Research Article

Speciation despite gene flow in two owls (Aegolius ssp.): Evidence from 2,517 ultraconserved element loci

Kevin Winker,1,* Travis C. Glenn,2 Jack Withrow,1 Spencer G. Sealy,3 and Brant C. Faircloth4

1University of Alaska Museum, Fairbanks, Alaska, USA
2Department of Environmental Health Science, University of Georgia, Athens, Georgia, USA
3Department of Biological Sciences, University of Manitoba, Winnipeg, Manitoba, Canada
4Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana, USA

*Corresponding author: kevin.winker@alaska.edu

Submission Date: April 24, 2018; Editorial Acceptance Date: January 23, 2019; Published April 12, 2019

ABSTRACT

New study systems and tools are needed to understand how divergence and speciation occur between lineages with gene flow. Migratory birds often exhibit divergence despite seasonal migration, which brings populations into contact with one another. We studied divergence between 2 subspecies of Northern Saw-whet Owl (Aegolius acadicus), in which a sedentary population on the islands of Haida Gwaii, British Columbia (A. a. brooksi), exists in the presence of the other form (A. a. acadicus) during migration but not during the breeding season. Prior research showed fixed mtDNA divergence but left open the question of nuclear gene flow. We used 2,517 ultraconserved element loci to examine the demographic history of this young taxon pair. Although we did not observe fixed single nucleotide polymorphism differences between populations among our genotyped individuals, 100% of the birds were diagnosticable and δaδi analyses suggested the demographic model best fitting the data was one of split-bidirectional-migration (i.e. speciation with gene flow). We dated the split between brooksi and acadicus to ~278 Kya, and our analyses suggested gene flow between groups was skewed, with ~0.7 individuals per generation coming from acadicus into brooksi and ~4.4 going the opposite direction. Coupled with an absence of evidence of phenotypic hybrids and the birds’ natural history, these data suggest brooksi may be a young biological species arising despite historic gene flow.

Keywords: population genomics, seasonal migration, speciation

INTRODUCTION

The predominant model of avian speciation involves allopatry, which enables population divergence to proceed by preventing gene flow through isolation (Mayr 1963, Coyne and Orr 2004, Price 2008). Although decades of work demonstrate the importance of allopatric speciation, it is increasingly clear that divergence followed by speciation can occur despite the presence of gene flow (Feder et al. 2012, Nosil 2012, Seehausen et al. 2014, Zarza et al. 2016). These occurrences have given
rise to a variety of speciation-with-gene flow models, which consider how populations can diverge without long-term isolation (Gavrilets 2003, Winker 2010, Nosil 2012).

Migration is a common life-history strategy that is exhibited, for example, by >50% of the birds of the USA (338 of 650 species; Rappole et al. 1995). Migratory lineages are interesting for studying speciation because the great distances that these birds transit can increase the opportunity for gene flow between lineages, and this can mute the effects of population divergence (Montgomery 1896, Paradis et al. 1998, Belliure et al. 2000). In migratory lineages, diverging populations often have parapatric or heteropatric distributions. Among migrants, parapatry generally occurs when breeding ranges abut, and heteropatry occurs when 2 populations have allopatric breeding ranges with some seasonal sympatry occurring, especially during migration and wintering (Winker 2010). The distributional proximities in both of these situations give closely related populations enhanced opportunities for gene flow beyond the simple increases due to dispersal distance alone.

One potential example of speciation-in-progress that departs from traditional models of speciation in strict allopatry occurs in the Northern Saw-whet Owl (Aegolius acadicus), which has 2 subspecies, A. a. acadicus and A. a. brooksi. A. a. acadicus is largely migratory, breeds from southern Alaska to Nova Scotia south to California and Maryland, and is largely invariable in size or color across its range (Rasmussen et al. 2008). The subspecies A. a. brooksi is a resident (nonmigratory) population endemic to Haida Gwaii (Queen Charlotte Islands), British Columbia, that has distinctly darker, diagnostically different plumage (Fleming 1916, Withrow et al. 2014; Figure 1) and unique feeding habits (Hobson and Sealy 1991; Sealy 1998, 1999, 2013) relative to A. a. acadicus. The subspecies A. a. brooksi is considered threatened, whereas acadicus is not of conservation concern across its range (COSEWIC 2006, Rasmussen et al. 2008). These 2 taxa have a heteropatric distribution: nominate A. a. acadicus occur sympatrically with A. a. brooksi in small numbers during migration and winter (~6.1% of specimens; Sealy 1998, 2013, Withrow et al. 2014), although no hybrids are

**FIGURE 1.** Ventral and dorsal views of *Aegolius acadicus brooksi* (top pair) and *A. a. acadicus* (bottom pair). Top-to-bottom: female, male, female, male. Photo credit: K. Winker.
known from specimen records. Prior genetic research has shown shallow, fixed differences in mitochondrial DNA (mtDNA) sequences (Topp and Winker 2008, Withrow et al. 2014) that suggested these 2 groups split ~16,000 ya (Withrow et al. 2014), and amplified fragment length polymorphism (AFLP) data have suggested 2 distinct groups with 78% of individuals diagnosable (Withrow et al. 2014). These genetic and phenotypic (Figure 1) differences between A. a. acadicus and A. a. brooksi are likely related to Pleistocene glacial cycles: A. a. brooksi is thought to have been isolated in a forested Haida Gwaii refugium during (at least) the last glacial maximum, similar to other bird populations on Haida Gwaii that show genetic attributes consistent with refugial occupation (Burg et al. 2005, Pruett and Winker 2005, Burg et al. 2006, Topp and Winker 2008, Pruett et al. 2013). The degree of gene flow between A. a. acadicus and A. a. brooksi is unknown, as is the relative importance of ecological, behavioral, and geographic factors in their divergence.

Here, we use thousands of nuclear DNA markers to investigate the genetic differences between populations of A. a. acadicus and A. a. brooksi, and to estimate the occurrence and rate of gene flow between the 2 subspecies. We also test the fit of these genetic data to a variety of demographic models to determine whether allopatric or speciation-with-gene-flow frameworks apply to this system and to obtain a better understanding of the genetic factors that underlie the divergence of these 2 forms. Our prediction was that, given their distributions and life histories, these 2 forms would exhibit characteristics of divergence associated with speciation-with-gene-flow rather than classic allopatry.

METHODS

We extracted whole genomic DNA from 13 specimens (7 acadicus and 6 brooksi) used by Withrow et al. (2014) and 2 from an outgroup lineage, A. funereus (Figure 2; Appendix Table 2). A. a. acadicus were represented by University of Alaska Museum (UAM) numbers 8,990, 9,180, 13,949, 13,996, 17,882, 17,953, and 17,957; A. a. brooksi by 10,153, 19,042, 19,472, 19,474, 19,485, and 26,388; and A. funereus by 7,626 and 15,084. Because our bioinformatics pipeline (more below) genotypes and phases single nucleotide polymorphism (SNP) in each locus, this approach produces 2 sequences per individual at each locus. This exceeds the sample size of 8 haplotypes (= 4 diploid individuals if both haplotypes can be determined) deemed to be optimal for coalescent-based and population genomics analyses (Felsenstein 2005, Nazareno et al. 2017). After DNA extraction, we prepared dual-indexed DNA libraries from each extract following Glenn et al. (2017), quantified libraries using a Qubit fluorimeter (Invitrogen, Waltham, Massachusetts, USA), and we combined 8 libraries into equimolar pools of 500 ng each (62.5 ng per library) prior to enrichment. We enriched pools of samples for 5,060 ultraconserved element (UCE) loci using the Tetrapods-UCE-5Kv1 kit from MYcroarray following UCE enrichment protocol 1.5 and post-enrichment amplification protocol 2.4 (ultraconserved.org) with HiFi HotStart polymerase (Kapa Biosystems, Wilmington, Massachusetts, USA) and 14 cycles of post-enrichment PCR. We then quantified the fragment size distribution of the enriched pool on a Bioanalyzer (Agilent, Santa Clara, California, USA), and we qPCR-quantified the enriched pool using a commercial kit (Kapa Biosystems). We combined the enriched owl samples with enriched pools from other birds at equimolar ratios, and we sequenced the pool-of-pools using PE150 sequencing on an Illumina HiSeq 2500.

Following sequencing, we demultiplexed the resulting reads using Bcl2fastq 1.8.4 (Illumina), and we trimmed
demultiplexed reads for adapter contamination and low-quality bases using a parallel wrapper (Faircloth 2013) around Trimmomatic (Bolger et al. 2014). To create a reference set of sequences against which to call SNPs of individual birds, we chose 2 individuals of each subspecies (4 in total) having moderate fastq file sizes. Our reasoning was that these birds with moderate numbers of sequencing reads would optimize data gains vs. data losses in the bioinformatics pipeline rather than simply choosing the individual(s) with the most (or the least) sequence data as the reference. For example, a lot of high-quality data would be lost if calling SNPs against references created from the lowest quality data, and lower-quality loci and even individuals would be lost if trying to call SNPs against overly long reference sequences (due to lower coverage and greater uncertainties away from the UCE core). The 4 birds making up our reference were: brooksi (KSW3087, KSW3338) and acadicus (UAMX2975, UAMX2119). For these 4 individuals, we combined singleton reads that lost their mate with read 1 files, then combined the entire read 1 files and the 4 individual read 2 files into 2 separate read 1 and read 2 files, then we assembled these 2 read 1 and read 2 files de novo using Trinity 2.0.6 (Grabherr et al. 2013) on Galaxy (Afgan et al. 2016). Following assembly, we used Phyluce 1.4.0 (Faircloth 2016) to identify FASTA sequences from orthologous UCEs and remove FASTA sequences from non-UCE loci or potential paralogs. The resulting file was our reference set of UCE loci.

Next, we used Phyluce and programs that it calls (BWA, Li and Durbin 2010; SAMtools, Li et al. 2009; Picard, http://broadinstitute.github.io/picard) to align raw reads from individual libraries to our reference set of UCE loci. This workflow performed alignments of raw reads on a sample-by-sample basis using the bwa-mem algorithm (Li 2013); added header information to identify alignments from individual samples, cleaned, validated, and marked duplicates in the resulting Binary Alignment/Map (BAM) file using Picard; and merged all individuals into a single BAM file using Picard. Next, we used GATK 3.4-0 (McKenna et al. 2010) to identify and realign indels, call and annotate SNPs and indels, and mask SNP calls around indels using a part of a population genomics pipeline for UCEs developed by Faircloth and Michael Harvey (https://github.com/mgharvey/seqcap_pop). This process includes restricting data to high-quality SNPs (Q30) and read-back phasing in GATK. After calling and annotating SNPs, we followed Winker et al. (2018) and used VCFtools 0.1.12b (Danecek et al. 2011) to filter the resulting variant call format (VCF) file with the --max-missing (1.0) and --minGQ (10.0) parameters, which created a complete data matrix (all individuals had SNP calls at all loci) with a minimum genotype quality (GQ) of 10. We also used GATK’s “emit all confident sites” function to ensure that we only retained invariant loci with high-quality, rather than missing, data. Then we removed variable and invariant loci with incomplete data from downstream analyses and retained only loci with complete data. This finished the creation of our complete VCF file.

We calculated nucleotide diversity by creating a concatenated FASTA file of all loci at both genotyped alleles for all individuals using Catfasta2phyml by Johan Nylander (https://github.com/nylander/catfasta2phyml); this produced 2 complete UCE sequences (all loci concatenated) for each individual. We then analyzed these data in MEGA 6 (Tamura et al. 2013) using the maximum composite likelihood method. Next, using VCFtools on the complete VCF file, we calculated coverage depths, SNP positions within loci, and SNP-specific and locus-specific FST values. We thinned the VCF file to one SNP per locus, converted it to STRUCTURE format using PGDSpider 2.1.0.3 (Lischer and Excoffier 2012), then performed tests of Hardy-Weinberg equilibrium and computed observed and expected heterozygosities, homogeneity of variance, population structure (population FST, including the G-test; see Goudet et al. 1996), and individual assignment probabilities to populations using adegenet 2.0.1 (Jombart and Ahmed 2011).

We used the program Diffusion Approximations for Demographic Inference (δaδi; 1.7.0 (Gutenkunst et al. 2009) to infer demographic parameters under different divergence models. Z-linked loci were excluded from these demographic analyses (although included in other analyses) because they have a different inheritance scalar from autosomal loci and sample population sex ratios affect allele frequency estimates (e.g., Jorde et al. 2000, Garrigan et al. 2007). We identified Z-linked loci with a script from Jessica McLaughlin (https://github.com/jfmclaughlin92/thesis), which uses BLASTN 2.3.1 (Zhang et al. 2000) searches of the reference set of UCE loci against the chicken (Gallus gallus) genome (NCBI Gallus_gallus-5.0 reference Annotation Release 103), and we excluded UCE loci that strongly matched (E-values ~0.0) the chicken Z chromosome. After removing Z-linked loci from our complete VCF file, we converted the dataset to biallelic format and thinned the data to one SNP per locus using VCFtools (to minimize effects of linkage, as recommended in the δaδi user manual). We then converted this new, smaller VCF file to the joint site frequency spectrum (SFS) format required by δaδi using a PERL script by Kun Wang (https://groups.google.com/forum/#!msg/dadi-user/p1WvTKRI9_0/1yQtcKqamPcJ).

Prior research showed that these 2 owl subspecies represented 2 different populations (Withrow et al. 2014). We used δaδi to examine general 2-population divergence models to determine which fit the data best before using that best-fit model to estimate several demographic parameters:
effective population sizes, split time, and migration (gene flow). We ran 7 different models, 6 basic ones and a derivative: (1) neutral (no divergence, or still strongly mixing), (2) split-migration, (3) split-no-migration, (4) isolation with migration and population growth, (5) isolation with population growth and no migration, (6) isolation and secondary contact, and (7) a custom split-bidirectional-migration model (a simple derivative of split-migration; https://doi.org/10.6084/m9.figshare.6179054.v3). Models 1, 2, and 4 are provided in the δaδi file Demographics2D.py. The split-no-migration and isolation-with-population-growth and no-migration models use models (2) and (4) with migration parameters set to zero. The secondary contact model is that of Rougemont et al. (2017), and the split-bidirectional-migration model (figshare link above) adds bidirectional migration to the second model (split-migration) to account for potential asymmetry in gene flow.

We began δaδi analyses using a series of optimization runs for each basic model. In these runs, we adjusted parameters (grid points, upper and lower bounds) until repeated runs yielded the highest log composite likelihood values (within each basic model). Once we optimized these parameters within each model type, we performed additional runs within each model using the optimized parameters. We ran each model repeatedly with optimized parameters perturbed (as recommended in the δaδi user manual) until we observed the best likelihood value for that model 3 times. That is the value we report, except for poorer models, when a good fit could not be achieved and results always varied, in which case we averaged and report the highest 5 values. After identifying the best-fit model based on likelihood values over successive runs and confirming it using the Akaike information criterion (AICc, Akaike 1974, Burnham and Anderson 2002), we ran this model 10 times each with 66 jackknifed datasets to estimate the 95% confidence interval (CI) for each parameter.

Interpreting δaδi parameter estimates in biological terms requires estimates of the substitution rate of our loci and of the generation time of the owls. To obtain a value for the average per-site substitution rate within our UCE loci, we BLASTed our owl reference FASTA file against the genomes of 3 of the closest available relatives to obtain an average substitution rate (reasoning that multiple estimates are better than one). These genomes included Carmine Bee-Eater (Merops nubicus; NCBI annotation release 100), Rhinoceros Hornbill (Buceros rhinoceros; assembly ASM71030v1), and Barn Owl (Tyto alba; NCBI annotation release 100). We used time to most recent common ancestor (TMRCA) date estimates of 55.719 Ma (Strix-Tyto) and 63.482 Ma (Strix-Buceros/Merops) to obtain 3 rate estimates (Claramunt and Cracraft 2015; we used Strix in their tree as equivalent to Aegolius). Claramunt and Cracraft (2015) used clocklike DNA sequence and fossil calibrations to derive a new time tree for birds. We imported BLAST results into a spreadsheet, removed duplicate, lower-affinity hits, and we calculated base-pairs, mutations, and substitutions per site. This value of substitutions per site was converted to an annual rate by multiplying it by 2 TMRCA. The resulting estimates of substitutions per site per year were Merops $1.84 \times 10^{-10}$, Buceros $5.14 \times 10^{-10}$, and Tyto $4.23 \times 10^{-10}$. We used the average rate ($3.73 \times 10^{-10}$) to convert parameter estimates from δaδi analyses into biologically relevant estimates of effective population sizes and split times. Variations in this rate do not affect gene flow estimates but do affect other estimates (Appendix Table 3). We converted mutation rates to substitutions/site/generation using a generation time of 3 yr for Northern Saw-whet Owls following Withrow et al. (2014).

RESULTS

Assembly of the 4 specimens used to create a reference yielded 230,616 contigs (min length = 224 base pairs [bp], max = 27,918 bp) with a mean length of 377.3 bp ($\pm 0.65$ bp 95% CI), for a total of 87.0 million bp. Of these contigs, 4,357 were >1 Kb. Following the identification of UCE loci and the removal of paralogs, 4,300 UCE loci remained.

After we brought the full dataset through the bioinformatics pipeline, applying quality-control filtering, calling SNPs, phasing loci (reconstructed haplotypes), and applying genotype-quality filtering, 2,517 loci remained with quality data for all individuals. These loci comprised 2.7 million bp with mean length 1,068 bp ($\pm 7.73$ bp 95% CI). This complete data matrix contained 2,210 variable loci and 307 invariant loci, with a total of 5,616 SNPs (averaging 2.54 SNPs per locus). Per-site sequencing depth for these SNPs was 28.4 ($\pm 16.4$ SD). Of the 2,210 variable loci, 1,282 were variable among A. a. acadicus and brooksi individuals, and 928 more loci were variable with inclusion of the outgroup A. funereus. Of the 1,282 variable ingroup loci, 145 loci were Z-linked; these were only excluded from the δaδi analyses.

Nucleotide diversity (π) was 0.00014 overall (including funereus), 0.00025 for acadicus, and 0.00017 for brooksi. A. a. brooksi had fewer alleles (2,609) than A. a. acadicus (3,162), which is concordant with the smaller population size of brooksi. Only 44 SNPs were not in Hardy-Weinberg equilibrium. Bartlett’s test (Jombart and Ahmed 2011) rejected homogeneity of variance between observed heterozygosity ($H_o = 0.089, 0.124$) and expected heterozygosity ($H_e = 0.076, 0.114$), but $H_o$ did not differ from $H_e$ ($t = –0.449, df = 2089, P = 0.67$).

No alleles had a fixed difference ($F_{ST} = 1.0$) between the 2 taxa, and few alleles showed strong segregation. Six loci had $F_{ST}$ values >0.70 (0.71–0.84); none were on the Z chromosome (Appendix). Overall, the 2 populations were genetically different ($F_{ST} = 0.093, P = 0.0003$).
The best-fit model under $\delta a \delta i$ was split-bidirectional-migration, with a maximum composite likelihood score averaging $-256.2$. The other models had successively lower scores: isolation and secondary contact ($-268.7$), split-with-migration ($-269.8$), isolation with migration and population growth ($-469.0$), neutral ($-729.3$), and isolation with migration and no migration ($-1149.8$). Split-bidirectional-migration was confirmed as the best-fit model using AIC$_{c}$ ($\Delta$ AIC$_{c} > 25$; other model likelihoods were all $< 3.4 \times 10^{-6}$). We were unable to find a stable configuration of the split with no migration model and could not get it to run to completion. Parameter estimates for the split-bidirectional-migration model and their CIs are given in Table 1. A key result with respect to our question of gene flow was that gene flow into brooksi is low ($\sim 0.74$ individuals per generation), whereas that from brooksi into nominate acadicus is higher ($\sim 4.4$ individuals per generation; Table 1). The effective population size of nominate acadicus is $\sim 179$K, whereas that of brooksi is $\sim 6$K (Table 1). Finally, our data suggest that the 2 populations split $\sim 278$ Kya (Table 1).

**DISCUSSION**

The Haida Gwaii owl *A. acadicus brooksi* is as distinctive genetically as it is phenotypically (100% diagnosable), and despite opportunity for gene flow from nominate *acadicus*, we found the levels of gene flow to be relatively low (Table 1). Our results show skewed levels of gene flow in exactly the opposite direction that one would predict given specimen evidence. From specimen ratios of subspecies represented in Haida Gwaii vs. other populations, we have a Haida Gwaii presence of 7: $\sim 115$ *acadicus: brooksi*, whereas elsewhere we have no *brooksi* and large numbers of *acadicus* (Sealy 1998, 2013, Withrow et al. 2014). This striking mismatch in directionality suggests that the genetic data (1) reflect historic conditions that are no longer present, (2) that current mechanisms (e.g., endogenous timing or direction of migration, dietary specialization) or selection prevent effective gene flow, or, most likely, (3) both. Dispersal from Haida Gwaii and gene flow are evident in both phenotype and genotype in at least one other endemic Haida Gwaii subspecies (Pine Grosbeak [*Pinicola enucleator carlottae*], Topp and Winker 2008).

The indication in our data of nuclear gene flow from brooksi into acadicus is not reflected in mtDNA (Withrow et al. 2014). The nuclear signal might arise from 2 possible scenarios: (1) introgression from postglacial expansion of acadicus into a former range of brooksi that was broader than its current range (e.g., Carrara et al. 2007), as has been found in Hermit and Townsend’s warblers (Krosby and Rohwer 2010) and Snow and McKay’s buntings (Maley and Winker 2010); or (2) dispersal of brooksi from Haida Gwaii to the range of acadicus. As noted, the latter has not been detected from either specimens or mtDNA (Sealy 1998, 2013, Withrow et al. 2014). If this species had male-biased dispersal, the latter pattern might develop (nuDNA vs. mtDNA mismatch in gene flow; e.g., Peters et al. 2014), but in most birds, including owls, female-biased dispersal is the norm (Konig et al. 2009, Lovette and Fitzpatrick 2016). There are no good data on dispersal in this species (Rasmussen et al. 2008), but there is some indication that females move more than males (Beckett and Proudfoot 2012, De Ruyck et al. 2012), and its congener *A. funerus* is known to have female-biased dispersal (Marks and Doremus 2000). At present, then, dispersal from Haida Gwaii seems very low or nonexistent from mtDNA and phenotypic evidence, so genomic evidence of gene flow might reflect historic events.

There is also a mismatch between mtDNA and nuDNA in allele fixation between the 2 taxa. This is likely because mtDNA has an effective population size a quarter that of nuclear alleles and will sort more rapidly due to the effects of genetic drift (Moore 1995). It is also worth noting that even although we assayed 2.7 million bp of DNA per individual, this only represents $\sim 0.25$% of the genome (assuming genome size is similar to the chicken, 1.05 billion bp; Hillier et al. 2004), and our data probably do not include portions of the genome under strong divergent selection or drift. But there is also a difference with respect to gene flow. Our nuclear genomic demographic estimates (Table 1) differ from earlier estimates using mtDNA (Withrow et al. 2014) in showing somewhat higher levels of gene flow ($\sim 0.74$ acadicus $\rightarrow$ brooksi and $\sim 4.4$ brooksi $\rightarrow$ acadicus individuals per generation here, vs. $\sim 0.0003$ and 0.136 using mtDNA) and a deeper divergence date ($\sim 297$ Kya vs.
TABLE 1. Demographic model parameters from the δαδι split-bidirectional-migration model and estimates in biological units, with 95% CIs determined by jackknifed datasets.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimated (± 95% CI)</th>
<th>Lower-upper bounds</th>
<th>Biological units</th>
</tr>
</thead>
<tbody>
<tr>
<td>nu1 (population size acadicus)</td>
<td>11.49 (± 3.22)</td>
<td>179,090 (± 50,275)</td>
<td>128,814–229,365</td>
</tr>
<tr>
<td>nu2 (population size brooksi)</td>
<td>0.39 (± 0.68)</td>
<td>6010 (± 10,527)</td>
<td>0–16,537</td>
</tr>
<tr>
<td>T (split time)</td>
<td>2.98 (± 0.69)</td>
<td>278,177 (± 64,252)</td>
<td>213,925–342,429</td>
</tr>
<tr>
<td>m12 (migration)</td>
<td>0.76 (± 0.41)</td>
<td>4.36 (± 2.37)</td>
<td>1.99–6.72</td>
</tr>
<tr>
<td>m21 (migration)</td>
<td>3.85 (± 2.44)</td>
<td>0.74 (± 0.47)</td>
<td>0.27–1.21</td>
</tr>
<tr>
<td>Θ</td>
<td>76.26 (± 14.04)</td>
<td>15,582 (± 2869)</td>
<td>12,713–18,452 Ancestral population individuals</td>
</tr>
</tbody>
</table>

* Nref (δαδι variable for reference population size; Θ = 4Nμ).

~16 Kya, respectively). Effective population size estimates were not as dissimilar, being of the same order of magnitude (although those from mtDNA represent females only), but the effective population size estimate from UCEs for brooksi is larger than current census-size estimates of ~1,900 individuals (COSEWIC 2006), perhaps reflecting a larger historical refugial population (Table 1).

Under a phylogenetic species concept, the Haida Gwaii owl brooksi is a species, given its fixed differences in plumage and mtDNA (Withrow et al. 2014). Under the biological species concept, which allows some degree of gene flow (Johnson et al. 1999, Winker et al. 2007, Price 2008), the issue is less clear cut. Two key questions arise: Is mating assortative (i.e. do we see nonrandom pairing of individuals?), and what levels of gene flow can be sustained while retaining evolutionary independence? The process of speciation requires very low levels of gene flow if it is to go to completion (Mayr 1963, Coyne and Orr 2004, Price 2008). If we consider that ~6% of Haida Gwaii specimens are A. a. acadicus individuals that might remain and breed, then the low levels of gene flow that we found indicate that nonrandom pairing (assortative mating) is occurring. Why these A. a. acadicus individuals do not stay and breed in a place that is clearly suitable for reproduction for the species is highly relevant to understanding divergence in this lineage. In this species, in particular, it seems surprising that they do not remain to breed more often. There is a migratory population of A. a. acadicus breeding in similar habitat in the Alexander Archipelago starting just ~50 km north of Haida Gwaii (Figure 2). Further, Marks and Doremus (2000: 299) suggested “that Northern Saw-Whet Owls are nomadic in some parts of their range, settling in to breed in areas of high food abundance that they encounter during the non-breeding season.”

From the observation that nominate acadicus individuals are not staying and reproducing at the frequency with which they might do so, we infer that this form of assortative mating is likely due to divergent selection operating on populations focused on resources heterogeneously distributed in time and space, as outlined in heteropatric speciation theory as a type of ecological speciation (including allochrony as a component; Winker 2010, Taylor and Friesen 2017). Possibly relevant is that A. a. acadicus individuals feed predominantly on small mammals, whereas the brooksi diet is more flexible and includes up to 50% intertidal invertebrates in winter (Hobson and Sealy 1991, Sealy 1999, Rasmussen et al. 2008). Selection pressures resulting from allopatric and allochronic breeding distributions (e.g., associated ecological factors such as food availability) might also be coupled with wintering factors such as competitive exclusion.

The second key question regarding biological species status focuses on the extent of gene flow and its effects. Levels of gene flow into brooksi from acadicus are estimated to be low in nuDNA and very low in mtDNA. Under neutral conditions, levels of gene flow below 1.0 individuals per generation result in populations continuing to diverge (Wright 1943, Cabe and Alstad 1994). The presence of divergent selection can accommodate somewhat higher levels of gene flow than this and still enable divergence to proceed (Rice and Hostert 1993, Hostert 1997; but see Postma and van Noordwijk 2005). This taxon pair seems to have low enough levels of gene flow that brooksi is effectively evolutionarily independent.

Reconstructing the exact model of speciation involved in the divergence between these owls is difficult, because we lack the ability to reliably recover the historic distributions and ecological contexts that preceded current environments in this glaciated region (contra Winker et al. 2013). Our results suggest that the divergence of these taxa did not rely on the long periods of isolation associated with classic allopatric processes. The speciation-with-gene-flow models likely to be most appropriate in this case include parapatric, heteropatric, and ecological speciation models, which are complementary in the ways they include both geographic and ecological factors contributing to divergence and its maintenance. Ecological speciation is the process of divergence in which barriers to gene flow evolve due to divergent selection; differences in behavior, ecology,
The environment and the distribution of Haida Gwaii, and the environment are common drivers of this process (Schluter 1996, 2001, McKinney et al. 2004, Rundle and Nosil 2005, Nosil 2012, Ruegg et al. 2012, Verzijden et al. 2012). Among migrants, these geographic and ecological aspects can be tightly coupled: in addition to overlapping and non-overlapping distributions, diverging migratory lineages are also often affected by ecological differences and/or exhibit behavioral differences, including differences in the timing of resource availability and/or degrees of partial migration or sedentariness.

We consider that in these owl lineages distribution and ecology together have likely played important roles. Strict isolation from a migratory lineage can be difficult to achieve. In this case, A. a. acadicus is a facultative migrant noted for its high dispersal rates (i.e. low philopatry and breeding site fidelity; Rasmussen et al. 2008, Marks et al. 2015). Individual acadicus also occur well outside their normal range, including Kodiak, St. Paul, and St. Lawrence islands in Alaska; Newfoundland, Canada; and Bermuda (Rasmussen et al. 2008, and UAM specimens). However, the enhanced isolation of being in a glacial refugium was probably important in providing an added degree of allopatry in this system, as opposed, for example, to divergences occurring in other migratory lineages that developed patterns of leapfrog migration (Winker 2010, Winker et al. 2013). It is noteworthy, though, that among other Haida Gwaii avian populations with evidence of refugial occupation (e.g., P. enucleator, Troglyodytes pacifica, and Melospiza melodia; Pruett et al. 2013), this enhanced isolation apparently did not prevent post-glacially expanding mainland forms from being able to introgress with Haida Gwaii populations. This adds additional evidence to a role for ecological factors being involved in the owls’ divergence. The Haida Gwaii owl appears to be maintaining phenotypic and genetic distinctiveness despite low levels of gene flow, and we suggest that this is likely due to divergent selection operating on aspects such as sedentariness, plumage coloration, and diet (Sealy 1998, 1999). It appears that this is a case of speciation with gene flow, and the Haida Gwaii owl (A. a. brooksi) might be considered a young biological species.

ACKNOWLEDGMENTS

We thank Fred Sheldon and several anonymous reviewers for comments on an earlier draft. We also thank Matthew Miller, Mike Harvey, and Jessica McLaughlin for providing help in bioinformatics. J.M. also helped in getting the secondary contact model to run. Kevin Hawkins, Phil Lavretske, Jeff Peters, and Ryan Gutenkunst provided help in running δaδi.

Funding statement: The National Science Foundation (DEB-1242267-1242241-1242260) supported the genomics aspect of this study. Funding for S.G.S.’s research on the fauna of Haida Gwaii was received from the Natural Sciences and Engineering Research Council of Canada. We thank the National Center for Genome Analysis Support (Trinity on Galaxy for assembly) for computational support. None of these funders had any influence on the content of the submitted or published manuscript, and none require approval of the final manuscript to be published.

Ethics statement: Archived museum specimens were used in this research.

Author contributions: (1) K.W., T.C.G., B.C.F., and J.W. formulated the questions; (2) S.G.S. provided specimens and historical and current background on the biology of Saw-whet owls on Haida Gwaii; (3) T.C.G. and B.C.F. generated the data; (4) K.W., B.C.F., and T.C.G. performed bioinformatics and analyses; (5) all authors wrote and edited drafts of the paper.

Data accessibility: Raw sequence data have been deposited in the NCBI sequence read archive (SRA); reference UCE contigs are deposited in GenBank; and the analyzed VCF files, reference sequence, and unique scripts used are available on FigShare (https://doi.org/10.6084/m9.figshare.6179054.v3).

LITERATURE CITED


APPENDIX 1. BLASTN RESULTS AGAINST CHICKEN GENOME FOR THE 6 LOCI WITH $F_{st} > 0.70$

(Gallus_gallus-5.0 reference Annotation Release 103).

**uce-7727** (length 982)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 8, Gallus_gallus-5.0 Length=29963013

Features in this part of subject sequence:
nuclear factor 1 A-type

Score = 902 bits (488), Expect = 0.0
Identities = 593/644 (92%), Gaps = 6/644 (1%)
Strand=Plus/Plus

**uce-5087** (length 1001)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 19, Gallus_gallus-5.0 Length=9979828

Features in this part of subject sequence:
cut-like homeobox 1 protein CASP isoform X1

Score = 329 bits (178), Expect = 6e-88
Identities = 189/194 (97%), Gaps = 1/194 (1%)
Strand=Plus/Plus

**uce-5227** (length 1032)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 1, Gallus_gallus-5.0 Length=196202544

Features in this part of subject sequence:
84813 bp at 5’ side: forkhead box protein P2
169970 bp at 3’ side: fork head domain-containing protein FD5-like

Score = 628 bits (340), Expect = 6e-178
Identities = 648/793 (82%), Gaps = 36/793 (5%)
Strand=Plus/Plus

**uce-601** (length 1335)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 2, Gallus_gallus-5.0 Length=149560735

Features flanking this part of subject sequence:
31972 bp at 5’ side: teashirt homolog 1
18408 bp at 3’ side: zinc-binding alcohol dehydrogenase domain-containing prot...

Score = 1694 bits (917), Expect = 0.0
Identities = 1181/1307 (90%), Gaps = 23/1307 (2%)
Strand=Plus/Plus

**uce-5371** (length 971)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 3, Gallus_gallus-5.0 Length=111302122

Features in this part of subject sequence:
serine/threonine-protein kinase MRCK alpha isoform X11
serine/threonine-protein kinase MRCK alpha isoform X7

Score = 501 bits (271), Expect = 1e-139
Identities = 646/820 (79%), Gaps = 54/820 (7%)
Strand=Plus/Plus

**uce-4278** (length 768)
Gallus gallus isolate RJF #256 breed Red Jungle fowl, inbred line UCD001 chromosome 2, Gallus_gallus-5.0 Length=149560735

Features in this part of subject sequence:
zinc finger CCCH domain-containing protein 3
zinc finger CCCH domain-containing protein 3 isoform X1

Score = 547 bits (296), Expect = 1e-153
Identities = 613/760 (81%), Gaps = 45/760 (6%)
Strand=Plus/Plus
**APPENDIX TABLE 2.** Specimens used in this study, associated data, and Sequence Read Archive (SRA) data deposits.

<table>
<thead>
<tr>
<th>UAM</th>
<th>Genus</th>
<th>Species</th>
<th>Subspecies</th>
<th>Age</th>
<th>Sex</th>
<th>Date</th>
<th>Country</th>
<th>State/province</th>
<th>County</th>
<th>SPECLOC</th>
<th>COLLCATNUM</th>
<th>SRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>8,990</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>U</td>
<td>M</td>
<td>Dec 13, 1997</td>
<td>USA</td>
<td>Alaska</td>
<td>Kenai Peninsula</td>
<td>Homer</td>
<td>UAMX538</td>
<td>SAMN09943900</td>
</tr>
<tr>
<td>9,180</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>U</td>
<td>M</td>
<td>Jan 22, 1997</td>
<td>USA</td>
<td>Alaska</td>
<td>Kenai Peninsula</td>
<td>Seward</td>
<td>KSW3084</td>
<td>SAMN09943901</td>
</tr>
<tr>
<td>13,949</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>U</td>
<td>M</td>
<td>Nov, 1999</td>
<td>USA</td>
<td>Alaska</td>
<td>Alexander Archipelago</td>
<td>Revillagigedo Island</td>
<td>UAMX2214</td>
<td>SAMN09943902</td>
</tr>
<tr>
<td>13,996</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>U</td>
<td>F</td>
<td>Dec. 1998</td>
<td>USA</td>
<td>Alaska</td>
<td>Upper Cook Inlet</td>
<td>Palmer</td>
<td>UAMX2119</td>
<td>SAMN09943903</td>
</tr>
<tr>
<td>17,882</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>AHY</td>
<td>U</td>
<td>May 12, 1999</td>
<td>USA</td>
<td>Alaska</td>
<td>Anchorage, 11 miles south</td>
<td>UAMX2953</td>
<td>SAMN09943904</td>
<td></td>
</tr>
<tr>
<td>17,953</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>AD</td>
<td>F</td>
<td>UNK, 199X</td>
<td>USA</td>
<td>Maine</td>
<td>Upper Cook Inlet</td>
<td>Palmer, Dry Lakes Graham Island</td>
<td>UAMX2975</td>
<td>SAMN09943905</td>
</tr>
<tr>
<td>17,957</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>acadicus</td>
<td>AD</td>
<td>M</td>
<td>Jul, 2002</td>
<td>USA</td>
<td>Alaska</td>
<td>Upper Cook Inlet</td>
<td></td>
<td>UAMX2980</td>
<td>SAMN09943906</td>
</tr>
<tr>
<td>10,153</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>JUV</td>
<td>F</td>
<td>UNK, 1999</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td></td>
<td>KSW3338</td>
<td>SAMN09943907</td>
</tr>
<tr>
<td>19,042</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>U</td>
<td>M</td>
<td>Nov 04, 1994</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td></td>
<td>KSW3087</td>
<td>SAMN09943908</td>
</tr>
<tr>
<td>19,472</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>U</td>
<td>F</td>
<td>Nov 09, 2005</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td></td>
<td>KSW5212</td>
<td>SAMN09943909</td>
</tr>
<tr>
<td>19,474</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>U</td>
<td>F</td>
<td>Nov 01, 1994</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td></td>
<td>KSW3089</td>
<td>SAMN09943910</td>
</tr>
<tr>
<td>19,485</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>JUV</td>
<td>M</td>
<td>Aug, 1997</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td>Skidegate</td>
<td>KSW3096</td>
<td>SAMN09943911</td>
</tr>
<tr>
<td>26,388</td>
<td>Aegolius</td>
<td>acadicus</td>
<td>brooksi</td>
<td>U</td>
<td>M</td>
<td>UNK, 2007</td>
<td>Canada</td>
<td>British Columbia</td>
<td>Queen Charlotte Islands</td>
<td></td>
<td>KSW5210</td>
<td>SAMN09943912</td>
</tr>
<tr>
<td>7,626</td>
<td>Aegolius</td>
<td>funereus</td>
<td></td>
<td>U</td>
<td>M</td>
<td>Jun, 1991</td>
<td>USA</td>
<td>Alaska</td>
<td>Kodiak Archipelago</td>
<td>Kodiak Island</td>
<td>RWD24036</td>
<td>SAMN09943913</td>
</tr>
<tr>
<td>15,084</td>
<td>Aegolius</td>
<td>funereus</td>
<td></td>
<td>AHY</td>
<td>M</td>
<td>Feb 22, 2001</td>
<td>USA</td>
<td>Alaska</td>
<td>Interior</td>
<td>Fairbanks</td>
<td>UAMX2638</td>
<td>SAMN09943914</td>
</tr>
</tbody>
</table>

* Specific locality; b Collector’s catalogue number; c Day and month unknown.

**APPENDIX TABLE 3.** Variation in parameter estimates under different substitution rates (gene flow estimates are not affected).

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Average (Table 1)</th>
<th>Buceros</th>
<th>Tyto</th>
<th>Merops</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Theta )</td>
<td>15,582.25</td>
<td>31,659.23</td>
<td>13,771.51</td>
<td>11,321.61</td>
<td>Ancestral population</td>
</tr>
<tr>
<td>Split time</td>
<td>278,177</td>
<td>565,186</td>
<td>245,851</td>
<td>202,115</td>
<td>Yr</td>
</tr>
<tr>
<td>Population sizes</td>
<td>179,090</td>
<td>363,865</td>
<td>158,278</td>
<td>130,121</td>
<td>acadicus individuals</td>
</tr>
<tr>
<td></td>
<td>6,010</td>
<td>12,211</td>
<td>5,311</td>
<td>4,367</td>
<td>brooksi individuals</td>
</tr>
</tbody>
</table>