 138, no. 2: ukab012.

Justen, H., J. A. Lee-Yaw, and K. E. Delmore. 2021. "Reduced Habitat Suitability and Landscape Connectivity in a Songbird Migratory Divide." *Global Ecology and Biogeography* 30, no. 10: 2043–2056.

Justen, H. C., W. E. Easton, and K. E. Delmore. 2024. "Mapping Seasonal Migration in a Songbird Hybrid Zone—Heritability, Genetic Correlations, and Genomic Patterns Linked to Speciation." *Proceedings of the National Academy of Sciences of the United States of America* 121, no. 18: e2313442121. https://doi.org/10.1073/pnas.2313442121.

Katoh, K., and D. M. Standley. 2013. "MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability." *Molecular Biology and Evolution* 30, no. 4: 772–780.

Klicka, J., G. Voelker, and G. M. Spellman. 2005. "A Molecular Phylogenetic Analysis of the "True Thrushes" (Aves: Turdinae)." *Molecular Phylogenetics and Evolution* 34, no. 3: 486–500.

Koppetsch, T., M. Malinsky, and M. Matschiner. 2024. "Towards Reliable Detection of Introgression in the Presence of Among-Species Rate Variation." *Systematic Biology* 73: 769–788.

Korunes, K. L., and K. Samuk. 2021. "Pixy: Unbiased Estimation of Nucleotide Diversity and Divergence in the Presence of Missing Data." *Molecular Ecology Resources* 21, no. 4: 1359–1368.

Kozlov, A. M., D. Darriba, T. Flouri, B. Morel, and A. Stamatakis. 2019. "RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference." *Bioinformatics* 35, no. 21: 4453–4455.

Krause, E. T., C. Brummel, S. Kohlwey, et al. 2014. "Differences in Olfactory Species Recognition in the Females of Two Australian Songbird Species." *Behavioral Ecology and Sociobiology* 68, no. 11: 1819–1827.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2017. "PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses." *Molecular Biology and Evolution* 34, no. 3: 772–773.

Lanfear, R., S. Y. W. Ho, D. Love, and L. Bromham. 2010. "Mutation Rate Is Linked to Diversification in Birds." *Proceedings of the National Academy of Sciences of the United States of America* 107, no. 47: 20423–20428. https://doi.org/10.1073/pnas.1007888107.

Li, G., H. V. Figueiró, E. Eizirik, and W. J. Murphy. 2019. "Recombination-Aware Phylogenomics Reveals the Structured Genomic Landscape of Hybridizing Cat Species." *Molecular Biology and Evolution* 36, no. 10: 2111–2126.

Li, H. 2011. "A Statistical Framework for SNP Calling, Mutation Discovery, Association Mapping and Population Genetical Parameter Estimation From Sequencing Data." *Bioinformatics* 27, no. 21: 2987–2993.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows–Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079.

Lundberg, M., A. Mackintosh, A. Petri, and S. Bensch. 2023. "Inversions Maintain Differences Between Migratory Phenotypes of a Songbird." *Nature Communications* 14, no. 1: 452.

Maddison, W. P., and D. R. Maddison. 2023. Mesquite, a Modular System for Evolutionary Analysis, v2.73. http://mesquiteproject.wikispaces.com/.

Mallet, J. 2005. "Hybridization as an Invasion of the Genome." *Trends in Ecology & Evolution* 20, no. 5: 229–237.

Martin, S. H., K. K. Dasmahapatra, N. J. Nadeau, et al. 2013. "Genome-Wide Evidence for Speciation With Gene Flow in Heliconius Butterflies." *Genome Research* 23, no. 11: 1817–1828.

Mayr, E. 1963. *Animal Species and Evolution*. Harvard University Press.

McKenna, A., M. Hanna, E. Banks, et al. 2010. "The Genome Analysis Toolkit: A MapReduce Framework for Analyzing Next-Generation DNA Sequencing Data." *Genome Research* 20, no. 9: 1297–1303.

McVean, G., and A. Auton. 2007. *LDhat 2.1: A Package for the Population Genetic Analysis of Recombination*. Oxford: Department of Statistics.

Montgomery, T. H. 1896. "Extensive Migration in Birds as a Check Upon the Production of Geographical Varieties." *American Naturalist* 30, no. 354: 458–464. https://doi.org/10.1086/276410.

Moran, B. M., C. Payne, Q. Langdon, D. L. Powell, Y. Brandvain, and M. Schumer. 2021. "The Genomic Consequences of Hybridization." *eLife* 10: e69016. https://doi.org/10.7554/eLife.69016.

Morjan, C. L., and L. H. Rieseberg. 2004. "How Species Evolve Collectively: Implications of Gene Flow and Selection for the Spread of Advantageous Alleles." *Molecular Ecology* 13, no. 6: 1341–1356. https://doi.org/10.1111/j.1365-294X.2004.02164.x.

Nichols, R. 2001. "Gene Trees and Species Trees Are Not the Same." *Trends in Ecology & Evolution* 16, no. 7: 358–364.

Nosil, P., and D. Schluter. 2011. "The Genes Underlying the Process of Speciation." *Trends in Ecology & Evolution* 26, no. 4: 160–167. https://doi.org/10.1016/j.tree.2011.01.001.

Orr, H. A. 1995. "The Population Genetics of Speciation: The Evolution of Hybrid Incompatibilities." *Genetics* 139, no. 4: 1805–1813. https://doi.org/10.1093/genetics/139.4.1805.

Orr, H. A. 2005. "The Genetic Basis of Reproductive Isolation: Insights From Drosophila." *Proceedings of the National Academy of Sciences of the United States of America* 102, no. S1: 6522–6526. https://doi.org/10.1073/pnas.0501893102.

Outlaw, D. C., G. Voelker, B. Mila, and D. J. Girman. 2003. "Evolution of Long-Distance Migration in and Historical Biogeography of Catharus Thrushes: A Molecular Phylogenetic Approach." *Auk* 120, no. 2: 299–310.

Oziolor, E. M., N. M. Reid, S. Yair, et al. 2019. "Adaptive Introgression Enables Evolutionary Rescue From Extreme Environmental Pollution." *Science* 364, no. 6439: 455–457. https://doi.org/10.1126/science.aav4155.

Pfennig, K. S., A. L. Kelly, and A. A. Pierce. 2016. "Hybridization as a Facilitator of Species Range Expansion." *Proceedings of the Royal Society B: Biological Sciences* 283, no. 1839: 20161329. https://doi.org/10.1098/rspb.2016.1329.

Pinho, C., and J. Hey. 2010. "Divergence With Gene Flow: Models and Data." *Annual Review of Ecology, Evolution, and Systematics* 41, no. 1: 215–230. https://doi.org/10.1146/annurev-ecolsys-102209-144644.

Rambaut, A., A. Drummond, D. Xie, G. Baele, and M. Suchard. 2018. "Posterior Summarisation in Bayesian Phylogenetics Using Tracer 1.7." *Systematic Biology* 67, no. 5: 901–904.

Raudvere, U., L. Kolberg, I. Kuzmin, et al. 2019. "G: Profiler: A Web Server for Functional Enrichment Analysis and Conversions of Gene Lists (2019 Update)." *Nucleic Acids Research* 47, no. W1: W191–W198.

Ravinet, M., R. Faria, R. K. Butlin, et al. 2017. "Interpreting the Genomic Landscape of Speciation: A Road Map for Finding Barriers to Gene Flow." *Journal of Evolutionary Biology* 30, no. 8: 1450–1477. https://doi.org/10.1111/jeb.13047.

Rieseberg, L. H., and B. K. Blackman. 2010. "Speciation Genes in Plants." *Annals of Botany* 106, no. 3: 439–455. https://doi.org/10.1093/aob/mcq126.

Rockwell, R., and G. Barrowclough. 1987. "Gene Flow and the Genetic Structure of Populations." In *Avian Genetics, a Population and Ecological Approach*, edited by F. Cooke and P. Buckley, 223–255. New York: Academic Press.

Rohwer, S., and D. E. Irwin. 2011. "Molt, Orientation, and Avian Speciation." *Auk* 128, no. 2: 419–425. https://doi.org/10.1525/auk.2011.10176.

Ronquist, F., and J. P. Huelsenbeck. 2003. "MrBayes 3: Bayesian Phylogenetic Inference Under Mixed Models." *Bioinformatics* 19, no. 12: 1572–1574.

Sanchez-Donoso, I., S. Ravagni, J. D. Rodríguez-Teijeiro, et al. 2022. "Massive Genome Inversion Drives Coexistence of Divergent Morphs in Common Quails." *Current Biology* 32, no. 2: 462–469.e6. https://doi.org/10.1016/j.cub.2021.11.019.

Seehausen, O., R. K. Butlin, I. Keller, et al. 2014. "Genomics and the Origin of Species." *Nature Reviews Genetics* 15, no. 3: 176–192.

Singhal, S., G. E. Derryberry, G. A. Bravo, E. P. Derryberry, R. T. Brumfield, and M. G. Harvey. 2021. "The Dynamics of Introgression Across an Avian Radiation." *Evolution Letters* 5, no. 6: 568–581.

Soini, H. A., D. J. Whittaker, D. Wiesler, E. D. Ketterson, and M. V. Novotny. 2013. "Chemosignaling Diversity in Songbirds: Chromatographic Profiling of Preen Oil Volatiles in Different Species." *Journal of Chromatography A* 1317: 186–192. https://doi.org/10.1016/j.chroma.2013.08.006.

Sokal, R. R., and F. J. Rohlf. 1995. *Biometry: The Principles of Statistics in Biological Research*. New York, NY: WH Freeman and Co.

Svedin, N., C. Wiley, T. Veen, L. Gustafsson, and A. Qvarnström. 2008. "Natural and Sexual Selection Against Hybrid Flycatchers." *Proceedings of the Royal Society B: Biological Sciences* 275, no. 1635: 735–744. https://doi.org/10.1098/rspb.2007.0967.

Swofford, D. 2001. *Phylogenetic Analyzing Using Parsimony and Others Methods*. Sunderland, MA: Sinauer Associates.

Taylor, S. A., and E. L. Larson. 2019. "Insights From Genomes Into the Evolutionary Importance and Prevalence of Hybridization in Nature." *Nature Ecology & Evolution* 3, no. 2: 170–177.

Ting, C.-T., S.-C. Tsaur, and C.-I. Wu. 2000. "The Phylogeny of Closely Related Species as Revealed by the Genealogy of a Speciation Gene, Odysseus." *Proceedings of the National Academy of Sciences of the United States of America* 97, no. 10: 5313–5316. https://doi.org/10.1073/pnas.090541597.

Toews, D. P., S. A. Taylor, H. M. Streby, G. R. Kramer, and I. J. Lovette. 2019. "Selection on VPS13A Linked to Migration in a Songbird." *Proceedings of the National Academy of Sciences of the United States of America* 116, no. 37: 18272–18274.

Turbek, S. P., E. S. Scordato, and R. J. Safran. 2018. "The Role of Seasonal Migration in Population Divergence and Reproductive Isolation." *Trends in Ecology & Evolution* 33, no. 3: 164–175.

Voelker, G., R. C. K. Bowie, and J. Klicka. 2013. "Gene Trees, Species Trees and Earth History Combine to Shed Light on the Evolution of Migration in a Model Avian System." *Molecular Ecology* 22, no. 12: 3333–3344. https://doi.org/10.1111/mec.12305.

Winker, K. 2010. "On the Origin of Species Through Heteropatric Differentiation: A Review and a Model of Speciation in Migratory Animals." *Ornithological Monographs* 69, no. 1: 1–30.

Winker, K., and C. L. Pruett. 2006. "Seasonal Migration, Speciation, and Morphological Convergence in the Genus Catharus (Turdidae)." *Auk* 123, no. 4: 1052–1068.

Wu, C.-I. 2001. "The Genic View of the Process of Speciation." *Journal of Evolutionary Biology* 14, no. 6: 851–865. https://doi.org/10.1046/j.1420-9101.2001.00335.x.

Zhang, C., M. Rabiee, E. Sayyari, and S. Mirarab. 2018. "ASTRAL-III: Polynomial Time Species Tree Reconstruction From Partially Resolved Gene Trees." *BMC Bioinformatics* 19, no. S6: 153. https://doi.org/10.1186/s12859-018-2129-y.

Zhang, D., F. E. Rheindt, H. She, et al. 2021. "Most Genomic Loci Misrepresent the Phylogeny of an Avian Radiation Because of Ancient Gene Flow." *Systematic Biology* 70, no. 5: 961–975.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.