Molecular Phylogenetics and Evolution 139 (2019) 106564

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

Speciation, gene flow, and seasonal migration in *Catharus* thrushes (Aves:Turdidae)

Kathryn M. Everson[a,b,c], Jessica F. McLaughlin[a,b,d], Iris A. Cato[b], Maryanne M. Evans[b], Angela R. Gastaldi[e], Kendall K. Mills[a,b], Katie G. Shink[e], Sara M. Wilbur[b], Kevin Winker[a,b,*]

[a] University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA
[b] Department of Biology and Wildlife, University of Alaska Fairbanks, PO Box 756100, Fairbanks, AK 99775, USA
[c] University of Kentucky, Department of Biology, 101 T. H. Morgan Building, 675 Rose St, Lexington, KY 40508, USA
[d] Sam Noble Oklahoma Museum of Natural History, 2401 Chautauqua Ave, Norman, OK 73072, USA
[e] College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, PO Box 757220, Fairbanks, AK 99775, USA

ARTICLE INFO

Keywords:
Ecological speciation
Heteropatric speciation
Speciation with gene flow
Long-distance migration
Mitochondrial capture
Turdidae

ABSTRACT

New World thrushes in the genus *Catharus* are small, insectivorous or omnivorous birds that have been used to explore several important questions in avian evolution, including the evolution of seasonal migration and plumage variation. Within *Catharus*, members of a clade of obligate long-distance migrants (*C. fuscescens*, *C. minimus*, and *C. bicknelli*) have also been used in the development of heteropatric speciation theory, a divergence process in which migratory lineages (which might occur in allopatry or sympatry during portions of their annual cycle) diverge despite low levels of gene flow. However, research on *Catharus* relationships has thus far been restricted to the use of small genetic datasets, which provide limited resolution of both phylogenetic and demographic histories. We used a large, multi-locus dataset from loci containing ultraconserved elements (UCEs) to study the demographic histories of the migratory *C. fuscescens-minimus-bicknelli* clade and to resolve the phylogeny of the migratory species of *Catharus*. Our dataset included more than 2000 loci and over 1700 variable genotyped sites, and analyses supported our prediction of divergence with gene flow in the fully migratory clade, with significant gene flow among all three species. Our phylogeny of the genus differs from past work in its placement of *C. ustulatus*, and further analyses suggest historic gene flow throughout the genus, producing genetically reticulate (or network) phylogenies. This raises questions about trait origins and suggests that seasonal migration and the resulting migratory condition of heteropatry is likely to promote hybridization not only during pairwise divergence and speciation, but also among non-sisters.

1. Introduction

Catharus (Aves: Turdidae) is a New World genus of thrushes of Neotropical origin (American Ornithologists' Union, 1998). This group has become important in the study of seasonal migration, divergence, and speciation in birds (e.g., Winker, 2000, 2010; Outlaw et al., 2003; Winker and Pruett, 2006; Voelker et al., 2013; Ruegg et al., 2014; Delmore et al., 2015, 2016). *Catharus* thrushes are mostly small, omnivorous or insectivorous, forest-related birds renowned for the males' fluty songs (Lowther et al., 2001). The genus contains twelve species, including long-distance migrants in North America and sedentary nightingale-thrushes in Central and South America (Clement, 2000). Breeding grounds for the five migrant species range from Siberia, Alaska, northern Canada, and the United States; and wintering distributions include Mexico, Central America, and South America. The seven nightingale-thrush species inhabit Mexico and the northern Neotropical region (Clement, 2000).

Molecular systematics research has shown that the migratory and tropical resident lineages in *Catharus* are intermixed in phylogenies (Outlaw et al., 2003; Winker and Pruett, 2006; Voelker et al., 2013). Among these studies, phylogenetic relationships within *Catharus* have been fairly concordant, with the noteworthy exception of the position of *C. ustulatus* (e.g., Outlaw et al., 2003; Winker and Pruett, 2006; versus Voelker et al., 2013). Although some studies have lumped the Wood Thrush (*Hylocichla mustelina*) with *Catharus* (Winker and Rappole, 1988), it is currently classified in a separate, monotypic genus (American Ornithologists' Union, 1998) that is sister to a monophyletic *Catharus* (Winker and Pruett, 2006; Voelker et al., 2013).

* Corresponding author at: University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA.
E-mail address: kevin.winker@alaska.edu (K. Winker).

https://doi.org/10.1016/j.ympev.2019.106564
Received 5 April 2019; Received in revised form 16 July 2019; Accepted 16 July 2019
Available online 19 July 2019

Across the class Aves, seasonal migration and the increased movements associated with it generally result in lower levels of divergence than among nonmigratory lineages (Montgomery, 1896; Belliure et al., 2000). Nevertheless, divergence and speciation commonly occur in these migratory lineages, despite the difficulty of their attaining long-term allopatric isolation. This causes a conundrum in avian speciation, because it has long been held that speciation in birds is "always allopatric" (Mayr, 2004:47). Diverging migratory lineages can occur either in allopatry or sympatry during their annual cycles. Adding this distributional condition (heteropatry, or seasonal sympatry) to speciation models enables us to focus on factors likely to allow divergence to proceed in migrants despite the likelihood of ongoing gene flow. Heteropatric speciation is a type of ecological speciation (and speciation with gene flow) in which barriers to gene flow evolve due to divergent selection (Schluter, 1996, 2001; Rundle and Nosil, 2005; Winker, 2010; Nosil, 2012). In avian migrants, this selection likely stems from ecological factors such as resource competition and the heterogeneous distributions of those resources in space and time (Winker, 2010). Key predictions from this theory are that migrants undergo a different mode of speciation than sedentary forms diverging in a classic allopatric manner, that gene flow should be present but low, and that population sizes should be fairly large (i.e., speciation is not enabled by rarity; Winker, 2010; Winker et al., 2013). Here we test these predictions in the genus *Catharus*, including all the migratory lineages and their closest relatives.

Catharus fuscescens, *C. minimus*, and *C. bicknelli* form a clade that originated within the last ~0.5–1 Mya based on mtDNA sequence data (Outlaw et al., 2003; Topp et al., 2013; Voelker et al., 2013). Despite minor phenotypic differences and long being recognized as a subspecies of *C. minimus* (Marshall, 2001), *C. bicknelli* was elevated to species level (American Ornithologists' Union, 1995) based on differences in morphology, vocalizations, habitat use, and migratory patterns (Ouellet, 1993; Evans, 1994). The breeding and wintering ranges of *C. minimus* occur to the north and south of the breeding and wintering ranges of *C. bicknelli*, a common distributional condition among diverging migratory lineages known as leapfrog migration. Leapfrog distributions between diverging lineages were considered important in the development of heteropatric speciation theory (because strict allopatry becomes difficult to achieve), and the case between *C. minimus* and *C. bicknelli* was used as a focal example (Winker, 2010). The breeding range of *C. fuscescens* is also traversed by migrating *C. minimus*, and it partially overlaps that of *C. bicknelli*. In this small, relatively recent three-species clade of obligate migrants we hypothesized that gene flow would be evident in their divergence history and that diverging populations sizes would not be especially small. We also hypothesized that with two other species of long-distance migrants in the genus (*C. ustulatus* and *C. guttatus*), evidence of gene flow during the group's diversification might be present among other lineages as well. Our results suggest extensive gene flow in the genus and indicate that this group will continue to play an important role in our developing understanding of avian speciation, gene flow, and trait origins in the presence of seasonal migration.

2. Materials and methods

Target enrichment of ultraconserved elements (UCEs; Faircloth et al., 2012) was our genomic tool of choice to simultaneously investigate phylogenetics, speciation, and gene flow at multiple levels in this genus. UCEs are a class of nuclear marker distributed broadly across the genome; they are typically transposon-free and are highly conserved across distantly related animals (Bejerano et al., 2004; Derti et al., 2006; Simons et al., 2006; McCormack et al., 2012; Faircloth et al., 2012; Harvey et al., 2016). Sequence data from UCEs and their flanking regions have proven useful for answering both deep and shallow phylogenetic questions (e.g., McCormack et al., 2013; White et al., 2017), and fine-scale questions below the species level, including estimation of gene flow between recently diverged lineages (e.g., Smith et al., 2014; Zarza et al., 2016; Winker et al., 2018, 2019). Large multilocus studies are finding increasing evidence of gene flow during the speciation process (Feder et al., 2012; Gagnaire et al., 2013; Martin et al., 2013; Mallet et al., 2015), departing from the classical framework of allopatric speciation without gene flow. UCE-based studies of non-model lineages in (or recently through) the process of speciation offer a promising opportunity to improve our understanding of this process. We focused our study at two scales, both narrowly on the *C. fuscescens-minimus-bicknelli* clade and more broadly on migratory lineages deeper in the phylogeny. We used sequence data and genotyped SNPs from > 2000 UCEs to accomplish this multi-level combination of phylogenetic and population genomics questions.

2.1. Data collection

We extracted whole genomic DNA from 24 frozen tissue samples from museum collections (Table S1). We chose multiple individuals of the focal migratory species and one or two representatives of sister sedentary species for each. For our main demographic analyses, our bioinformatics pipeline (see below) genotypes and phases SNPs in each locus, giving us sample sizes of two haplotypes per individual. This causes four individuals to produce 8 haplotypes, which is considered optimal for coalescent-based and population genomics analyses (Felsenstein, 2005, Nazareno et al., 2017). The 24 samples include representatives of 10 of the 12 recognized species of *Catharus* thrushes (missing *C. dryas* and *C. fuscater*, whose close relationship to other sedentary taxa has not been contested) and an outgroup *Hylocichla mustelina*.

To sequence a common set of UCEs in all 24 individuals, we used a probe set developed by Faircloth et al. (2012), which targets 5060 loci conserved across all tetrapods. Laboratory methods followed the protocols described by Glenn et al. (2019) and Winker et al. (2018). Multiplexing used the methods outlined by Glenn et al. (2019). Our target for sequencing depth was ~30×, and the average coverage achieved at SNP sites was 34.6× (±17.5). An Illumina HiSeq2000 (San Diego, CA) was used to sequence paired-end 150 bp reads.

2.2. Bioinformatics

The PHYLUCE bioinformatics pipeline (ver. 1.4.0; Faircloth, 2016) was used to process UCE data and generate sequence alignments for phylogenetic analyses. Reads were trimmed for adapter contamination and low-quality bases using Illumiprocessor (Faircloth, 2013), then assembled using Trinity (ver. 2.4.0; Grabherr et al., 2011) on the Galaxy web portal (Afgan et al., 2016). We isolated UCE loci from the resulting assemblies and discarded putative paralogs using PHYLUCE scripts, after which UCE loci were aligned and edge-trimmed using MAFFT (Katoh and Standley, 2013). We created a 95% complete dataset (each UCE locus contains data from at least 95% of individuals) and produced both locus-specific and concatenated nexus alignment files using PHYLUCE scripts.

We also performed SNP-based analyses. To call SNPs, we first created a reference sequence by combining the reads for the two *Hylocichla mustelina* individuals and assembling with Trinity as above. PHYLUCE was then used to match contigs to probes and extract a UCE-only reference sequence that was indexed with BWA (ver. 0.7.7; Li and Durbin, 2009) and SAMTOOLS (ver. 0.1.19; Li et al., 2009). The reads for each individual were aligned to the reference using BWA-MEM (Li 2013), and the resulting SAM alignments were converted to BAM with SAMTOOLS. Alignments were checked for BAM format violations, read groups header information was added, and PCR duplicates were marked for each individual using PICARD (ver. 1.106; http://broadinstitute.github.io/picard). We constructed two datasets consisting of: (1) the *C. fuscescens-minimus-bicknelli* clade; and (2) a final dataset containing all taxa. For each, we created a merged BAM file and then used the Genome Analysis Toolkit (GATK; ver. 3.4–0; McKenna et al., 2010) to

locate and realign around indels. SNPs were called using the UnifiedGenotyper, followed by annotation of SNPs and indels, then masking of indels. We restricted our datasets to high-quality SNPs (Q30) and performed read-backed phasing. We then filtered further with VCFtools (ver. 0.1.12b; Danecek et al., 2011), reducing our dataset to a complete matrix (100% representation of all individuals at all loci) with a minimum genotype quality (GQ) of 10 and thinning to one biallelic SNP per locus. Setting our quality control filters so stringently caused us to lose many of the targeted UCE loci but ensured high-quality data for subsequent analyses. These datasets were then used for multiple downstream analyses, primarily for demographic analyses.

2.3. Demographic analyses

For the three taxa in the *C. fuscescens-minimus-bicknelli* clade we estimated effective population sizes, times since divergence, and rates of gene flow using Diffusion Analysis for Demographic Inference (δaδi; Gutenkunst et al., 2009). A custom Python script (https://github.com/jfmclaughlin92/vcf_scripts/blob/master/find_chrom.py) was used to remove Z-linked loci by identifying loci aligned by BLASTn (Zhang et al., 2000) to the Z chromosome of *Taeniopygia guttata* (for Passeriformes; NCBI Annotation Release 103) and removing them from the VCF file. (In birds, the Z chromosome is the sex chromosome; it has a different inheritance scalar and its allele frequencies are affected by sample sex ratios, necessitating its removal from most population genomic analyses.) These SNP data were then converted into the joint site frequency spectrum (SFS) format using a perl script by Kun Wang (https://groups.google.com/forum/#!msg/dadi-user/p1WvTKRI9_0/1yQtcKqamPcJ), creating the input file for δaδi. Z-linked loci were included in other analyses.

Pairwise comparisons were performed for the three focal taxa (*C. bicknelli, C. minimus,* and *C. fuscescens*). We ran five two-population demographic models to determine what models best fit our data: (1) neutral (no divergence, or still strongly mixing; this is effectively a null, one-population model); (2) split with no gene flow (gene flow is often termed "migration" in population genetics and this analysis software, but we replace it with "gene flow" here so as not to confuse the term with seasonal migration, the life-history trait that we focus on in this study); (3) split with gene flow; (4) isolation with bidirectional gene flow and population growth; and (5) a custom split-bidirectional-gene-flow model (a derivative of split-with-gene-flow; Fig. 1). The neutral, split-with-gene-flow, and isolation-with-gene-flow-and-population-growth models are provided in the δaδi file Demographics2D.py as "snm," "split_mig," and "IM," respectively. The no-gene-flow model (2 above) uses the "split_mig" model with gene flow parameters set to zero. The split-bidirectional-gene-flow model (figshare https://doi.org/10.6084/m9.figshare.6453125.v1) adds bidirectional gene flow to the split-with-gene-flow model to examine potential asymmetry in gene flow. After identifying the best-fit model based on likelihood values over successive runs and confirming it using the Akaike Information Criterion (Table S2; Akaike, 1974; Burnham and Anderson, 2002), we ran this model ten times with 100 bootstrapped replicates to estimate the 95% confidence interval (CI) for each parameter using a custom Python script (https://github.com/jfmclaughlin92/thesis/blob/master/bootstrap_dadi.py). To translate model estimates to biological values, substitution rate was calculated by comparing the reference (*Hylocichla mustelina*) against the *Ficedula albicollis* genome (the most closely related high-quality genome available), and the length of sites adjusted by the proportion of SNPs used in the analyses. We used a generation time of 4, calculated using the typical passerine age at first reproduction of 1 year and an annual adult survival of 0.6.

2.4. Phylogenetic analyses

An unpartitioned maximum-likelihood (ML) phylogenetic analysis was performed on the 95% complete concatenated dataset using the GTR-GAMMA site-rate substitution model in RAxML v.7.0.4 (Stamatakis, 2014). We assessed node support by conducting 100 nonparametric bootstrapping replicates. SplitsTree4 (ver. 4.14.6; Huson and Bryant, 2005) was used as a complementary tool to visualize conflicting or non-treelike signals in the data (Huson & Bryant, 2005). Networks were constructed from the full concatenated dataset using uncorrected *p*-distances and the NeighborNet algorithm.

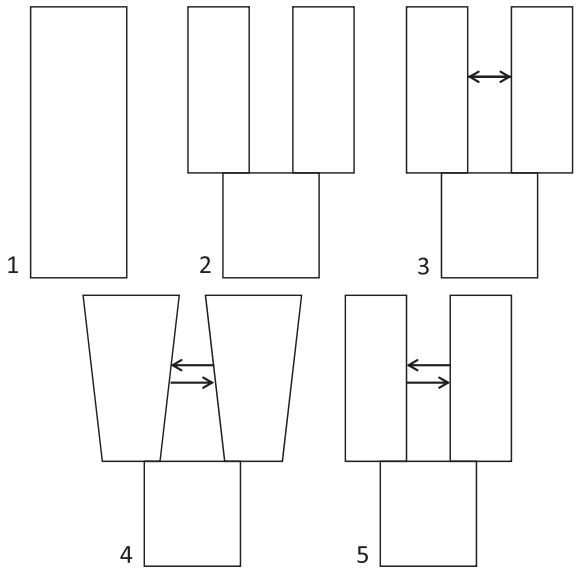

Fig. 1. The five population models tested using δaδi: (1) neutral (no divergence; a null, single-population model), then a series of two-population models with an ancestral population diverging into two populations (at time *T*): (2) split with no gene flow, (3) split with gene flow, (4) isolation with bidirectional gene flow and population growth, and (5) model 3 with bidirectional gene flow.

We also estimated a species phylogeny using a summary-based coalescent approach. Summary-based methods require well-resolved gene trees as input (Xi et al., 2015), and some UCE loci can contain weak phylogenetic signal. To address this potential problem, we calculated the number of parsimony-informative sites in each locus of our 95% complete dataset and isolated the loci with parsimony-informative sites in the upper quartile of the range. The best ML gene tree for each of these most-informative UCE loci was estimated using RAxML under the GTR-GAMMA site-rate substitution model. The 528 resulting gene trees were used as input to the summary-based species-tree program ASTRAL v4.10.12 (Mirarab et al., 2014). Support values in ASTRAL were calculated as Bayesian posterior probabilities on each species quartet, which has been shown to be a more robust measurement of support compared to bootstrapping (Sayyari and Mirarab, 2016).

A multispecies coalescent model in SNAPP (Bryant et al. 2012), implemented in BEAST2 (Bouckaert et al., 2014), was used to infer a species tree from biallelic SNP data (thinned to 1 SNP/locus). SNP data were converted from a VCF to a nexus file and uploaded into BEAUti, part of the BEAST2 package (Bouckaert et al., 2014), which was used to set prior parameter values and convert the data to an XML file. Default parameters were used and BEAST2 was set to run 50,000,000 generations, sampling every 1000 steps. Tracer (ver. 1.6; Rambaut et al., 2014) was used to view the MCMC output and check for convergence. DensiTree (ver. 2.2.6; Bouckaert and Heled, 2014) was used to visualize the full posterior distribution of tree outputs. TreeSetAnalyzer, which is part of SNAPP (Bryant et al., 2012), was used to produce the 95% credible set of trees output by SNAPP and to test whether published topologies (Winker and Pruett, 2006; Voelker et al., 2013) and an mtDNA topology were part of that set.

Unanticipated phylogenetic results from UCE loci (a novel relationship for *C. ustulatus*, detailed below) led us to re-examine lagniappe (non-targeted) sequences to determine what relationships the

single-locus mitogenome supported in these taxa. Using the mitogenome of the outgroup species *Turdus migratorius* (NCBI GenBank NC_024872-1), we followed a SNP-calling pipeline using GATK (v 3.4-0; McKenna et al., 2010) (https://gist.github.com/brantfaircloth/4315737) and PHYLUCE (Faircloth, 2016). The fasta files produced for each taxon were imported into Geneious (ver. 7.1.9; https://www.geneious.com) where they were aligned using MUSCLE (ver. 3.5; Edgar, 2004). We estimated an ML phylogeny using the GTR-GAMMA site-rate substitution model with 100 nonparametric bootstrapping replicates in RAxML v.7.0.4 (Stamatakis, 2014), as above.

Because our UCE dataset recovered a novel phylogenetic position of *C. ustulatus*, likely due to mtDNA-nuDNA topological discord (see below), we used our SNP dataset to perform a series of ABBA-BABA tests (Green et al., 2010; Durand et al., 2011) to assess whether gene flow might explain this. The ABBA-BABA test, also known as Patterson's *D* statistic, can be performed on any set of four lineages that are related in the form (((P1, P2), P3), P4). Using sequence data from one or a few samples per lineage (Zheng and Janke, 2018), the test counts the number of ABBA patterns (where P2 and P3 share a derived allele) and BABA patterns (where P1 and P3 share a derived allele). Under incomplete lineage sorting, we expect an equal number of ABBAs and BABAs; however, asymmetrical gene flow between P3 and either P1 or P2 would cause an excess of one pattern type (note that this does not test for gene flow between sister species). Because our dataset contained more than 4 species, we split the species tree into 10 well-supported, four-taxon subsets with *Hylocichla mustelina* always used as the outgroup (P4) and *C. ustulatus* always used as P3. Each four-taxon tree was analyzed separately using the used the R package evobiR (ver. 1.1; Blackmon and Adams, 2015). Because our dataset included multiple alleles for each species, in each analysis we calculated Patterson's *D* statistic for all possible combinations of alleles, then used a one-tailed student's *T*-test to determine whether the distribution of *D* statistics differed significantly from zero.

To further examine the prediction that gene flow has occurred in the presence of migratory lineages, we conducted additional sets of ABBA-BABA tests: (1) on the final migratory lineage not yet examined for gene flow (*C. guttatus*), (2) on non-migratory lineages to determine whether migration is a driving factor, and (3) on *C. fuscescens* and *C. minimus* to resolve the full pattern of ABBA-BABA-based inferences of gene flow in the group. In each case, the focal lineage was treated as P3. Tests were conducted using the above procedure, with *H. mustelina* again used as the outgroup (P4). Although *C. bicknelli* is also a migratory species, it was not included in this series of tests because its range does not overlap with any tropical resident *Catharus* species.

3. Results

3.1. Demographic analyses

Focusing on the *C. fuscescens-minimus-bicknelli* clade (Fig. 2), summary genetics statistics are given in Table 1. For both *C. minimus* and *C. fuscescens*, Tajima's *D* was negative, implying recent population contraction or purifying selection. F_{ST} values between each member of this triad were remarkably similar, being lowest between *fuscescens-minimus* (0.161), mid-range between *minimus-bicknelli* (0.176), and highest between *fuscescens-bicknelli* (0.188).

For all three pairwise comparisons in the *C. fuscescens-minimus-bicknelli* clade, our δaδi demographic analyses showed that the split-with-gene-flow model with a single gene flow parameter provided the best fit (Fig. 1, model 3), although we could not reject a bidirectional split-with-gene-flow model (Fig. 1, model 5; Table S2). We identified gene flow between lineages in all three of our pairwise comparisons, *C. bicknelli-minimus*, *C. bicknelli-fuscescens*, and *C. fuscescens-minimus* comparisons (Table 2), with *C. minimus* in both cases showing a somewhat lower signal of introgression from the other taxa than vice versa. Levels of gene flow inferred ranged from ~0.54 to 0.94 individuals per generation (Table 2). Effective population size estimates for each taxon were roughly consistent between pairwise comparisons (*C. bicknelli*: 30,307 and 20,935; *C. minimus*: 66,602 and 46,645; *C. fuscescens*: 183,004 and 169,953; Table 2). The split between *C. bicknelli* and *C. minimus* was estimated to have occurred 123,147 (± 19,123) yr ago, whereas the splits between these and *C. fuscescens* were estimated to be older: *C. minimus* 168,105 (± 13,892) yr; *C. bicknelli* 171,243 yr (± 22,198; Table 2).

3.2. Phylogenetic analyses

Our UCE dataset included 2111 UCE loci and 1,208,670 bp of aligned sequence data, roughly a 250-fold increase in the number of loci over the most recent *Catharus* study (Voelker et al., 2013). Our concatenated ML analysis of the 95% complete dataset converged on a strongly supported topology (Fig. 2). The summary-based species-tree reconstruction matched the concatenated topology, again with high support values (Fig. 3). The position of *C. ustulatus* in these trees has not been reported before (Outlaw et al., 2003; Winker and Pruett, 2006; Voelker et al., 2013).

Our SNAPP analysis recovered the same species tree, but also revealed extensive underlying conflict among gene trees, suggesting the possibility of gene flow among multiple lineages (Fig. 4). As a second way of visualizing conflict in the data, we estimated a neighbor network using the concatenated dataset in SplitsTree4 (Huson and Bryant, 2005; Fig. S1). This approach highlights most gene tree disagreements occurring early in the divergence of the fully migratory clade (*C. fuscescens*, *C. minimus,* and *C. bicknelli*) and among the four members of the clade containing *C. guttatus* (Fig. S1, contrasting with Figs. 2–4).

Our mitogenomic analysis produced a topology (Fig. S2) that did not match our UCE topology (Figs. 2 and 3). The mtDNA topology fell outside the 95% CI of the SNAPP topology; that of Voelker et al. (2013) did not, despite the different placement of *C. ustulatus* (sister to all other in-group members, Voelker et al., 2013: Fig. 1B and C) relative to our UCE topology (Figs. 2 and 3). This significant mismatch, and the mtDNA position of *C. ustulatus* (Fig. S2), suggests an old hybridization event (discussed below).

3.3. Hybridization

ABBA-BABA tests for hybridization between *C. ustulatus* and other species revealed significant gene flow ($P < 0.05$) with *C. gracilirostris*, and nearly significant gene flow ($P < 0.1$) with *C. frantzii* (Table 3). A second set of ABBA-BABA tests on *C. guttatus* suggested gene flow with *C. minimus* and perhaps also with *C. gracilirostris* (Table S3). Our third set of ABBA-BABA tests addressed whether or not gene flow has occurred among migratory and resident species in Middle America, and recovered gene flow among the following species : *C. aurantiirostris, C. mexicanus, C. frantzii, C. gracilirostris, C. guttatus,* and *C. ustulatus* (Fig. 5, Table S4). Finally, our fourth set of tests suggested that there may also have been gene flow between *C. minimus* and *C. frantzii* (Table S5). Suggestions of gene flow occurred in tests both with and without migratory species present (Tables S2 and S3, Fig. 5), and overall there was no evidence that gene flow events involved more migratory species than expected by chance (*G*-test with Williams' correction, $P > 0.5$). At this higher scale, then, beyond pairwise speciation events but more broadly in the genus, it is not yet clear what role seasonal migration plays in among-lineage hybridization.

4. Discussion

Our results provide several insights into speciation and diversification in the *Catharus* thrushes, including (1) the presence of speciation with gene flow in the most recent three-species migratory clade, none of which seem to have had especially small population sizes; (2) genomic confirmation of (remarkably distinct) species status for *C. bicknelli*; (3) a

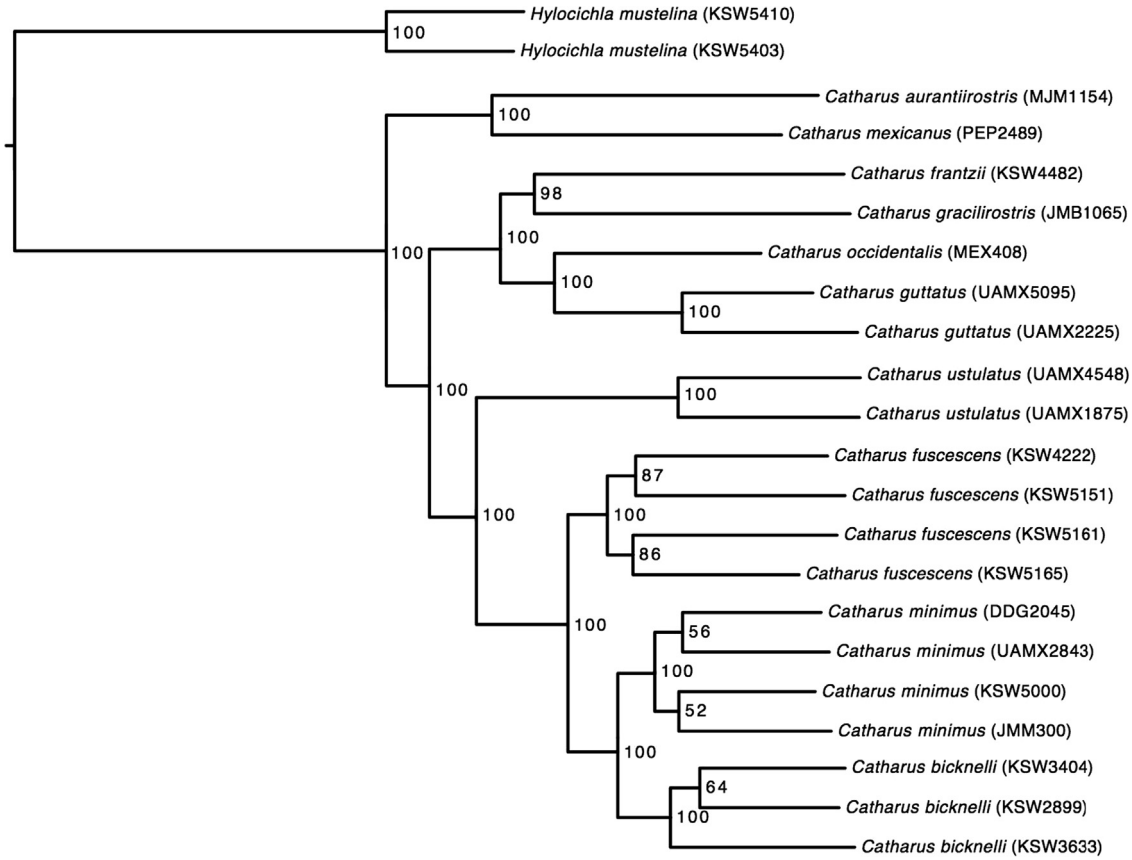

Fig. 2. Phylogeny of all individuals included in this study (see Table 1) based on a concatenated maximum-likelihood analysis of 2111 UCE loci. Node values denote bootstrap support.

Table 1
Summary genetics statistics for the *C. fuscescens-minimus-bicknelli* clade.

	Heterozygosity (H)	Watterson's estimator (Θ_W)	Nucleotide diversity (π)	Tajima's D
C. fuscescens	0.00020	0.00173	0.00166	−0.43823
C. minimus	0.00015	0.00170	0.00167	−0.22465
C. bicknelli	0.00014	0.00160	0.00161	n/a*
All three species	**0.00017**	**0.00180**	**0.00132**	**−1.27860**

* Calculation requires at least four individuals.

previously unrecognized phylogenetic position of *C. ustulatus,* with evidence for an old hybridization event (implying mitochondrial capture) driving mitonuclear gene-tree discord; and (4) suggestions of numerous occurrences of gene flow among *Catharus* lineages during diversification in the clade. In addition, our results allow us to consider the natural extension of heteropatric speciation theory beyond a basic two-lineage framework. Contra Mayr (2004), our results indicate that classic allopatric speciation has not been the only process involved in the divergence of this group.

4.1. Speciation with gene flow

Our results support theory developed specifically for migratory lineages (Winker, 2010) that suggested gene flow was likely during speciation in the *C. fuscescens-minimus-bicknelli* clade (Table 2). Our analysis of demographic histories in these lineages indicated low levels of gene flow in all three pairwise comparisons. Notably, during our study the first F_1 hybrid between *C. fuscescens* and *C. bicknelli* was discovered (Martinsen et al., 2018), and the first hybrid between *C. minimus* and *C. bicknelli* was also described (Fitzgerald et al., 2017), providing additional support for our findings.

Table 2
Estimates of population parameters from δaδi split-migration models for each pairwise comparison among *C. bicknelli, C. minimus,* and *C. fuscescens*. Interpreted values are presented first, with raw best-fit parameters as estimated by δaδi given below in parentheses, with 95% CI. Sample sizes (N, first column) represent individuals; however, because both alleles have been called, the sample sizes involved in model calculations are 2N.

	N	N_{ref} (θ)	N_1 (ν_1)	N_2 (ν_2)	T (years) (t)	M (indiv/gen) (m)
C. bicknelli/C. minimus	3:4	12,295 ± 1019 (91.29 ± 7.57)	30,307 ± 1827 (2.465 ± 0.149)	66,602 ± 4,495 (5.417 ± 0.366)	123,147 ± 19,123 (1.252 ± 0.194)	0.55 ± 0.13 (0.28 ± 0.066)
C. bicknelli/C. fuscescens	3:4	11,761 ± 1391 (87.33 ± 10.33)	20,935 ± 1652 (1.78 ± 0.140)	183,004 ± 9268 (15.56 ± 0.788)	171,243 ± 22,198 (1.82 ± 0.236)	0.94 ± 0.21 (0.218 ± 0.048)
C. fuscescens/C. minimus	4:4	11,546 ± 803 (85.73 ± 5.96)	169,953 ± 9471 (14.72 ± 0.820)	46,645 ± 2676 (4.04 ± 0.232)	168,105 ± 13,892 (1.82 ± 0.150)	0.54 ± 0.076 (0.116 ± 0.016)

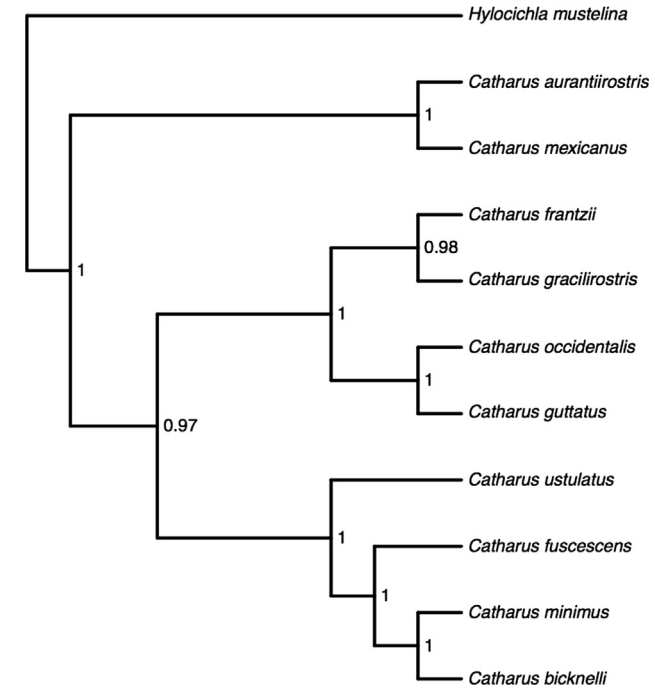

Fig. 3. Summary-based (ASTRAL) species topology of *Catharus* thrushes. Node values denote the Bayesian posterior probabilities of each species quartet. Branch lengths are not scaled. *Hylocichla mustelina* was used to root the phylogeny.

4.1.1. Population sizes

Heteropatric speciation theory posits that speciation involving migratory lineages is not enabled by especially small population sizes, and that in these situations population sizes are likely to be relatively large (Winker, 2010). It appears that population sizes of the two more widespread thrush species studied here, *C. fuscescens* and *C. minimus*, have fluctuated enormously in the past. Our long-term effective population size estimates of these species are two-three orders of magnitude smaller than their current census sizes: *C. fuscescens* is ~170 K from UCEs (Table 2) versus 12 M (Heckscher et al., 2017); *C. minimus* is ~56 K (Table 2) versus 16–20 M (Whitaker et al., 2018). Our long-term effective population size estimate for *C. bicknelli* (~25.6 K; Table 2) is much closer to its census size of 98–126 K (Townsend et al., 2015). For each case we would expect UCE-based estimates of effective population size to be smaller than census size. Purifying selection on UCEs (Katzman et al., 2007) will cause background selection on linked variation in the flanking regions (where the bulk of our SNPs occur), thus reducing (through hitchhiking) the estimated effective population size (Charlesworth and Charlesworth, 2016). In addition, long-term population size fluctuations caused by glacial cycles will also diminish effective size relative to census size during interglacials (i.e., now) in these species. Gene flow, however, can produce the opposite effect, inflating effective population size estimates (Leaché et al., 2014). Although it is possible that eventually corrections might be made to offset the distorting effects of selection on population size estimates made using loci with UCEs (Ewing and Jensen, 2015), for the present we consider the relative, among-lineage values to be informative. Of the three species, *C. bicknelli* populations appear to show the most long-term stability. Notably, none of the population sizes were especially small.

4.1.2. Split times and species status

Our estimates of split-time depths in the three-species clade (Table 2) ranged from 123 to 171 Kyr and are thus much shallower than the ~0.4–1 Mya estimates based on mtDNA (Outlaw et al., 2003; Topp et al., 2013; Voelker et al., 2013). Variation in divergence time

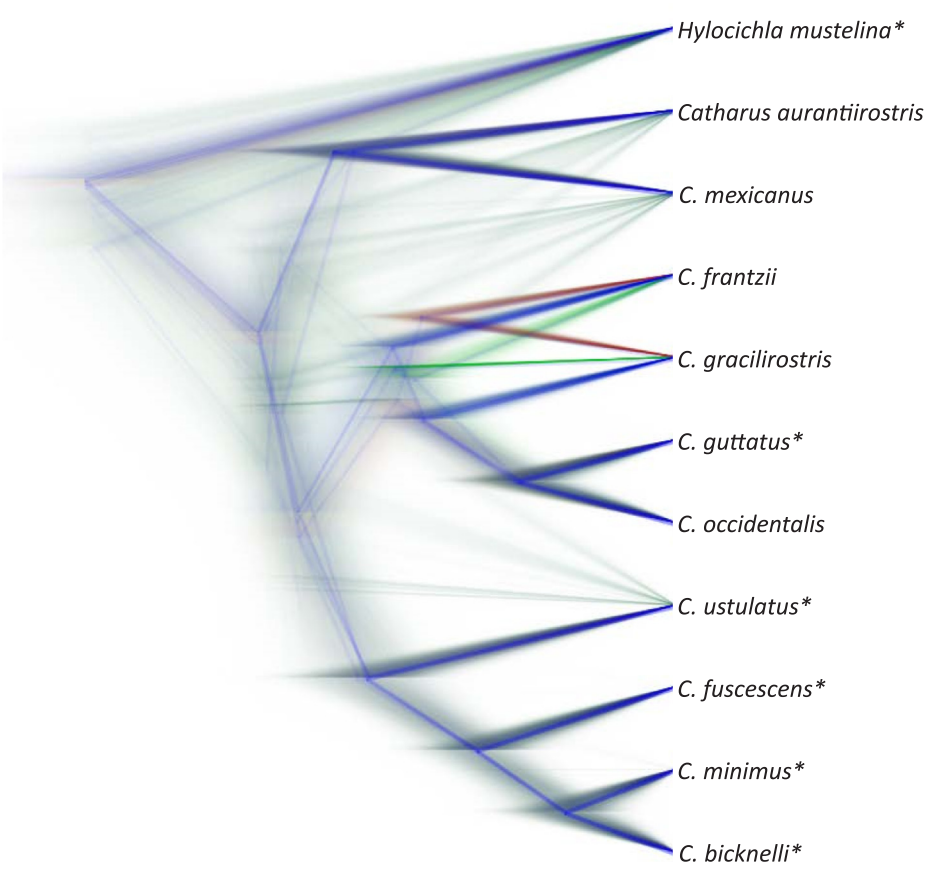

Fig. 4. A 50 M-generation SNAPP analysis visualized using DensiTree. Asterisks denote migratory species. Blue indicates the most popular topology in the tree dataset, red the secondmost, green the thirdmost, and dark green the rest. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Results of ABBA-BABA tests for gene flow in *C. ustulatus*. Significant *P*-values ($P < 0.05$) are highlighted in green, while nearly significant *P*-values ($P < 0.1$) are highlighted in yellow. Note that the ABBA-BABA test can only be used to detect gene flow in non-sister taxa, and that it will only detect gene flow that is asymmetrical between Species 1 – Species 3 and Species 2 – Species 3. The outgroup (P4) was *Hylocichla mustelina* in all analyses. A positive *D* statistic indicates that P2 and P3 share more alleles than would be expected by incomplete lineage sorting, while a negative *D* statistic indicates an excess of shared alleles between P1 and P3.

P1	P2	P3	total ABBA/ BABA sites	avg. # pairwise ABBAs (s.e.)	avg. # pairwise BABAs (s.e.)	D statistic	Z score	p-value
minimus	*fuscescens*	*ustulatus*	43	3.5977 (±0.0784)	4.5977 (±0.1404)	0.0291	0.0543	0.9567
frantzii	*gracilirostris*	*ustulatus*	15	3.1094 (±0.1776)	2.6719 (±0.1638)	0.0692	0.184	0.854
frantzii	*guttatus*	*ustulatus*	21	2.656 (±0.1224)	6.3906 (±0.1202)	-0.5175	-1.648	0.0993
frantzii	*occidentalis*	*ustulatus*	18	1.5156 (±0.1153)	4.8281 (±0.1095)	-0.596	-1.5375	0.1242
gracilirostris	*guttatus*	*ustulatus*	18	1.3125 (±0.1583)	5.875 (±0.1853)	-0.699	-2.0141	0.044
gracilirostris	*occidentalis*	*ustulatus*	18	2.1875 (±0.1552)	5.9375 (±0.1674)	-0.4914	-1.8035	0.0713
guttatus	*occidentalis*	*ustulatus*	15	2.6563 (±0.1144)	1.8438 (±0.1199)	0.2876	0.5409	0.5885
aurantiirostris	*mexicanus*	*ustulatus*	11	1.25 (±0.2159)	2.625 (±0.2649)	-0.3383	-0.4947	0.6208

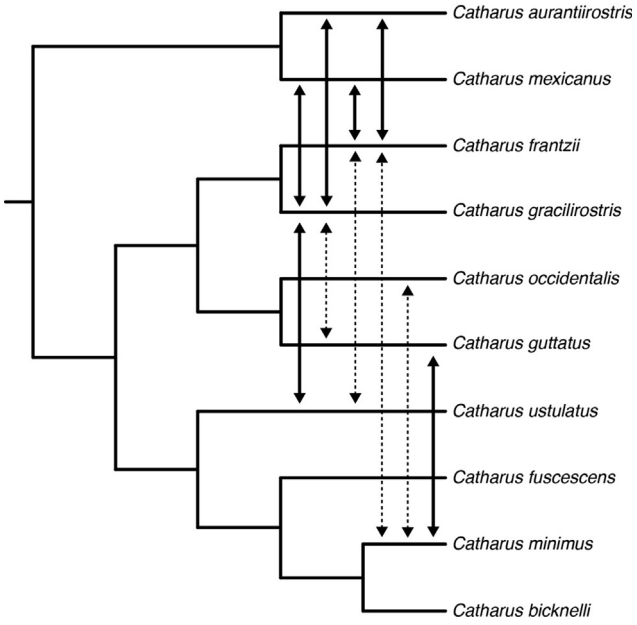

Fig. 5. ABBA-BABA test results. Solid arrows show the directions of inferred gene flow ($P < 0.05$), while dashed arrows show the directions of possible gene flow ($P < 0.1$). Note that the ABBA-BABA method tests only for gene flow among non-sister lineages. Branch lengths are not to scale. Tables 3 and S2–S4 show the results of all ABBA-BABA tests.

estimates can be caused by differences in dataset composition or analytical approach; for example, mtDNA datasets tend to produce older divergence time estimates compared to nuclear datasets (Zheng et al., 2011), and divergence times can be incorrectly estimated when gene flow is not accounted for (Leaché et al., 2014). As our study is the first to use a fully nuclear dataset while simultaneously parameterizing gene flow, this could explain the discrepancy in split-time estimates compared to previous work.

F_{ST} quantifies the fraction of differentiation attributable to population structure. Given the sister relationship and the close morphological similarities between *C. bicknelli* and *C. minimus*, we expected to find the lowest F_{ST} value between these two species; however, we instead found moderate and comparable levels of differentiation in all three pairwise comparisons. Given that phenotypically *C. fuscescens* is fairly distinct from both of the others, the similarity of F_{ST} values between each species pair is surprising.

Although *C. bicknelli* was long considered a subspecies of *C. minimus* (Wallace, 1939; AOU, 1957; Marshall, 2001), our results show that it is about as distinct genomically from *minimus* as both are from *fuscescens* (e.g., Table 2 and F_{ST} values), indicating that *bicknelli* is indeed a full species. Although low levels of gene flow were found among these three closely related species, it is generally accepted that full biological species can form and persist despite low levels of introgression; indeed, the morphological, behavioral, and ecological evidence (not evaluated in this study) continue to indicate the presence of three distinct species. Species delimitation, despite its fundamental importance to biological investigation and understanding, continues to be contentious (Coyne and Orr, 2004; de Queiroz, 2007; Sukumaran and Knowles, 2017). It is important to look at gene flow in recently diverged lineages, as we did here, while also considering phenotypic attributes in species delimitation. The role of gene flow during the speciation process in this group warrants continued study, given how important members of the genus were in the development of heteropatric speciation theory (Winker, 2000, 2010).

4.2. Phylogenetic analyses and mitochondrial capture

Phylogenetic analysis produced a strongly supported phylogeny for the 10 species of *Catharus* thrushes included in our study (Figs. 2 and 3). Most relationships agreed with previous phylogenetic hypotheses: *C. occidentalis* and *C. guttatus* were recovered as sister species (as in Outlaw et al., 2003; Winker and Pruett, 2006; Voelker et al., 2013), *C. frantzii* and *C. gracilirostris* were recovered as sister species (Voelker et al., 2013), and *C. bicknelli* and *C. minimus* form a clade sister to *C. fuscescens* (Voelker et al., 2013). *C. fuscater* and *C. dryas* were missing from our analysis and still await phylogenetic placement using UCEs, but their position as sister species in a clade with *C. aurantiirostris* and *C. mexicanus* (Winker and Pruett, 2006; Voelker et al., 2013) seems unlikely to change.

The only relationship we found that has not been reported by

previous phylogenetic studies is the placement of *C. ustulatus* as sister to the *C. fuscescens-minimus-bicknelli* clade. Prior studies have placed it elsewhere (Outlaw et al., 2003, Winker and Pruett, 2006; Voelker et al., 2013), likely, as our results show, due to the dominance of mtDNA in historic datasets and the mismatch between the mtDNA tree and the species tree (Fig. S2 versus Figs. 2 and 3). The relationship we found here may not come as a surprise to traditional taxonomists, who have long recognized a number of morphological and behavioral similarities between *C. ustulatus* and the *C. fuscescens-minimus-bicknelli* clade, including long-distance migratory behavior, a North American breeding distribution, and several pattern and color similarities (Winker and Pruett, 2006). Indeed, the specimen later designated as the holotype of *C. ustulatus* was initially mistaken as *C. fuscescens* (Nuttall, 1840).

Mitochondrial capture is the most likely cause of the disagreement between mitochondrial and nuclear phylogenies with regard to *C. ustulatus* (Fig. S2 versus Fig. 4); i.e., the mtDNA lineage in *C. ustulatus* was likely obtained from an older ancestor (i.e., deeper in the tree) than the seemingly more recent relationships found in most of its nuclear genome. Indeed, we found evidence of introgression between *C. ustulatus* and a non-sister clade, *C. frantzii* + *C. gracilirostris* (see below). Multiple other orders and families of birds show evidence of mitochondrial capture and disagreement between mtDNA and nuDNA trees (Shipham et al., 2017; Drovetski et al., 2018; Harris et al., 2018; Ferreira et al., 2018).

4.3. Additional gene flow and heteropatric speciation theory

Our SNAPP analysis recovered a large amount of topological discord throughout *Catharus* (Figs. 4 and S1), and our ABBA-BABA tests suggest that at least some of this is due to historic gene flow or hybridization (Fig. 5, Tables 3, S2–S4). All of the species for which we inferred gene flow through ABBA-BABA tests are either residents (*C. aurantiirostris, C. mexicanus, C. frantzii,* and *C. gracilirostris*) or migrants (*C. guttatus*, *C. minimus*, and *C. ustulatus*) in Middle America, so inferring that they have hybridized is not unreasonable geographically.

Heteropatric speciation theory (Winker, 2010) does not presently consider the likelihood that the gene flow promoted by seasonal migration might be an important feature not only during speciation, but also at a higher scale among related, non-sister lineages coming into seasonal contact. Our results suggest that this distributional condition of seasonal sympatry could promote hybridization and result in reticulate or networked genetic evolution among congeners. This is a logical extension of the theory, especially in birds, in which many species perform cyclic annual trans- and inter-continental movements and thus often occur together seasonally with members of the same genus. Successful avian hybridization (i.e., producing reproductively capable offspring) can occur long after the emergence of normally effective prezygotic reproductive isolating mechanisms (Price, 2008).

Our UCE dataset does not provide as much power over questions of gene flow using ABBA-BABA tests as methods that sample more of the genome and often produce an order of magnitude more data (e.g., GBS or RAD-seq). For this reason, we did not include multiple-test corrections to our four different sets of ABBA-BABA tests: it would likely cause rampant Type II error (incorrectly accepting the null hypothesis of no gene flow). From a conservative perspective, there are two fairly clear indications of gene flow, that of *C. ustulatus*, probably with *C. gracilirostris* and perhaps with an older ancestor (mtDNA and Table 3), and that of *C. aurantiirostris* and *C. gracilirostris* (Table S4). From this same statistically conservative perspective, the other suggestions of gene flow (Tables S2–S4, Fig. 5) are appropriate hypotheses for further testing. Much more detailed genomic study is needed to ascertain the frequencies of gene exchange among these lineages. Also, because this is to our knowledge the first examination of heteropatric speciation at two different scales (population genomics and systematics) in a genus of birds, other avian clades with and without migratory lineages are needed to determine the relative frequencies with which migratory versus non-migratory forms contribute to such patterns more broadly. It should also be noted that migratory species are not necessarily uninvolved in gene flow among sedentary species in our study (e.g., Table S4), because they might be important bridge taxa in these events, as has been found in bears (Kumar et al., 2017).

5. Conclusions

Heteropatric speciation theory considers that because seasonal migration promotes gene flow, when divergence progresses despite this inhibitor it will do so leaving signatures not predicted by allopatric speciation theory. Allopatry (and allochrony) remain important parts of heteropatric speciation theory, but, unlike allopatric speciation theory, the presence of gene flow is expected (together with phenotypic evidence that divergent selection is also occurring; Winker, 2010). Our expectation of low levels of gene flow in diverging migratory lineages was upheld in the three-species clade of obligate migrants (*C. fuscescens*, *minimus*, and *bicknelli*). Moreover, gene flow appears to have occurred (and might have been prevalent) elsewhere in the *Catharus* phylogeny as well, suggesting a logical extension of heteropatric speciation theory: seasonal migration is likely to promote hybridization not only during pairwise divergence and speciation, but also at a higher scale, among non-sisters (Fig. 5, Tables 3, and S2–S4). The result is a reticulate gene phylogeny, with disagreement among many gene trees influenced by hybridization events (Figs. 4 and S1). These findings add *Catharus* thrushes to an increasing number of vertebrate and invertebrate taxa exhibiting substantial levels of gene reticulation in their phylogenies (e.g., Mallet et al., 2015; Arnold, 2016; Kumar et al., 2017). This raises two important unanswered questions in the genus *Catharus*: Have we accurately recovered the species tree? And what genes have been passed among lineages?

As Maddison (1997:535) observed, "…the history of genetic descent does not take the form of a simple tree with sticklike branches." This issue is exacerbated by gene flow, because a phylogeny becomes less tree-like with more successful hybridizations (Mallet et al., 2015). We typically resolve conflicts among gene trees to produce a species tree using analyses that are effectively a "winner-take-all democracy" (Maddison, 1997:533; e.g., Figs. 2 and 3). However, with enough gene flow, the true species tree can be difficult to recover. And as a case in *Anopheles* mosquitoes showed, the genomic consensus tree may not be an accurate reflection of the species tree (Fontaine et al., 2015). Although we hope that a similar situation is not occurring in *Catharus* thrushes, and that the species trees in Figs. 2 and 3 are accurate, the indications of gene flow in complicating the group's history (Figs. 4 and 5, S1; Tables 3, and S2–S4) suggest that caution and further study are needed.

The network of trees produced by gene flow also makes it important to know what genes have been passed among lineages. Mallet et al. (2015:147) noted that "The origins of traits, and the genes that determine them can have very different histories from that of the species tree." There are numerous candidate genes associated with avian migration (Ruegg et al., 2014), including the presence of concentrated SNPs related to migratory orientation in *C. ustulatus* (Delmore et al., 2016). These genes (and/or their associated regulation; Lugo Ramos et al., 2017) could be readily passed among lineages and undergo selective sweeps (Bay and Ruegg, 2017). This leads to the intriguing possibility that *Catharus* thrushes are in some ways like *Heliconius* butterflies and *Anopheles* mosquitoes, in which key traits (e.g., wing patterns and vector attributes) have been shared among lineages through hybridization (Hines et al., 2011; Fontaine et al., 2015).

Thus, there is still work to be done in the *Catharus* thrushes to have confidence in the bifurcating species tree and in reconstructing the origins of phenotypic traits among these species. Previous work has suggested that several phenotypic characteristics, including long-distance migration, arose multiple times in this group (Winker and Pruett, 2006; Voelker et al., 2013). Full genomes and full representation of

genus members will be required to resolve not just the true phylogeny but also whether widely shared phenotypic characters (and associated gene complexes) arose multiple times as that previous work suggested, or rather arose just once and moved among lineages through hybridization.

6. Data accessibility

Our short-read sequence data are available from the NCBI sequence read archive (SRA; PRJNA553799, and Table S1). Additional data are available on figshare (https://doi.org/10.6084/m9.figshare.8980439): (a) the reference UCE sequence data fasta file, (b) the vcf file used in δaδi analyses, and (c) the individual UCE sequences in a fasta file.

Author contributions

KW conceived and designed the study. KW, KME, and JFM performed the bioinformatics. All authors participated in data analysis and interpretation and in writing the manuscript.

Acknowledgments

Funding: This project was partially supported by the U.S. National Science Foundation (DEB-1242267-1242241-1242260). The National Center for Genome Analysis and Support and the University of Alaska Fairbanks Research Computing Systems provided computational support.

We thank the field collectors and museum staff members who made this work possible, including at the University of Alaska Museum, the Field Museum of Natural History, and the Louisiana State University Museum of Natural Science. We also thank Brant Faircloth and Travis Glenn for production and invaluable assistance with data, and Chris Rimmer and Kent McFarland of the Vermont Center for Ecostudies for *C. bicknelli* specimens. We thank Michael Harvey and Matthew Miller for bioinformatics assistance and advice. Additional assistance in running δaδi was provided by Kevin Hawkins, Phil Lavretsky, Jeff Peters, and Ryan Gutenkunst. Helpful comments on a previous draft were given by Robert Gabbitas, Adam Haberski, Fern Spaulding, Giovanni Tundo, and two anonymous reviewers.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2019.106564.

References

Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., Chilton, J., Clements, D., Craor, N., Eberhard, C., Grüning, B., Guerler, A., Hillman-Jackson, J., Von Kuster, G., Rasche, E., Soranzo, N., Turaga, N., Taylor, J., Nekrutenko, A., Goecks, J., 2016. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses. Nucleic Acids Res. 44, 3–10. https://doi.org/10.1093/nar/gky379.

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Cont. 19, 716–723. https://doi.org/10.1109/TAC.1974.1100705.

American Ornithologists' Union, 1957. Check-list of North American Birds, fifth ed. American Ornithologists' Union, Baltimore, Maryland.

American Ornithologists' Union, 1995. Fortieth supplement to the American Ornithologists' Union check-list of North American birds. The Auk 112, 819–830.

American Ornithologists' Union, 1998. Check-list of North American Birds, seventh ed. American Ornithologists' Union, Washington, D.C.

Arnold, M.L., 2016. Divergence with Genetic Exchange. Oxford University Press, United Kingdom.

Bay, R.A., Ruegg, K., 2017. Genomic islands of divergence or opportunities for introgression? Proc. Royal Soc. B 284, 20162414. https://doi.org/10.1098/rspb.2016.2414.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325. https://doi.org/10.1126/science.1098119.

Belliure, J., Sorci, G., Møller, A.P., Clobert, J., 2000. Dispersal distances predict subspecies richness in birds. J. Evol. Biol. 13, 480–487. https://doi.org/10.1046/j.1420-9101.2000.00178.x.

Blackmon, H., Adams, A.A., 2015. EvobiR: tools for comparative analyses and teaching evolutionary biology. https://doi.org/10.5281/zenodo.30938.

Bouckaert, R.R., Heled, J., 2014. DensiTree 2: seeing trees through the forest. bioRxiv. https://doi.org/10.1101/012401.

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.-H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10, e1003537. https://doi.org/10.1371/journal.pcbi.1003537.

Bryant, D., Bouckaert, R., Felsenstein, J., Rosenberg, N.A., RoyChoudhury, A., 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29, 1917–1932. https://doi.org/10.1093/molbev/mss086.

Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer-Verlag, New York.

Charlesworth, B., Charlesworth, D., 2016. Population genetics from 1966 to 2016. Heredity 118, 2–9. https://doi.org/10.1038/hdy.2016.55.

Clement, P., 2000. Thrushes. Princeton University Press, Princeton, New Jersey.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates Inc, Sunderland, Massachusetts.

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Delmore, K.E., Hübner, S., Kane, N.C., Schuster, R., Andrew, R.A., Câmara, F., Guigó, R., Irwin, D.E., 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol. Ecol. 24, 1873–1888. https://doi.org/10.1111/mec.13150.

Delmore, K.E., Toews, D.P.L., Germain, R.R., Owens, G.L., Irwin, D.E., 2016. The genetics of seasonal migration and plumage color. Curr. Biol. 26, 2167–2173. https://doi.org/10.1016/j.cub.2016.06.015.

de Queiroz, K., 2007. Species concepts and species delimitation. Syst. Biol. 56, 879–886. https://doi.org/10.1080/10635150701701083.

Derti, A., Roth, F.P., Church, G.M., Wu, C.-T., 2006. Mammalian ultraconserved elements are strongly depleted among segmental duplications and copy number variants. Nature Genet. 38, 1216–1220. https://doi.org/10.1038/ng1888.

Drovetski, S.V., Reeves, A.B., Red'kin, Y.A., Fadeev, I.V., Koblik, E.A., Sotnikov, V.N., Voelker, G., 2018. Multi-locus reassessment of a striking discord between mtDNA gene trees and taxonomy across two congeneric species complexes. Mol. Phylogenet. Evol. 120, 43–52. https://doi.org/10.1016/j.ympev.2017.11.023.

Durand, E.Y., Patterson, N., Reich, D., Slatkin, M., 2011. Testing for ancient admixture between closely related populations. Mol. Biol. Evol. 28, 2239–2252. https://doi.org/10.1093/molbev/msr048.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340.

Evans, W.R., 1994. Nocturnal flight call of Bicknell's Thrush. Wilson Bull. 106, 55–61.

Ewing, G.B., Jensen, J.D., 2015. The consequences of not accounting for background selection in demographic inference. Mol. Ecol. 25, 135–141. https://doi.org/10.1111/mec.13390.

Faircloth, B.C., 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. https://doi.org/10.6079/J9ILL.

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646.

Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61, 717–726. https://doi.org/10.1093/sysbio/sys004.

Feder, J.L., Egan, S.P., Nosil, P., 2012. The genomics of speciation-with-gene-flow. Trends Genet. 28, 342–350. https://doi.org/10.1016/j.tig.2012.03.009.

Felsenstein, J., 2005. Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? Mol. Biol. Evol. 23, 691–700.

Ferreira, M., Fernandes, A.M., Aleixo, A., Antonelli, A., Olsson, U., Bates, J.M., Cracraft, J., Ribas, C.C., 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. Mol. Phylogenet. Evol. 129, 149–157. https://doi.org/10.1016/j.ympev.2018.07.007.

Fitzgerald, A.M., Whitaker, D.M., Ralston, J., Kirchman, J.J., Warkentin, I.G., 2017. Taxonomy and distribution of the imperiled Newfoundland Gray-cheeked Thrush, *Catharus minimus minimus*. Avian Conserv. Ecol. 12, 10. https://doi.org/10.5751/ACE-00976-120110.

Fontaine, M.C., Pease, J.B., Steele, A., Waterhouse, R.M., Neafsey, D.E., Sharakhov, I.V., Jiang, X., Hall, A.B., Catteruccia, F., Kakani, E., Mitchell, S.N., Wu, Y.-C., Smith, H.A., Love, R.R., Lawniczak, M.K., Slotman, M.A., Emrich, S.J., Hahn, M.W., Besansky, N.J., 2015. Extensive introgression in a malaria vector species complex revealed by phylogenomics. Science 347, 1258524. https://doi.org/10.1126/science.1258524.

Gagnaire, P.A., Pavey, S.A., Normandeau, E., Bernatchez, L., 2013. The genetic architecture of reproductive isolation during speciation-with-gene-flow in lake whitefish species pairs assessed by RAD sequencing. Evolution 67, 2483–2497. https://doi.org/10.1111/evo.12075.

Glenn, T.C., Nilsen, R., Kieran, T.J., Finger Jr., J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., García-De León, F.J., Portilla, M.A.D.R., Reed, K.D., Anderson, J.L., Meece, J.K., Aggrey, S.E., Rekaya, R., Alabady, M., Bélanger, M., Winker, K., Faircloth, B.C., 2019. Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). PeerJ (in review).

https://www.biorxiv.org/content/10.1101/049114v2https://doi.org/10.1101/049114.
Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N., Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., Regev, A., 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29, 644–652. https://doi.org/10.1038/nbt.1883.
Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H.Y., Hansen, N.F., 2010. A draft sequence of the Neandertal genome. Science 328, 710–722. https://doi.org/10.1126/science.1188021.
Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLOS Genet. 5, e1000695. https://doi.org/10.1371/journal.pgen.1000695.
Harris, R.B., Alström, P., Ödeen, A., Leaché, A.D., 2018. Discordance between genomic divergence and phenotypic variation in a rapidly evolving avian genus (*Motacilla*). Mol. Phylogenet. Evol. 120, 183–195. https://doi.org/10.1016/j.ympev.2017.11.020.
Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., Brumfield, R.T., 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. Syst. Biol. 65, 910–924. https://doi.org/10.1093/sysbio/syw036.
Heckscher, C.M., Bevier, L.R., Poole, A.F., Moskoff, W., Pyle, P., Patten, M.A., 2017. Veery (Catharus fuscescens), version 3.0. In: Rodewald, P.G. (Ed.), The Birds of North America. Cornell Lab of Ornithology, Ithaca, NY.
Hines, H.M., Counterman, B.A., Papa, R., de Moura, P.A., Cardoso, M.Z., Linares, M., Maller, J., Reed, R.D., Jiggins, C.D., Kronforst, M.R., McMillan, W.O., 2011. Wing patterning gene redefines the mimetic history of *Heliconius* butterflies. Proc. Natl. Acad. Sci. 108, 19666–19671. https://doi.org/10.1073/pnas.1110096108.
Huson, D.H., Bryant, D., 2005. Application of phylogenetic networks in evolutionary studies. Mol. Biol. Evol. 23, 254–267. https://doi.org/10.1093/molbev/msj030.
Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.
Katzman, S., Kern, A.D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R.K., Salama, S.R., Haussler, D., 2007. Human genome ultraconserved elements are ultraselected. Science 317, 915. https://doi.org/10.1126/science.1142430.
Kumar, V., Lammers, F., Bidon, T., Pfenninger, M., Kolter, L.K., Nilsson, M.A., Janke, A., 2017. The evolutionary history of bears is characterized by gene flow across species. Sci. Rep. 7, 46487. https://doi.org/10.1038/srep46487.
Leaché, A., Harris, R.B., Rannala, B., Yang, Z., 2014. The influence of gene flow on species tree estimation: a simulation study. Syst. Biol. 63, 17–30. https://doi.org/10.1093/sysbio/syt049.
Li, H., 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 2009. The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
Lowther, P.E., Rimmer, C.C., Kessel, B., Johnson, S.L., Ellison, W.G., 2001. Gray-cheeked Thrush (*Catharus minimus*), version 2.0. In: Rodewald, P.G. (Ed.), The Birds of North America. Cornell Lab of Ornithology, Ithaca, NY.
Lugo Ramos, J.S., Delmore, K.E., Liedvogel, M., 2017. Candidate genes for migration do not distinguish migratory and non-migratory birds. J. Comp. Physiol. A 203, 383–397. https://doi.org/10.1007/s00359-017-1184-6.
Maddisson, W.P., 1997. Gene trees in species trees. Syst. Biol. 46, 523–536. https://doi.org/10.1093/sysbio/46.3.523.
Mallet, J., Besansky, N., Hahn, M.W., 2015. How reticulated are species? BioEssays 38, 140–149. https://doi.org/10.1002/bies.201500149.
Marshall, J.T., 2001. The Gray-cheeked Thrush, *Catharus minimus*, and its New England subspecies, Bicknell's Thrush, *Catharus minimus bicknelli*. Nuttall Ornithological Club 28, 1–136.
Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., Blaxter, M., Manica, A., Mallet, J., Jiggins, C.D., 2013. Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. Genome Res. 23, 1817–1828. https://doi.org/10.1101/gr.159426.113.
Martinsen, E.S., McFarland, K.P., Rimmer, C.C., 2018. Documentation of a hybrid Bicknell's Thrush (*Catharus bicknelli*) × Veery (*C. fuscescens*) using vocalization and genetic data. Wilson J. Ornithol. 130, 70–80. https://doi.org/10.1676/16-061.1.
Mayr, E., 2004. 80 years of watching the evolutionary scenery. Science 305, 46–47.
McCormack, J.E., Faircloth, B.C., Crawford, N.G., Gowaty, P.A., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements are novel phylogenomic markers that resolve placental mammal phylogeny when combined with species tree analysis. Genome Res. 22, 746–754. https://doi.org/10.1101/gr.125864.111.
McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8, e54848. https://doi.org/10.1371/journal.pone.0054848.
McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., DePristo, M.A., 2010. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303. https://doi.org/10.1101/gr.107524.110.
Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, 541–548. https://doi.org/10.1093/bioinformatics/btu462.
Montgomery Jr., T.H., 1896. Extensive migration in birds as a check upon the production of geographical varieties. Am. Nat. 458–464.
Nazareno, A.G., Bemmels, J.B., Dick, C.W., Lohmann, L.G., 2017. Minimum sample sizes for population genomics: an empirical study from an Amazonian plant species. Mol. Ecol. Res. 17, 1136–1147. https://doi.org/10.1111/1755-0998.12654.
Nosil, P., 2012. Ecological Speciation. Oxford University Press, Oxford, England, UK.
Nuttall, T., 1840. A Manual of the Ornithology of the United States and Canada. Hilliard, Gray, and Co., Boston, Massachusetts.
Ouellet, H., 1993. Bicknell's Thrush: taxonomic status and distribution. Wilson Bull. 105, 545–754.
Outlaw, D.C., Voelker, G., Mila, B., Girman, D.J., 2003. Evolution of long-distance migration in and historical biogeography of *Catharus* thrushes: a molecular phylogenetic approach. Auk 120, 299–310. https://doi.org/10.2307/4090182.
Price, T., 2008. Speciation in Birds. Roberts and Company, Greenwood Village, Colorado.
Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from http://beast.bio.ed.ac.uk/Tracer.
Ruegg, K., Anderson, E.C., Boone, J., Pouls, J., Smith, T.B., 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol. Ecol. 23, 4757–4769. https://doi.org/10.1111/mec.12842.
Rundle, H.D., Nosil, P., 2005. Ecological speciation. Ecol. Lett. 8, 336–352. https://doi.org/10.1111/j.1461-0248.2004.00715.x.
Sayyari, E., Mirarab, S., 2016. Fast coalescent-based computation of local branch support from quartet frequencies. Mol. Biol. Evol. 33, 1654–1668. https://doi.org/10.1093/molbev/msw079.
Schluter, D., 1996. Ecological speciation in postglacial fishes. Philos. Trans. R. Soc. Lond. B 351, 807–814. https://doi.org/10.1098/rstb.1996.0075.
Schluter, D., 2001. Ecology and the origin of species. Trends Ecol. Evol. 16, 372–380. https://doi.org/10.1016/S0169-5347(01)02198-X.
Shipham, A., Schmidt, D.J., Joseph, L., Hughes, J.M., 2017. A genomic approach reinforces a hypothesis of mitochondrial capture in eastern Australian rosellas. The Auk 134, 181–192. https://doi.org/10.1642/AUK-16-31.1.
Simons, C., Pheasant, M., Makunin, I.V., Mattick, J.S., 2006. Transposon-free regions in mammalian genomes. Genome Res. 16, 164–172. https://doi.org/10.1101/gr.4624306.
Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements (UCEs) for comparative studies at shallow evolutionary time scales. Syst. Biol. 63, 83–95. https://doi.org/10.1093/sysbio/syt061.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Sukumaran, J., Knowles, L.L., 2017. Multispecies coalescent delimits structure, not species. Proc. Natl. Acad. Sci. 114, 1607–1612. https://doi.org/10.1073/pnas.1607921114.
Topp, C.M., Pruett, C.L., McCracken, K.G., Winker, K., 2013. How migratory thrushes conquered northern North America: a comparative phylogeography approach. PeerJ 1, e206. https://doi.org/10.7717/peerj.206.
Townsend, J.M., McFarland, K.P., Rimmer, C.C., Ellison, W.G., Goetz, J.E., 2015. Bicknell's Thrush (*Catharus bicknelli*), version 2.0. In: Rodewald, P.G. (Ed.), The Birds of North America. Cornell Lab of Ornithology, Ithaca, NY.
Voelker, G., Bowie, R.C., Klicka, J., 2013. Gene trees, species trees and Earth history combine to shed light on the evolution of migration in a model avian system. Mol. Ecol. 22, 3333–3344. https://doi.org/10.1111/mec.12305.
Wallace, G.J., 1939. Bicknell's Thrush, its taxonomy, distribution, and life history. Proc. Boston Soc. Nat. Hist. 41, 211–402.
Whitaker, D.M., et al., 2018. Gray-cheeked Thrush (*Catharus minimus*), version 2.0. In: Rodewald (Ed.), Cornell Lab of Ornithology, Ithaca, NY. https://doi.org/10.2173/bna.gycthr.02.
White, N.D., Mitter, C., Braun, M.J., 2017. Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae). J. Avian Biol. 48, 872–880. https://doi.org/10.1111/jav.01313.
Winker, K., 2000. Migration and speciation. Nature 404, 36. https://doi.org/10.1038/35003651.
Winker, K., 2010. On the origin of species through heteropatric differentiation: a review and a model of speciation in migratory animals. Ornithol. Monogr. 69, 1–30. https://doi.org/10.1525/om.2010.69.1.1.
Winker, K., Glenn, T.C., Faircloth, B.C., 2018. Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. PeerJ 6, e5735. https://doi.org/10.7717/peerj.5735.
Winker, K., Glenn, T.C., Withrow, J., Sealy, S.G., Faircloth, B.C., 2019. Speciation despite gene flow in two owls (*Aegolius* ssp.): Evidence from 2517 ultraconserved element loci. The Auk 136https://doi.org/10.1093/auk/ukz012. (in press).
Winker, K., McCracken, K.G., Gibson, D.D., Peters, J.L., 2013. Heteropatric speciation in a duck, *Anas crecca*. Mol. Ecol. 22, 5922–5935. https://doi.org/10.1111/mec.12525.
Winker, K., Pruett, C.L., 2006. Seasonal migration, speciation, and morphological convergence in the genus *Catharus* (Turdidae). The Auk 123, 1052–1068. https://doi.org/10.1642/0004-8038(2006) 123[1052:SMSAMC]2.0.CO;2.
Winker, K., Rappole, J.H., 1988. The relationship between *Hylocichla* and *Catharus* (Turdinae). The Auk 105, 392–394. https://doi.org/10.2307/4087513.
Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol. Phylogenet. Evol. 92, 63–71. https://doi.org/10.1016/j.ympev.2015.06.009.
Zarza, E., Faircloth, B.C., Tsai, W.L., Bryson, R.W., Klicka, J., McCormack, J.E., 2016. Hidden histories of gene flow in highland birds revealed with genomic markers. Mol. Ecol. 25, 5144–5157. https://doi.org/10.1111/mec.13813.

Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7, 203–214. https://doi.org/10.1089/10665270050081478.

Zheng, Y., Peng, R., Kuro-o, M., Zeng, X., 2011. Exploring patterns and extent of bias in estimating divergence time from mitochondrial DNA sequence data in a particular lineage: a case study of salamanders (order Caudata). Mol. Biol. Evol. 9, 2521–2535. https://doi.org/10.1093/molbev/msr072.

Zheng, Z., Janke, A., 2018. Gene flow analysis method, the *D*-statistic, is robust in a wide parameter space. BMC Bioinform. 19, 1–19. https://doi.org/10.1186/s12859-017-2002-4.