

Journal Pre-proofs

Population genomics indicate three different modes of divergence and speciation with gene flow in the green-winged teal duck complex

Fern Spaulding, Jessica F. McLaughlin, Rebecca G. Cheek, Kevin G. McCracken, Travis C. Glenn, Kevin Winker

PII: S1055-7903(23)00033-7
DOI: <https://doi.org/10.1016/j.ympev.2023.107733>
Reference: YMPEV 107733

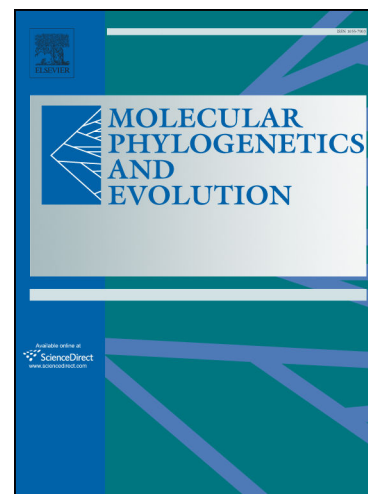
To appear in: *Molecular Phylogenetics and Evolution*

Received Date: 19 August 2021
Revised Date: 31 January 2023
Accepted Date: 9 February 2023

Please cite this article as: Spaulding, F., McLaughlin, J.F., Cheek, R.G., McCracken, K.G., Glenn, T.C., Winker, K., Population genomics indicate three different modes of divergence and speciation with gene flow in the green-winged teal duck complex, *Molecular Phylogenetics and Evolution* (2023), doi: <https://doi.org/10.1016/j.ympev.2023.107733>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Published by Elsevier Inc.



Population genomics indicate three different modes of divergence and speciation with gene flow in the green-winged teal duck complex

Fern Spaulding ^{a, b}, Jessica F. McLaughlin ^c, Rebecca G. Cheek ^d, Kevin G. McCracken ^{a, e},
Travis C. Glenn ^f, Kevin Winker ^{a, b}

^a University of Alaska Museum, University of Alaska Fairbanks, Fairbanks, AK, USA

^b Department of Biology and Wildlife, University of Alaska Fairbanks, Fairbanks, AK, USA

^c Department of Environmental Science, Policy, and Management, University of California Berkeley, Berkeley, CA, USA

^d Graduate Degree Program in Ecology, Department of Biology, Colorado State University, Fort Collins, CO, USA

^e Department of Biology, University of Miami, Coral Gables, FL, USA

^f Department of Environmental Health Science, University of Georgia, Athens, GA, USA

Corresponding author: Fern Spaulding; frspaulding@alaska.edu

Abstract

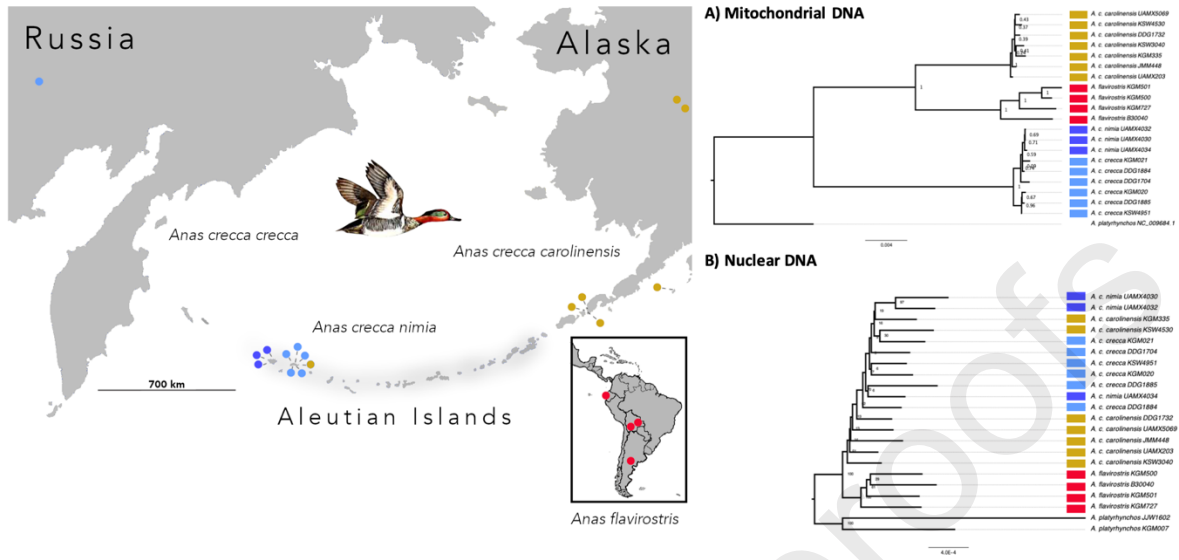
The processes leading to divergence and speciation can differ broadly among taxa with different life histories. We examine these processes in a small clade of ducks with historically uncertain relationships and species limits. The green-winged teal (*Anas crecca*) complex is a Holarctic species of dabbling duck currently categorized as three subspecies (*Anas crecca crecca*, *A. c. nimia*, and *A. c. carolinensis*) with a close relative, the yellow-billed teal (*Anas flavirostris*) from South America. *A. c. crecca* and *A. c. carolinensis* are seasonal migrants, while the other taxa are

sedentary. We examined divergence and speciation patterns in this group, determining their phylogenetic relationships and the presence and levels of gene flow among lineages using both mitochondrial and genome-wide nuclear DNA obtained from 1,393 ultraconserved element (UCE) loci. Phylogenetic relationships using nuclear DNA among these taxa showed *A. c. crecca*, *A. c. nimia*, and *A. c. carolinensis* clustering together to form one polytomous clade, with *A. flavirostris* sister to this clade. This relationship can be summarized as (*crecca*, *nimia*, *carolinensis*)(*flavirostris*). However, whole mitogenomes revealed a different phylogeny: (*crecca*, *nimia*)(*carolinensis*, *flavirostris*). The best demographic model for key pairwise comparisons supported divergence with gene flow as the probable speciation mechanism in all three contrasts (*crecca*—*nimia*, *crecca*—*carolinensis*, and *carolinensis*—*flavirostris*). Given prior work, gene flow was expected among the Holarctic taxa, but gene flow between North American *carolinensis* and South American *flavirostris* ($M \sim 0.1 - 0.4$ individuals/generation), albeit low, was not expected. Three geographically oriented modes of divergence are likely involved in the diversification of this complex: heteropatric (*crecca*—*nimia*), parapatric (*crecca*—*carolinensis*), and (mostly) allopatric (*carolinensis*—*flavirostris*). Our study shows that ultraconserved elements are a powerful tool for simultaneously studying systematics and population genomics in systems with historically uncertain relationships and species limits.

Keywords

Anatidae, waterfowl phylogenetics, Beringia, speciation with gene flow, parapatric speciation, allopatric speciation

Graphical Abstract



1. Introduction

A fundamental goal of evolutionary biology is to understand the different modes of divergence and speciation that generate biodiversity. Speciation is the process of evolutionary divergence that often results in species, which are distinct groups of organisms that are essentially reproductively isolated from each other (Mayr, 1942; Price, 2008). This process has historically been categorized in three main modes: sympatric, parapatric, and allopatric, each of which is based on spatial relationships between the diverging populations (Fitzpatrick et al., 2008). Classic allopatric speciation is the divergence of populations that are geographically separated from each other with no movement of individuals between them (Mayr, 1942; Gavrilets, 2003, 2004; Coyne & Orr, 2004; Fitzpatrick et al., 2008; Price, 2008). Sympatric speciation involves the evolution of reproductive isolation while the ranges of populations overlap (Coyne & Orr, 2004). Parapatric speciation is the origin of new species in which gene flow occurs across a spatially restricted contact zone such that only a fraction of each population has a high probability of emigrating or of interacting with immigrants (Smith, 1955; Endler, 1977; Futuyma & Mayer, 1980, Fitzpatrick et al., 2008). These geographic modes of speciation can be considered to exist on a continuum, on which allopatric and sympatric speciation represent the endpoints of different amounts of gene flow (zero to maximum), whereas parapatric speciation occupies the space representing intermediate gene flow between these extremes (Butlin et al., 2008; Gavrilets, 2014). In speciation theory, if the connectivity between two populations is not broken, and gene flow persists, then parapatric speciation models apply (speciation with gene flow, in a non-sympatric distribution; Gavrilets, 2004). The frequency of parapatric speciation in nature is uncertain (Coyne & Orr, 2004; Price, 2008), and it has been a relatively neglected area in speciation research (Gavrilets, 2004).

The study of speciation has often focused on these geographic modes of divergence. There are major debates about whether allopatric speciation is predominant (e.g. Coyne & Orr, 2004), whether sympatric speciation is possible, and if so how frequently it might occur (e.g. Berlocher & Feder, 2002; Bolnick & Fitzpatrick, 2007). In some cases, geographic isolation alone has been shown to drive divergence when isolated populations reside in allopatry. For example, allopatric speciation is common in molluscs (78% of species pairs), driven in part by their low dispersal rates (Hernández-Hernández et al., 2021). Similarly, among terrestrial snails that live on island archipelagos, allopatric species pairs predominate (96%), with most species being endemic to a single island (Holland & Cowie, 2009; Rundell, 2008; Jordaens et al., 2009; Hernández-Hernández et al., 2021).

Though the geographic distribution of populations is important for understanding speciation, other factors such as ecological, environmental, and behavioral differences between populations are increasingly seen as important drivers of population divergence and speciation (Schluter, 2001; McKinnon et al., 2004; Rugg et al., 2012; Verzijden et al., 2012, Withrow et al., 2014). While these factors can operate within each category of the geographically oriented modes of speciation (allopatric, parapatric, and sympatric; Gavrillets, 2003), their presence is particularly important for enabling parapatric or sympatric speciation to progress (speciation with gene flow; Rundle & Nosil, 2005; Nosil, 2008; Feder et al., 2012). For example, ecological and sexual selection are considered major evolutionary forces that often drive insect speciation (Arnqvist et al., 2000; Forbes et al., 2017; Hernández-Hernández et al., 2021), and spatial relationships can become less important. An example of ecological divergence occurs in *Timema* walking stick insects where ecotypes live on different host plants (Nosil et al., 2008), in which greater reproductive isolation evolves between populations adapting to contrasting environments

than between populations adapting to similar environments (Rice & Hostert, 1993; Schluter & Nagel, 1995; Schluter, 2009). Environmental factors are also often an important driver of divergence (Hernández-Hernández et al., 2021), and this occurs in many vertebrate groups, including salamanders (Kozak & Wiens, 2010), frogs (Moen & Wiens, 2017), birds (Cooney et al., 2016), and mammals (Castro-Insua et al., 2018). Finally, behavior is commonly involved in prezygotic isolating barriers, including ecological and behavioral differences between species (Hernández-Hernández et al., 2021). For example, species-specific vocalization and communication are often important reproductive isolating mechanisms in birds and frogs (e.g., Edwards et al., 2005; Hoskin et al., 2005; Boul et al., 2007; Uy et al., 2018).

In birds, allopatric speciation has historically been thought to be the main route to speciation (Mayr, 1963; Coyne & Orr, 2004; Price, 2008). However, genomic data increasingly identify groups that do not fit this model and instead indicate that speciation has progressed with at least some gene flow (Mallet et al., 2016; Morales et al., 2017; Peñalba et al., 2019; Rheindt & Edwards, 2011; Zarza et al., 2016, McLaughlin et al., 2020; Winker, 2021). One example of how the allopatric speciation model is not an ideal fit for speciation in birds is that many long-distance seasonally migratory birds often exhibit semiannual transcontinental and transoceanic movements that can prevent diverging populations from undergoing long periods of strict allopatry (Winker, 2010; Peters et al., 2012), increasing the likelihood of divergence with gene flow.

Holarctic avian taxa breeding across the Northern Hemisphere provide an excellent opportunity to study possible speciation with gene flow, given that many of these birds are long-distance migrants that seasonally migrate to the Northern Hemisphere to breed. Here we focus on divergence in an avian taxonomic complex in Beringia, a geographic region that extends from

the Russian Far East across Alaska and into western Canada that has experienced dynamic fluctuations in climate throughout the Pleistocene. While much of the Northern Hemisphere was covered by ice sheets during past glaciations, the lowlands of Beringia remained free of ice, providing a refuge for high-latitude flora and fauna (Elias & Brigham-Grette, 2013). These cyclic fluctuations in climate have had genetic consequences on Beringian taxa (Hewitt, 1996), including creating diverse patterns of differentiation in Holarctic birds (e.g. Zink et al., 1995; Drovetski et al., 2004; Buehler & Baker, 2005; Humphries & Winker, 2011; Peters et al., 2014). Previous work on speciation and divergence in birds across Beringia has shown that speciation with gene flow is common (Winker et al., 2013, 2018; Peters et al., 2014; McLaughlin et al., 2020). In the green-winged teal (*Anas crecca* and its subspecies), cyclic climatic fluctuations seem to have caused incomplete parapatric speciation (divergence with gene flow) between Asian and North American populations (*A. c. crecca* and *carolinensis*; Peters et al., 2012). A sedentary population in the Aleutian Islands (*A. c. nimia*) is also undergoing divergence with gene flow from the migratory Asian mainland subspecies (*A. c. crecca*), which performs seasonal migrations through the range of *nimia* in a heteropatric (seasonally sympatric) geographic relationship (Winker et al., 2013). Finally, the closest relative of these lineages is the South American yellow-billed teal (*A. flavirostris*), whose exact relationship with *A. crecca* (*sensu lato*) remains unresolved (Peters et al., 2012).

In this study, we examined the green-winged teal complex comprising three subspecies of green-winged teal in the Northern Hemisphere (the migratory Eurasian common teal (*Anas crecca crecca*) and North American green-winged teal (*A. c. carolinensis*), and the sedentary Aleutian green-winged teal (*A. c. nimia*)), and a closely related sister species (or species complex; more below), the yellow-billed teal (*A. flavirostris*) in the Southern Hemisphere (Fig.

1). This latter South American species is known to be closely related to *A. crecca*, but it has previously demonstrated conflicting relationships between mitochondrial and nuclear phylogenies, and it likely represents a classic case of allopatric speciation (Peters et al., 2012). Taxonomically, the current species tree is (*A. c. crecca*, *A. c. nimia*, *A. c. carolinensis*)(*A. flavirostris*). However, the phylogeny based on mitochondrial DNA is (*A. c. crecca*, *A. c. nimia*)(*A. c. carolinensis*, *A. flavirostris*) (Johnson & Sorenson, 1999; Gonzalez et al., 2009). Here we bring a genomic-scale dataset to this system, asking first what are the phylogenetic relationships in the group? Second, how much gene flow has occurred? And, finally, what modes of divergence and speciation are prevalent in this group?

2. Materials and methods

2.1. Study design

We used ultraconserved elements (UCEs) from the nuclear genome as our primary genetic markers. UCEs allow us to examine thousands of orthologous loci, providing insight into the divergence and speciation processes from populations to deeper relationships (Faircloth et al., 2012; Everson et al., 2019). In addition to obtaining thousands of nuclear loci, a benefit of producing UCE data is that high-quality complete mitogenomes are also generated for each individual. Pairing UCEs with the complete mitogenomes for each individual allowed us to compare the phylogenetic relationships between nuclear and mitogenomic data. Given that there is often a lack of concordance between mitochondrial and nuclear DNA (mtDNA and nuDNA) estimates of divergence (Humphries & Winker, 2011; Peters et al., 2014), using both types of molecular markers can improve our understanding of lineage relationships and species limits (Rubinoff & Holland, 2005; Edwards & Bensch, 2009; Humphries & Winker, 2011). Using both

mtDNA and nuDNA simultaneously can also resolve potential discrepancies between previously published phylogenies.

2.2. Sampling, DNA extraction, and sequencing

We sampled the following numbers of individuals from each taxon: 6 *Anas c. crecca*, 7 *A. c. carolinensis*, 3 *A. c. nimia*, 4 *A. flavirostris* (Fig. 1), and 2 *A. platyrhynchos* were used as an outgroup. In this study, we consider each named taxon a population, so we are conducting comparisons at both the subspecies and species level. The South American *A. flavirostris* has historically been treated as one species, but it is more recently being considered as two (*A. flavirostris* and *A. andium*; e.g., del Hoyo & Collar, 2014; Remsen et al., 2022). Insofar as sample availability precluded our consideration of speciation in that group, our overarching study design treats *A. flavirostris sensu lato* as a single-lineage, superspecies group (though we do include some secondary analyses delving further into this group, where our samples included 3 *A. flavirostris oxyptera/flavirostris* and 1 *A. f. andium*).

Our study design and sampling required consideration of sample sizes. Theory considers an optimal sample size for coalescent-based genomic analyses to be 8 individuals (or haplotypes) per population when estimating population parameters (e.g., $\theta = 4N_e\mu$; Felsenstein, 2005). But it has been demonstrated that key demographic parameter estimates in lineages with divergences of these depths (i.e., subspecies and species) are generally resilient to lower sample sizes (McLaughlin & Winker, 2020). It has also been shown that, in general, relatively small sample sizes are sufficient when estimating interpopulation divergence and genetic diversity when using thousands of loci (Nazareno et al., 2017; McLaughlin & Winker, 2020). In our population-level analyses, each allele is called, which effectively doubles the sample size of each taxon (two haplotypes per individual). Thus, a minimum reliable sample size varies from taxon to taxon

(though is frequently much lower than 8), and is dependent on the demographic parameters of interest and the divergence levels between the taxa (McLaughlin & Winker, 2020). Here we are particularly interested in the possible presence and levels of gene flow (m), which when using our methods of analysis appear to be relatively consistent even when sample sizes are small (McLaughlin & Winker, 2020).

Each sample was obtained from high-quality, vouchered tissue samples from wild individuals (Table S1). DNA extractions followed the standard protocol for animal tissues using the QIAGEN DNeasy Blood + Tissue Extraction Kit (QIAGEN, 2006).

Our sequencing procedures followed Glenn et al. (2019). In short, we prepared dual-indexed DNA libraries which were quantified using a Qubit fluorimeter (Invitrogen, Inc., Carlsbad, CA, USA). We then enriched the samples with 5,060 UCE loci using the Tetrapods-UCE-5Kv1 kit from MYcroarray following version 1.5 of the UCE enrichment protocol and version 2.4 of the post-enrichment amplification protocol (<http://ultraconserved.org>). The resulting pool was then sequenced using a paired-end 150 bp (PE150) protocol on an Illumina HiSeq 2500 using three lanes (Illumina, Inc., San Diego, CA, USA; UCLA Neuroscience Genomics Core).

2.3. Bioinformatics and UCE pipeline

Our bioinformatics pipeline followed that of Winker et al. (2018). Briefly, raw and untrimmed FASTQ sequence data that contained low-quality bases were removed using Illumiprocessor (v.2.0.6; Faircloth, 2013), which incorporates Trimmomatic (v.0.32-1; Bolger et al., 2014). Our next steps used the package PHYLUCE (v.1.5.0; Faircloth, 2016), which identified conserved orthologous loci that were then used as our reference set of UCE loci to call variants in the focal, ingroup individuals. We used the mallard (*Anas platyrhynchos*; Table S1) to

build a UCE reference. We first combined sequence read files from two individuals into two read files (Table S1). We assembled these reads de novo using Trinity (v.2.4.0; Grabherr et al., 2011) on Galaxy (v.2.4; Afgan et al., 2016), then found and extracted UCE loci using PHYLUCE by matching the contigs to the probe set used. The resulting sequences were saved as a reference FASTA file. Next, for each ingroup individual, we combined singletons (reads that lost their pair) with read1 files. The PHYLUCE dependencies such as BWA-MEM (v.0.7.7; Li & Durbin, 2009; Li, 2013), SAMtools (v.0.1.19; Li et al., 2009), and PICARD (v.1.106; <http://broadinstitute.github.io/picard>) were used to index the reference sequence and align the unassembled raw reads of each individual against the mallard UCE reference using default parameters. We followed the population genomics pipeline developed by Harvey et al. (2016), which includes using Genome Analysis Toolkit (GATK, v.3.3.0; McKenna et al., 2010) to call and restrict data to high-quality SNPs (Q30).

VCFtools (v.0.1.13; Danecek et al., 2011) was used to cull sites without complete data to extract the high-quality SNPs and to create a complete matrix (all individuals represented at all loci) with a minimum genotype quality (Phred) score of 10.0 (which equates to 90% confidence). Our final FASTA file contained only high-quality, confidently sequenced loci, producing a 100% complete data matrix for analyses. The high-quality, complete-data VCF file was thinned to one SNP per locus using VCFtools and made biallelic by filtering out loci that had more than two alleles. The thinned, biallelic VCF file then was used in our demographic model analysis. We used BLASTn on NCBI to identify sex linked loci (Z-linked) using our high-quality FASTA data. We used the mallard genome (*Anas platyrhynchos*, IASCAAS_PekingDuck_PBH1.5) to identify hits for Z-linked loci, which were removed from our thinned biallelic VCF using the script `find_chrom.py` (v.1.2; https://github.com/jfmclaughlin92/beringia_scripts).

The biallelic VCF dataset was converted using PGDSpider (v.2.0.9.1; Lischer & Excoffier, 2012) into the appropriate format for analysis using the R package adegenet (v.2.0.1; Jombart & Ahmed, 2011). Adegnet was used to calculate Hardy-Weinberg equilibrium, observed (H_o) and expected heterozygosities (H_e), pairwise population level F_{ST} values, principal component analysis (PCA), and perform assignment tests using discriminant analysis of principal components (DAPC). A paired t -test was used to test for differences between expected and observed heterozygosity (Jombart & Ahmed, 2011). Although we tested demographic models of divergence and gene flow, Hardy-Weinberg equilibrium calculations offer another perspective on potential gene flow because unequal values for observed and expected heterozygosity suggest that one of the assumptions of Hardy-Weinberg equilibrium is not met (e.g., absence gene flow). To calculate nucleotide diversity (π), we created a concatenated FASTA file of all individuals using catfasta2phyml.pl (<https://github.com/nylander/catfasta2phyml>). The resulting file was then imported into MEGA (v.10; Kumar et al., 2018), and nucleotide diversity was calculated using the maximum composite likelihood method.

2.4. Mitogenomic pipeline

Mitogenomic sequence data were obtained as part of our UCE sequencing. The backbone of our mitogenomic analysis is similar to that of our UCE dataset and follows many of the same steps. Our pipeline again used PHYLUCE and followed the mitogenomic pipeline used by Everson et al. (2019). Briefly, we used the complete mitogenome of the mallard (*Anas platyrhynchos*, NC_009684.1) as a reference. The read1 and read2 files were mapped to the reference and indexed using BWA-MEM and SAMtools. This was followed by using PICARD to clean the alignments, add read group header information, and remove PCR and sequencing duplicates. Using the GATK module UnifiedGenotyper, SNPs were called for each individual

against the reference, and GATK was used to call and realign around indels, call and annotate SNPs, and filter SNPs around indels using the IndelRealigner module, which incorporates the merged interval output created by the RealignerTargetCreator module. We restricted the data to high-quality SNPs by adding a quality filter (Q30) before converting the resulting VCF file to a FASTA file using GATK. Mitochondrial divergence estimates were also calculated using whole mitogenomes to enable comparison with previous studies (Johnson & Sorenson, 1999; Humphries & Winker, 2011). Divergence estimates were calculated in MEGA using the between-group mean distance option.

2.5. Phylogenetic analysis

We used the PHYLUCE phylogenetic systematics pipeline (Faircloth, 2016) to generate a single UCE consensus sequence per locus/individual (with heterozygotes coded using IUPAC codes). We next used RAxML (using raxmlGUI v.1.5; Silvestro & Michalak, 2012) to reconstruct a maximum likelihood phylogeny. We used ModelTest-NG (v.0.1.7; Darriba et al., 2020) to determine that the GTR+GAMMA substitution model was best. This model was run with 100 bootstrap replicates. We used SNAPP (through BEAST v.2.5; Bouckaert et al., 2019) to reconstruct a phylogeny that integrates across all possible gene trees to illustrate the group's history using UCE data. We ran our SNAPP analysis for 30 million generations, sampling every 1000 steps, with a burn-in of 1 million generations. Tracer (v.1.7.2; Rambaut et al., 2018) was used to view the MCMC output to check for convergence and to ensure that no large-scale fluctuations were present in later trace trends. TreeSetAnalyzer (through SNAPP v.2.4.7; Bouckaert et al., 2019) was used to analyze the inferred gene trees to produce the 95% credible set of trees and to examine whether our mitochondrial topology was part of that set. SNAPP results were visualized using DensiTree (v.2.1.11; Bouckaert, 2010). For both our SNAPP and

RAXML trees we included the mallard (*Anas platyrhynchos*) as an outgroup, giving us a confident root. For the mitogenomes we used MEGA to reconstruct a phylogeny using maximum likelihood with 100 bootstrap replicates. The mallard was again used as an outgroup. FigTree (v.1.4.2; Rambaut, 2006) was used to visualize resulting tree files.

2.6. Demographic analysis

Diffusion Approximations for Demographic Inference ($\delta\text{a}\delta\text{i}$, v.1.7.0; Gutenkunst et al., 2009) was used to estimate demographic parameters under different models of pairwise divergence. Under a coalescent framework, $\delta\text{a}\delta\text{i}$ predicts the joint frequency spectrum of genetic variation among populations enabling statistically rigorous assessments of user-defined demographic models focused on population size, gene flow rates, and divergence times (Gutenkunst et al., 2009). We tested eight models of divergence (Fig. S1): A) no divergence (neutral, populations never diverge); B) split with no migration (divergence without gene flow); C) split with migration (divergence with gene flow that is bidirectionally symmetric, 1 migration parameter); D) split with bidirectional migration (divergence with gene flow that is bidirectionally asymmetric, 2 migration parameters); E) split with exponential population growth, no migration; F) split with exponential population growth and migration; G) secondary contact with migration (1 migration parameter); and H) secondary contact with bidirectional migration (2 migration parameters). The scripts for these models are available here: <https://doi.org/10.6084/m9.figshare.14327252.v1>. The neutral, split with migration, and exponential population growth models are provided in the $\delta\text{a}\delta\text{i}$ file Demographics2D.py (as snm, splitmig, and IM, respectively). The models split-with-no-migration and split-with-exponential-growth-no-migration are versions of the split-with-migration (splitmig) and exponential population growth (IM) models with the migration parameters set to zero. The split-with-

bidirectional-gene-flow model is a custom script that is a derivative of the split-with-migration (splitmig) model used to examine asymmetric gene flow. The secondary-contact model with one migration parameter (symmetric gene flow) is from Rougemont et al. (2017), and the secondary-contact model with two migration parameters is a derivative of that model to account for potential asymmetry in gene flow.

For each pairwise comparison, we ran a series of optimization runs, which consisted of running each model repeatedly under different parameters to find the most stable configuration with the lowest negative maximum log likelihood score (Table 2). Following this series of optimization runs, the best five log-likelihood scores from each set of subsequent runs were averaged to summarize that model, and we used the Akaike Information Criterion (AIC; Akaike, 1974; Burnham & Anderson, 2004) to determine the best-fit model. We then ran the best-fit model repeatedly and used demographic parameter estimates from this model's top three runs to interpret values (as appropriate for each pairwise model) for ancestral population size (N_{ref} , derived from θ), migration rates (m or $m12$, $m21$), current effective population sizes (N_e for each population, $nu1$ and $nu2$), time since divergence (T), and time of secondary contact (T_{sc}), depending on model. The best-fit model was then bootstrapped to provide a 95% confidence interval around each parameter estimated (bootstrap_dadi.py v.1.1; https://github.com/jfmclaughlin92/beringia_scripts/).

To calculate substitution rates, we used BLASTn to compare our mallard reference FASTA sequence to a fossil-calibrated node within the lineage of Anseriformes, the swan goose (*Anser cygnoides*; AnsCyg_PRJNA183603_v1.0) with a dated node of ~28 Ma (Claramunt & Cracraft, 2015). Calculations for substitution rates, generation time, and adjusted length of sequences surveyed (Table S2) were then used with the best-fit model parameter estimates

obtained from our demographic analyses to provide biological estimates of ancestral population size (N_{ref}), size of populations ($nu1$, $nu2$), time since split (T), migration (gene flow in individuals/generation; derived from m), migration from population 1 into population 2 (individuals/generation; derived from $m12$), migration from population 2 into population 1 (individuals/generation; derived from $m21$), and time of secondary contact (T_{sc}) as appropriate (based on the best-fit model). The estimate for ancestral population size (Table S2) is derived from the output of θ from $\delta a \delta i$; where $\theta = 4 * (N_{ref}) * (\text{substitution rate}) * (\text{adjusted length of sequences})$.

3. Results

3.1. Assembly and quality control

After assembly and filtering for UCE loci, we recovered 1,905 loci to create our *A. platyrhynchos* reference FASTA sequence. Following the bioinformatics pipeline for all individuals of read mapping, SNP calling, and quality filtering, the resulting dataset containing only high-quality data at all loci for all individuals (100% complete matrix) had 1,393 UCE loci, with a total length of 710,532 bp and a mean per-locus length of 510.1 bp (± 3.9 bp, 95% CI). The data matrix contained 1,204 variable loci and 189 invariable loci, and a total of 4,940 SNPs. For demographic analyses using $\delta a \delta i$, this high-quality dataset was thinned to be biallelic and contain one SNP per locus, which retained 1,202 variable loci. Identifying and removing Z-linked loci further reduced the number to 1,118 loci. Nucleotide diversity (π) was similar among the taxa (Table S3).

3.2. mtDNA and nuDNA phylogenies, and inferred gene tree history

The maximum likelihood phylogeny from the complete mitogenomes shows *A. c. carolinensis* sister to *A. flavirostris* and *A. c. crecca* sister to *A. c. nimia* (Fig. 2A). This topology

is consistent with previous studies using mtDNA (Johnson & Sorenson, 1999, Gonzalez et al., 2009). Estimates of total mtDNA divergence were 3.7% between *A. c. crecca* and *A. c. carolinensis* and 2.2% between *A. c. carolinensis* and *A. flavirostris*. These divergences are commensurate with those reported in the literature (Johnson & Sorenson, 1999; Humphries & Winker, 2011) that indicate a deeper mtDNA split between *A. c. crecca* and *A. c. carolinensis* than between the latter and *A. flavirostris*.

In contrast to mtDNA, which places *A. c. crecca* subspecies into two different well-resolved clades with one of these sister to *A. flavirostris* (Fig. 2A), our nuDNA (UCEs) phylogeny shows a much different topology, with all *A. crecca* individuals (*A. c. crecca*, *A. c. nimia*, and *A. c. carolinensis*) forming a polytomous clade that is sister to *A. flavirostris* (Fig. 2B). In this phylogeny, there was low bootstrap support for the nodes within *A. crecca*, resulting in a lack of structure at the subspecies level, whereas *A. flavirostris* individuals formed their own clade with 100% bootstrap support (Fig. 2B). The inferred UCE gene tree history from our SNAPP analysis showed the same sister relationship between *A. crecca* (*sensu lato*) and *A. flavirostris*, with considerable gene-tree conflicts evident among the *A. crecca* subspecies (Fig. 3). The mtDNA topology (Fig. 2A) thus falls outside of the 95% CI of the nuDNA SNAPP topology (Fig. 3), indicating that the two genomic histories are significantly different and reflecting mitochondrial-nuclear discord.

3.3. Divergence models and gene flow

F_{ST} values for each pairwise comparison (Table 1) ranged from a low of 0.050 (*A. c. crecca*—*A. c. nimia*; $P = 0.02$) to a high of 0.331 (*A. c. carolinensis*—*A. flavirostris*; $P = 0.01$). A principal components analysis (PCA) of *A. flavirostris* and the three subspecies of *A. crecca* using our thinned biallelic VCF file (Fig. 4) reflected the F_{ST} values, with the three *A. crecca*

subspecies clustering closely together and *A. flavirostris* divergent from this group. The DAPC analysis also showed the *A. crecca* complex clustering together, with *A. flavirostris* being divergent (Fig. S2). In contrast to the nuDNA and mtDNA phylogenetic trees (Figs. 2, 3) and the DAPC analysis (Fig. S2), however, *A. flavirostris* shows pronounced within-taxon differentiation in the nuDNA PCA space (Fig. 4), likely reflecting its superspecies status, with our single sample of *A. [f.] andium* being distinct from our other samples of the *A. flavirostris* superspecies (Fig. 4).

Our PCA plot (Fig. 4) and the DAPC analysis (Fig. S2) clustered all three subspecies of *A. crecca* together (with a small separation of *carolinensis* from *crecca+nimia* in PCA space; Fig. 4). The results for expected heterozygosity (H_e) versus observed heterozygosity (H_o) showed all comparisons to have significant differences, indicating a deviation from Hardy-Weinberg equilibrium and suggesting gene flow (Table S4).

The best-fit models for our demographic analyses in $\delta a d i$ found gene flow present in all pairwise comparisons (Fig. 5), as indicated by the AIC values (Table 2). When selecting best-fit models, some models had a ΔAIC of <10 , which indicates models that are not statistically separable in their likelihood of explaining the data. Calculating ΔAIC causes the best model to have $\Delta AIC = 0$, while the rest of the models have positive values (Burnham & Anderson, 2004). All of the statistically best-fit models included gene flow (Tables 2, S5). Among the *A. crecca* subspecies, there were effectively statistical ties between split-with-migration models (i.e., divergence with ongoing gene flow) and secondary-contact models (i.e., divergence with some isolation before resumption of gene flow; Tables 2, S5). For demographic analyses, we chose the best-fit model to be the one for each pairwise comparison with the lowest AIC value, a $\Delta AIC = 0$, and a weighted AIC = 1 (Tables 2, S5). The best-fit models chosen were: split-with-

symmetric-migration (Fig. S1C) for *A. c. carolinensis*–*A. c. nimia* and *A. c. crecca*–*A. c. nimia*; secondary-contact-with-symmetric-migration (Fig. S1G) for *A. c. crecca*–*A. c. carolinensis*; and split-with-symmetric-migration-and-exponential-population-growth (Fig. S1F) for *A. c. carolinensis*–*A. flavirostris*.

The magnitude of inferred gene flow between these pairwise comparisons varied from ~1 to ~26 individuals per generation within the three *A. crecca* subspecies, with the highest levels of gene flow from *A. c. carolinensis* into *A. c. nimia* (Table 3, Fig. 5). The lowest levels of gene flow were ~0.4 to ~0.1 individuals per generation, which occurred in the *A. c. carolinensis*–*A. flavirostris* comparison (Table 3, Fig. 5). Estimates for ancestral population sizes (N_{ref}) ranged from ~19,700 individuals (in *A. c. crecca*–*A. c. carolinensis*) to 180,500 individuals (in *A. c. carolinensis*–*A. flavirostris*; Table 3). Values for effective population sizes (N_e) ranged from a low of ~366,400 individuals (for *A. c. crecca* in *A. c. crecca*–*A. c. carolinensis*) to ~16,900,000 individuals (for *A. c. carolinensis* in *A. c. carolinensis*–*A. flavirostris*; Table 3). UCE-based time since divergence (T) ranged from ~105,000 years in *A. c. crecca*–*A. c. carolinensis* to ~744,000 years in *A. c. carolinensis*–*A. flavirostris* (Tables 3, S6). The *A. c. crecca*–*A. c. carolinensis* comparison was the only one with the very best-fit model being one of secondary contact, in which time of secondary contact (T_{sc}) occurred ~43,000 years ago coinciding with the end of the last glacial maximum (McLaughlin et al., 2020; Table 3, Fig. 5).

Finally, secondary $\delta\text{a}\delta\text{i}$ demographic modeling analyses were run to tease out more information about suggested gene flow between *A. c. carolinensis* and *A. flavirostris*, using our 3 *A. f. oxyptera/flavirostris* and 1 *A. f. andium* specimens. Removing population structure from these analyses can lower variability among model runs, probably making outcomes more accurate despite lower sample sizes (e.g., McLaughlin and Winker, 2020). All of the best-fit

models included gene flow (Table S7). The best-fitting model between *A. c. carolinensis* and *A. f. andium* was the same as that between *A. c. carolinensis* and *A. flavirostris sensu lato*: split with exponential population growth and two migration parameters (Table S7, Figs. 1, S1F). The *A. c. carolinensis*—*A. f. oxyptera/flavirostris* pairwise comparison showed effectively a statistical tie among four models: split with migration and secondary contact with one- and two- migration parameters (Table S7; Fig. S1C, D, G, and H). The parameter estimates for gene flow in these models were commensurate with those in the original contrast, including the directionality in the model supporting two migration parameters (more *A. c. carolinensis* into *A. f. andium* than the reverse; Tables 3, S6, and S8).

4. Discussion

We found strong discordance between nuclear and mitochondrial phylogenies in the green-winged teal and yellow-billed teal complex (Figs. 2, 3). Our mitochondrial phylogeny using complete mitogenomes showed a sister relationship between *A. crecca carolinensis* and *A. flavirostris* and deeply divergent haplogroups within *A. crecca* (Fig. 2A), which agreed with previous, more limited mtDNA studies (Johnson & Sorenson, 1999; Gonzalez et al., 2009; Humphries & Winker, 2011; Peters et al., 2012; Peters et al., 2014). In contrast, nuclear phylogenies showed the *A. crecca* subspecies complex as an unresolved polytomy, with *A. flavirostris* as its sister (Fig. 2B, Fig. 3). Our other major finding is that divergence occurred with gene flow in all pairwise demographic analyses. We found no evidence of classic allopatric speciation, although levels of gene flow between *A. c. carolinensis* and *A. flavirostris* were low (less than one individual per generation), approaching the classic allopatric condition of no gene flow (as defined by Mayr, 1942; Gavrilets, 2003, 2004; Coyne & Orr, 2004; Fitzpatrick et al., 2008; Price, 2008). Periods of allopatry will foster divergence by allowing for reproductive

isolation to develop. However, given the absence of divergence models with secondary contact being supported as best-fit for the *A. c. carolinensis* and *A. flavirostris* comparison (Tables 2, S7), strict allopatric speciation (where the level of gene flow is zero) does not seem to be occurring.

4.1. Phylogenetic analyses of mtDNA and nuDNA reveal discord

There is strong evidence for mitonuclear discord in this group. The mitochondrial phylogeny showed *A. c. carolinensis* and *A. flavirostris* to be sister taxa, with *A. c. crecca* and *A. c. nimia* being sister to those (Fig. 2A). In contrast, our UCE phylogeny (Fig. 2B) and inferred UCE gene tree history (Fig. 3) showed *A. flavirostris* to form its own clade, sister to the *A. crecca* complex rather than to the subspecies *A. c. carolinensis* as indicated by the mitochondrial phylogeny (Fig. 2A). Furthermore, our mitogenomic phylogeny (Fig. 2A) matched previous mtDNA studies on this group (Johnson & Sorenson, 1999; Gonzalez et al., 2009). Després (2019) suggested that mitonuclear discord is a consequence of demographic fluctuations, for example when a large population is fragmented into small isolates, such as during glacial periods, which can result in founder events. Divergence in mitochondrial DNA can be stochastic with high potential to not reflect the underlying species tree even though it has one-quarter the effective population size (N_e) of autosomal nuclear loci and is less prone to incomplete lineage sorting (Moore, 1995; Hudson & Turelli, 2003; Zink & Barrowclough, 2008), making it more likely to track the species tree than a single nuclear locus. Given that Beringia has historically experienced cyclic fluctuations of climate change during the Pleistocene, climate fluctuations could have contributed to the mitonuclear discord within this group. Mitonuclear discord is prevalent in other Holarctic avian taxa that occur in Beringia, such as the mallard (*Anas platyrhynchos*; Peters et al., 2014), the Eurasian and American wigeons (*Mareca penelope* – *M.*

americana; Humphries & Winker, 2011; Peters et al., 2014), and the *Eremophila* lark complex (Drovetski et al., 2014). However, unlike most birds, ducks have male-biased dispersal, and male-mediated intercontinental nuclear gene flow combined with female philopatry might be a leading contributor to this case of mtDNA and nuDNA discord (Peters et al., 2012, 2014; Fig. 3, Table 3).

We also note differences between the branch lengths in the mtDNA phylogeny (Fig. 2A) versus the nuDNA phylogeny (Fig. 2B). In the latter, the long terminal branch lengths could be due in part to recombination, or, more likely, to uncertainties in analyses of branch length estimation in shallow trees (Madison & Knowles, 2006; Kubatko et al., 2009; Lanier & Knowles, 2012). Recombination has not been effectively measured for UCE loci, but Winker et al. (2018) found loci exhibiting patterns indicative of recombination in around 15-25% of variable loci, suggesting that there are likely some recombinant haplotypes among our UCE sequences. Recombination within a locus could be a problem in sequence-based species tree reconstruction (Fig. 2B) because it might cause a mismatch between estimated gene trees and the true genealogies, and concatenating loci ignores recombination and makes the assumption that all loci have the same genealogy. Lanier & Knowles (2012) found that recombination had little effect on species-tree accuracy using both Bayesian and ML-based analyses, and that other factors such as total tree depth and sample sizes of individuals and loci were much more important.

While the full effects of recombination on tree accuracy remain unclear, their effects appear to be small at the shallow evolutionary depths of our study and are thus not considered to unduly affect species-tree accuracy (Lanier & Knowles, 2012; Edwards et al., 2016; Springer & Gatesy 2018; Van Dam et al., 2021). Our inferred UCE gene tree history obtained through the SNAPP analysis (Fig. 3) uses one SNP per locus and therefore is not affected by recombination

(although it might be affected by linkage disequilibrium; Bryant et al., 2012). In contrast, our UCE tree (Fig. 2B) provides a “winner take all” perspective through concatenation (Maddison, 1997), and while it does not account for recombination within a locus the majority of loci are likely not affected. Despite these two different methodologies, the two topologies are very similar, with the *A. crecca* complex forming a polytomous clade with *A. flavirostris* being sister to this complex.

Estimates of mitochondrial divergence show *A. c. crecca* and *A. c. carolinensis* to be more divergent (3.7%), than *A. c. carolinensis* and *A. flavirostris* (2.2%). Using smaller mtDNA datasets and substitution rates appropriate for those data (Peters et al., 2008; 2014), the dates of these divergences have been estimated at ~2.6 Mya between *A. c. crecca*—*A. c. carolinensis* and ~1.1 Mya between *A. c. carolinensis* and *A. flavirostris*. We note the differences here between estimates from mtDNA versus UCEs, with shallower divergences found in UCEs (Table 3). Nuclear DNA generally has lower substitution rates than mitochondrial DNA, and so it might be concluded that nuDNA has lower utility in estimating population divergence times (Arbogast et al., 2002). But male-mediated gene flow among *A. crecca* subspecies has also likely affected the nuclear *A. c. crecca*—*A. c. carolinensis* divergence time estimate, making it shallower (males disperse farther than females, and mtDNA is maternally inherited, causing nuDNA to disperse farther than mtDNA in this system). The relative depths of the nuclear divergences (Table 3) also seem to be different than those depicted in the UCE SNAPP phylogeny (Fig. 3, blue, red, green), but we note that the pairwise comparisons with the youngest divergence dates also have the highest rates of gene flow inferred. Calibration points for estimating substitution rates also affect dating estimates (Arbogast et al., 2002), and might affect these mtDNA and nuDNA divergence estimates. If there had been no gene flow between Eurasian and North American populations of

A. crecca, and substitution rate calibrations between marker types were accurate, we would expect to see the nuclear phylogeny match the mtDNA phylogeny, barring uncertainties stemming from incomplete lineage sorting. It has been established that deep mtDNA lineages do not always represent significant population divergences (Irwin, 2002; Zink & Barrowclough, 2008; Collins & Cruickshank, 2012; Morgan-Richards et al., 2017). This study highlights how biological species can have paraphyletic mitochondrial relationships, adding to many other cases of this taxonomically widespread phenomenon (Funk & Olmand, 2003).

The discordant phylogenies in these ducks make the biogeographic history uncertain. Lower female dispersal rates and male-mediated nuclear gene flow make it likely that the mtDNA history (Fig. 2A) is a more accurate reflection of the biogeographic history of these lineages, but the nuclear genome adds important information to our understanding of their subsequent divergence and speciation. One hypothesis is that after Eurasia and North America were colonized during the Pleistocene (~2.6 Mya) and mtDNA divergence between *A. c. crecca* and *A. c. carolinensis* was well established, South America was colonized by the proto-*carolinensis* ancestor of *A. flavirostris* (this is effectively the same hypothesis of Johnson & Sorenson, 1999:802). Under this scenario, the intermittent or ongoing nuclear gene flow between Eurasian and North American populations confounded this biogeographic history, resulting in a single Holarctic biological species, while a later intercontinental colonist (*A. flavirostris*) emerged from this Holarctic complex, achieved a much higher degree of isolation, and became its own well-differentiated biological species that currently resides in South America and continues to differentiate (Fig. 4). Other hypotheses can be given, however (e.g., including mtDNA capture), and Johnson & Sorenson's (1999) biogeographic reconstruction of ancestral areas was equivocal for this group.

Modern *A. c. carolinensis* winter south almost to South America, and numerous records show individuals on the continent in Venezuela, Colombia, and Ecuador, and as far east as French Guiana (Meyer de Schauensee, 1966; Scott & Carbonell, 1986; Botero & Rusch, 1988; Renaudier et al., 2010; eBird, 2021). This wintering range might have been even farther south during prior glacial maxima (a period spanning ~740 Kyr; Table 3). The phenomenon of migratory lineages dropping out new populations that establish new breeding grounds on or near tropical wintering grounds and subsequently differentiating and becoming new taxa occurs fairly often among numerous orders and families of birds (e.g., Bildstein, 2004; Winker, 2010; Winkler et al., 2017; Gómez-Bahamón et al., 2020). This seems to be a likely mechanism for the initial colonization of South America by the ancestors of *A. flavirostris*. It is also likely that a similar phenomenon at an individual level caused the ongoing or intermittent gene flow suggested between *A. c. carolinensis* and *A. flavirostris*. These birds form pair bonds on the wintering grounds, and it is plausible that wintering *A. c. carolinensis* males occasionally pair bond with *A. flavirostris* females and remain there to reproduce rather than return to northern breeding grounds. Our data suggest that over evolutionary time this occurs ~4 times every ten generations, which is not inconceivable for a highly mobile source population with an effective population size of over 16 million birds (Table 3). This, again, is only a hypothesis, however, and it bears further testing with larger sample sizes (in loci and/or individuals). These species are known to hybridize in captivity, even with our outgroup, *A. platyrhynchos* (Gray 1958, McCarthy 2006). Further, most of the best-fitting models in our series of three pairwise comparisons between *A. c. carolinensis* and *A. flavirostris sensu lato* (*A. f. oxyptera/flavirostris*, and *A. f. andium*) support asymmetric gene flow in the predicted direction: more from *A. c. carolinensis* into *A. flavirostris* than the opposite (Tables 2, S7, and S8).

4.2. Divergence models reveal gene flow

We expected divergence with gene flow to occur among the three subspecies of *Anas crecca* given previous work, but our findings extend this understanding using a large number of orthologous nuclear loci and place the results in a directly comparable framework (Peters et al., 2012; Winker et al., 2013).

Prior work by Peters et al. (2012) on these taxa showed that Eurasian and North American populations of *A. crecca* are genetically rather distinct, with gene flow occurring between the two continental populations where they come into contact in Beringia. Because our interest is in levels of gene flow in this region, this was the source of our *A. c. crecca* and *A. c. carolinensis* specimens (Fig. 1). While this might be surmised to predispose our results to showing higher levels of gene flow than more widespread sampling might show, our results are concordant with those from the range-wide study of Peters et al. (2012), with overall modest levels of gene flow recovered in both studies. This concordance is likely due to each individual's genome being an amalgamation of its entire lineage metapopulation, including individuals and populations that have not been sampled (Mazet et al., 2015). For thousands of loci with low substitution rates in a highly migratory duck in which mate selection occurs on the wintering grounds, we expect strong population mixing and thus to recover strong lineage-specific demographic attributes.

We detected gene flow between the three subspecies of *A. crecca*, with substantial levels between *A. c. crecca* and *A. c. carolinensis*, which have deeply diverged mtDNA lineages (Table 3, Fig. 2A). We could not statistically separate models of secondary contact (one or two migration parameters) for these subspecies (Tables 2, S5). Our F_{ST} results show significant values in each pairwise comparison. Further work with more loci and perhaps larger sample sizes

might provide better resolution among competing divergence-with-gene-flow models (Table 2) and for parameter estimates that are quite variable (e.g., effective population sizes; Table 3).

4.3. Speciation modes

The pairwise comparisons within the *A. crecca* complex reaffirm the importance of parapatric and heteropatric (seasonally sympatric) divergence processes in this group, given the lineages' geographic relationships and that gene flow was detected in all comparisons (Fig. 5; Table 2; Peters et al., 2012, Winker et al., 2013). Parapatric speciation, here occurring between Eurasian and North American continental lineages, is thought in this case to be driven by divergent selection stemming from sexual selection and site fidelity to different wintering grounds, although countered here by rather high levels of nuclear gene flow (Hartl & Clark, 1989; Rice & Hostert, 1993; Hostert, 1997; Price, 2008). This has resulted in only partial reproductive isolation between these populations, causing them to be stalled short of complete speciation, despite deeply divergent mitochondrial DNA (Peters et al., 2012). Heteropatric speciation is a type of ecological speciation driven by divergent selection occurring between lineages that are in sympatry and allopatry at different times during cyclic seasonal migrations (Winker, 2010; Winker et al., 2013). Both parapatric and heteropatric speciation involve gene flow between populations. However, periods of allopatric isolation (migration ≈ 0 ; Harrison, 2012), associated with glacial cycles at high latitudes and also seasonally at low latitudes, might have been important in lineage-specific evolutionary change, as seems possible in the *A. c. crecca-carolinensis* split and especially likely in *A. flavirostris* (Tables 2, 3). Finally, it is somewhat surprising that full allopatric isolation does not seem to have occurred in *A. flavirostris*, but that instead speciation with gene flow (albeit low) was strongly supported. The breeding ranges of *A. flavirostris* and *A. c. carolinensis* are fully allopatric and on different

continents. But this isolation is apparently not complete, and from these data it seems that strict allopatric divergence might not have occurred. This could be the result of periodic recolonization of South America by migratory *A. c. carolinensis*, likely from their wintering grounds, which currently extend to northern Central America and the Northern Caribbean, with occasional birds recorded from South America. This hypothesis bears further investigation.

4.4. Conclusions

We found mitonuclear discord within this group when comparing phylogenies reconstructed from the mitochondrial and nuclear genomes (Figs. 2, 3). Overall, our mitochondrial topology (Fig. 2A) matched relationships found in previous studies, in which *A. c. carolinensis* was sister to *A. flavirostris*, and *A. c. crecca* was sister to *A. c. nimia*. However, our nuclear phylogenies (Figs. 2B, 3) disagreed with the mitochondrial results, indicating that *A. crecca* subspecies formed one polytomy with *A. flavirostris* being sister to this group. This result could be explained by male-biased dispersal and the substantial levels of nuclear gene flow that we found among *A. crecca* subspecies (~1–26 individuals per generation depending on pairwise comparison). We also found gene flow occurring between *A. c. carolinensis* and *A. flavirostris*, although these levels were low (~0.1–0.4 individuals per generation). This suggests that while *A. flavirostris* is largely allopatric, *A. c. carolinensis* individuals might occasionally recolonize South America, for example through males pair-bonding in winter with a female *A. flavirostris* and remaining with her to reproduce rather than following a more species-appropriate mate back to her northern breeding grounds. We thus found no evidence for classic allopatric speciation (divergence without gene flow) in this group. Divergence with gene flow appears to be the predominant mode in this group, and the patterns of this gene flow have likely caused *A. crecca*

(*sensu lato*) to be, mitochondrially, a paraphyletic biological species with respect to *A. flavirostris*.

Journal Pre-proofs

Acknowledgments

We thank the University of Alaska Museum and the Louisiana State University Museum of Natural Science for loans of tissue specimens and the specimen collectors who made it possible to do this research. We also thank the Kessel Fund for Northern Ornithology and the Friends of Ornithology for financial support, and Brant Faircloth for his input early in the project. This project was funded in part by the National Science Foundation [grant number DEB-1242267-1242241-1242260], by an Institutional Development Award (IDeA) from the National Institute of General Medical Sciences of the National Institutes of Health [grant number 2P20GM103395] and by the National Institute of General Medical Sciences of the National Institutes of Health [grant number RL5GM118990] through the University of Alaska Fairbanks Biomedical Learning and Student Training program. Naoki Takebayashi, Devin Drown, and two anonymous reviewers provided helpful comments and feedback on earlier drafts of the manuscript.

Data Availability Statement

Original sequence data have been deposited in the NCBI Sequence Read Archive (SRA; Table S1; projects PRJNA741698 and PRJNA393740).

Additional data are available on figshare: (a) the individual UCE sequences in a PHYLIP file (<https://figshare.com/s/f678d42ae94f520aef8e>); (b) the reference UCE sequence data FASTA file (<https://figshare.com/s/4b22451c51ece33e4e60>); and (c) the VCF file used in $\delta a \delta i$ analyses (<https://figshare.com/s/245b3e7843b7c0622114>).

Author Contributions

The study was conceived by Kevin Winker, Travis Glenn, Brant Faircloth, and Kevin McCracken. The data were generated by Travis Glenn and Brant Faircloth. The data were analyzed by Fern Spaulding and Rebecca Cheek. The first draft was written by Fern Spaulding. All authors contributed to the final version.

Figures

Figure 1. Distribution of samples used in this study, focusing on Beringia and South America (inset). *A. c. crecca* occurs throughout most of Eurasia and seasonally migrates through the western Aleutians (light blue; 6 individuals), *A. c. nimia* occurs as a resident in the Aleutian Islands (dark blue; 3 individuals), and *A. c. carolinensis* occurs across North America (gold; 7 individuals). The yellow-billed teal, *Anas flavirostris* (red; 4 individuals) is a South American sister taxon; our samples include two members of this superspecies complex, *A. [f.] flavirostris* and *A. [f.] andium*. See Supplementary Table S1 for specimen details. Illustration of green-winged teal courtesy of USFWS (Hines, 1963).

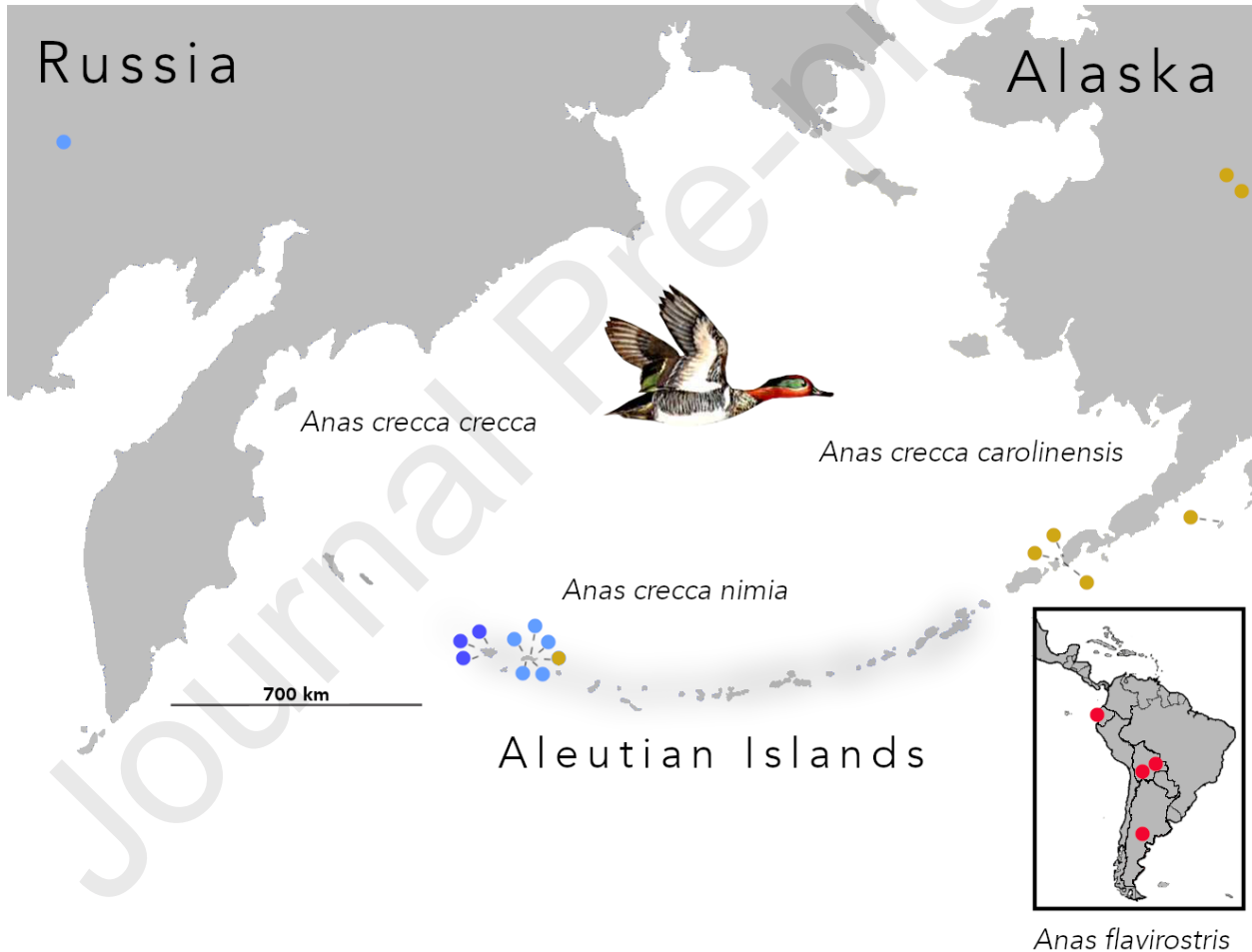


Figure 2. Maximum likelihood phylogenies of mitogenomic and nuclear DNA sequences. Taxon colors correspond to Figure 1. **A)** Phylogeny of complete mitogenomes with 100 bootstrap replicates using MEGA (node values: 0-1). **B)** Phylogeny of UCEs with 100 bootstrap replicates (node values: 1-100), reconstructed using a 95% complete data matrix using RAxML. Values on internal nodes differ due to different programs being used for phylogenetic reconstruction. Adjacent to each phylogeny is a sketch illustration highlighting the discordance between the topologies, with the relationship of the *A. crecca* subspecies (within the circle) relative to the sister taxon *A. flavirostris*. The mallard (*Anas platyrhynchos*) was used as an outgroup.

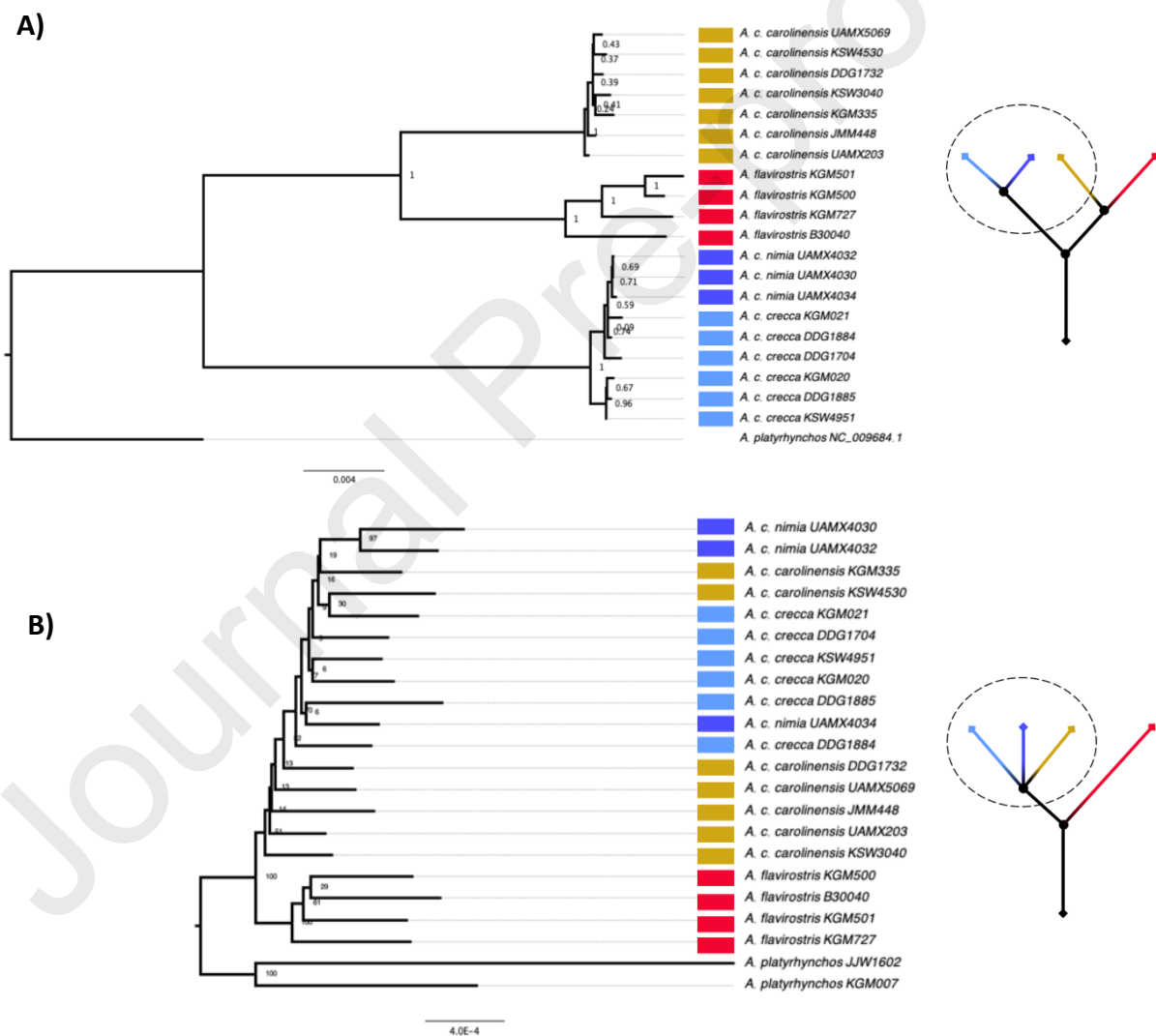


Figure 3. Inferred UCE gene tree history from a 30-million generation SNAPP analysis. Colors in the tree correspond to 1st- (blue), 2nd- (red), and 3rd-most (green) supported topologies. All individuals within the *A. crecca* complex form a polytomy and are illustrated by the colored gradient box, whereas the *A. flavirostris* clade is illustrated by the red box. The mallard (*A. platyrhynchos*) was used as an outgroup. Taxon colors correspond to Figure 1.

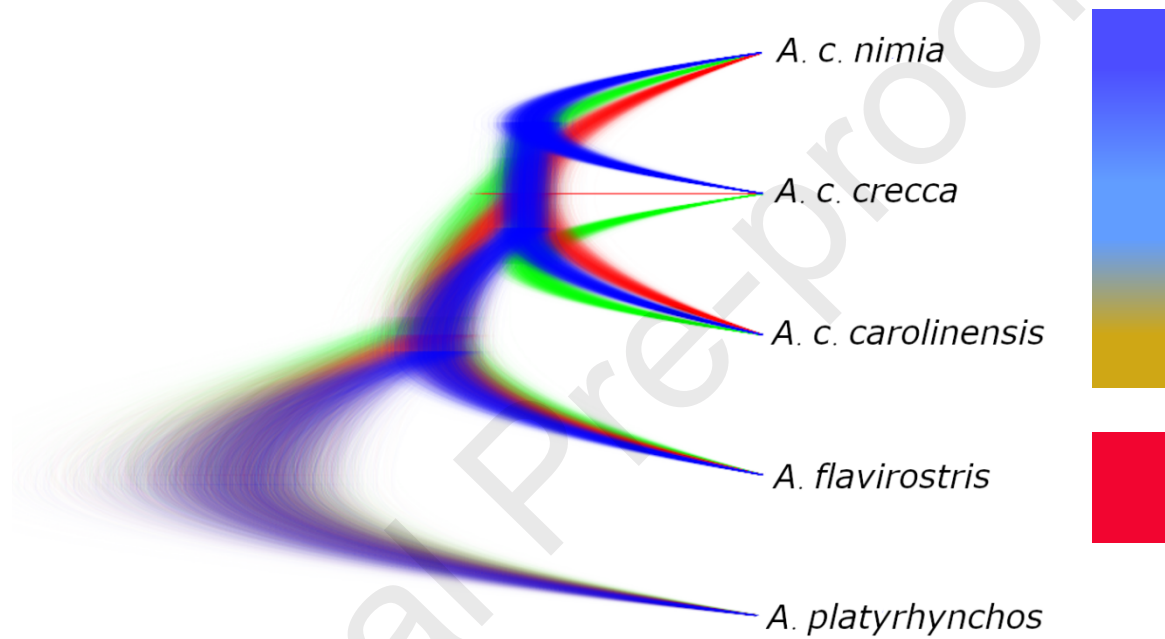


Figure 4. Principal components analysis (PCA) of *A. flavirostris* and all three subspecies of *A. crecca*. The one *A. flavirostris* outlier is the individual from Ecuador (Table S1), a different subspecies, *A. flavirostris andium*, likely indicating geographic isolation playing a role in the genetic divergence found within *A. flavirostris* across its range in South America.

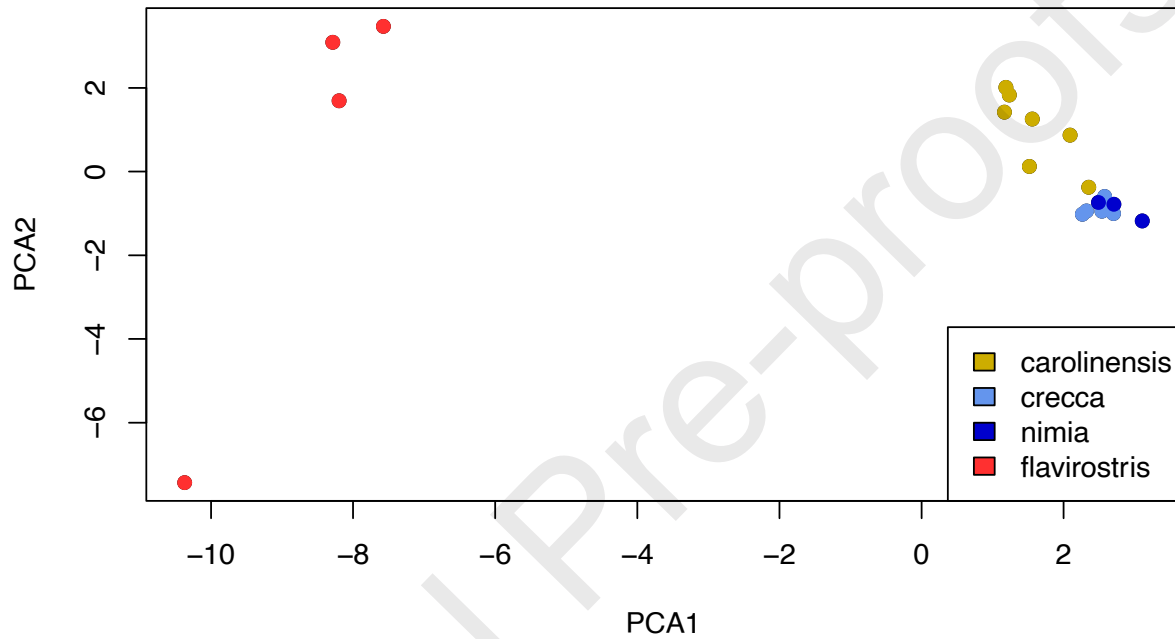
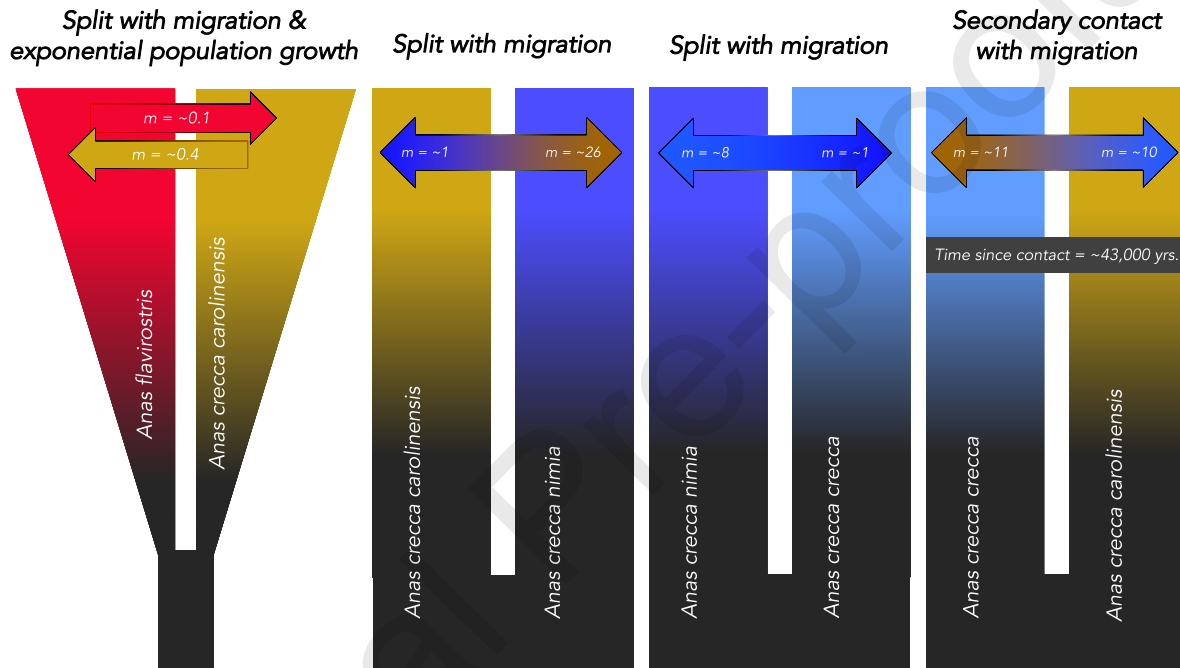


Figure 5. Visual representations of the best-fit demographic model for each pairwise comparison using $\delta a\delta i$ (Gutenkunst et al., 2009). Migration (gene flow) rate estimates (m) are in individuals per generation. All analyses indicated divergence with gene flow. Technically, all fit in parapatric speciation theory, but, geographically, allopatry and heteropatry are involved. Full biological estimates are given in Table 3. Color coding corresponds to Figure 1.



Tables

Table 1. UCE-based population pairwise comparisons for between-population F_{ST} and their corresponding P -values. Asterisks (*) indicate significant differences.

Pairwise Comparison	Between-population F_{ST}	P-value (F_{ST})
<i>A. c. carolinensis</i> — <i>A. flavirostris</i>	0.331	0.01*
<i>A. c. crecca</i> — <i>A. c. nimia</i>	0.050	0.02*
<i>A. c. carolinensis</i> — <i>A. c. nimia</i>	0.073	0.02*
<i>A. c. crecca</i> — <i>A. c. carolinensis</i>	0.057	0.02*

Table 2. AIC values and negative log-likelihood values for each pairwise comparison for the eight demographic models tested (“migration” = gene flow). For each model, AIC values are paired with the negative log-likelihood values in parentheses. The negative log-likelihood values were averaged from the five best runs. Best-fit models with a weighted AIC of 1 are in bold, while runner-up models ($\Delta AIC < 10$) are italicized. See Table S5 for ΔAIC values and weighted AIC values.

Pairwise comparison (nu1/nu2)	Neutral (Fig. S1A)	split with no migration (Fig. S1B)	split with migration, 1 parameter (Fig. S1C)	split with bidirectional migration, 2 parameters (Fig. S1D)	split with exponential population growth, no migration (Fig. S1E)	split with exponential population growth, migration (Fig. S1F)	secondary contact, 1 migration parameter (Fig. S1G)	secondary contact, 2 migration parameters (Fig. S1H)
<i>A. c. carolinensis</i> — <i>A. flavirostris</i>	1,311.27 (-653.64)	391.61 (-192.80)	371.74 (-181.87)	373.72 (-181.86)	422.01 (-207.01)	296.33 (-143.17)	373.51 (-181.75)	375.04 (-181.52)
<i>A. c. crecca</i> — <i>A. c. nimia</i>	590.79 (-293.40)	338.56 (-166.28)	229.11 (-110.55)	387.56 (-188.78)	590.37 (-291.19)	395.92 (-192.96)	<i>231.10</i> <i>(-110.55)</i>	<i>235.26</i> <i>(-111.63)</i>
<i>A. c. carolinensis</i> — <i>A. c. nimia</i>	508.67 (-252.34)	472.31 (-233.16)	220.39 (-106.20)	<i>223.05</i> <i>(-106.52)</i>	447.47 (-219.73)	303.57 (-146.79)	<i>227.97</i> <i>(-108.99)</i>	235.59 (-111.79)
<i>A. c. crecca</i> — <i>A. c. carolinensis</i>	882.88 (-439.44)	682.97 (-338.49)	478.58 (-235.29)	532.43 (-261.21)	887.25 (-439.62)	335.55 (-162.78)	310.90 (-150.45)	<i>311.95</i> <i>(-149.97)</i>

Table 3. Biological estimates obtained from the best-fit $\delta\alpha\delta i$ models for each pairwise comparison. Here we report ancestral population size (N_{ref} , in number of individuals), size of population 1 (N_e or $nu1$, in number of individuals), size of population 2 (N_e or $nu2$, in number of individuals), migration (gene flow) from population 1 into population 2 ($m12$ as individuals/generation), migration from population 2 into population 1 ($m21$, individuals/generation), time since split (T , in years), and time of secondary contact (T_{sc} , in years). See Figure 1.5 for a representation of the best-fit models. Values in parenthesis are the $\pm 95\%$ confidence interval around the biological estimates.

Pairwise comparison (nu1/nu2)	Best fit model via $\delta\alpha\delta i$	N_{ref}	N_e ($nu1$)	N_e ($nu2$)	Migration ($m12$)	Migration ($m21$)	Time since split (T)	Time of secondary contact (T_{sc})
<i>A. c. carolinensis</i> — <i>A. flavirostris</i>	split with migration & exponential population growth	180,524 ($\pm 9,731$)	16,898,942 ($\pm 231,289$)	747,090 ($\pm 59,819$)	0.44 (± 0.09)	0.10 (± 0.04)	743,758 ($\pm 38,436$)	-
<i>A. c. crecca</i> — <i>A. c. nimia</i>	split with migration	134,102 ($\pm 7,596$)	4,334,075 ($\pm 153,214$)	544,900 ($\pm 34,824$)	7.76 (± 0.38)	0.98 (± 0.05)	576,663 ($\pm 40,593$)	-
<i>A. c. carolinensis</i> — <i>A. c. nimia</i>	split with migration	148,577 ($\pm 11,955$)	11,611,924 ($\pm 210,055$)	457,268 ($\pm 64,171$)	25.61 (± 1.08)	1.01 (± 0.04)	436,583 ($\pm 113,537$)	-
<i>A. c. crecca</i> — <i>A. c. carolinensis</i>	secondary contact with migration	19,759 (± 779)	366,441 ($\pm 18,679$)	388,404 ($\pm 8,144$)	10.17 (± 1.84)	10.78 (± 1.95)	105,026 ($\pm 12,363$)	42,850 ($\pm 23,386$)

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. DOI: 10.1093/nar/gkw343.
- Akaike H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control.* 19, 716–723. DOI: 10.1109/TAC.1974.1100705.
- Arbogast, B.S., Edwards, S.V., Wakeley, J., Beerli, P., Slowinski, J.B. 2002. Estimating divergence times from molecular data on phylogenetic and population genetic timescales. *Annual Review of Ecology and Systematics.* 33, 707–740. DOI: 10.1146/annurev.ecolsys.33.010802.150500.
- Arnqvist, G., Edvardsson, M., Friberg, U., Nilsson, T. 2000. Sexual conflict promotes speciation in insects. *Proceedings of the National Academy of Sciences of the United States of America.* 97, 10460–10464. DOI: 10.1073/pnas.97.19.10460.
- Bellrose, F.C. 1976. Ducks, geese, and swans of North America. Harrisburg, PA: Stackpole Books.
- Berlacher, S.H., Feder, J.L. 2002. Sympatric speciation in phytophagous insects: moving beyond controversy? *Annual Review of Entomology.* 47, 773–815. DOI: 10.1146/annurev.ento.47.091201.145312.
- Bolger, A. M., Lohse, M., Usadel, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics.* 30, 2114–2120. DOI: 10.1093/bioinformatics/btu170.
- Bildstein, K. 2004. Raptor migration in the Neotropics: patterns, processes, and consequences. *Ornitologia Neotropical.* 15, 83–99.
- Bolnick, D.I., Fitzpatrick, B.M. 2007. Sympatric speciation: models and empirical evidence. *Annual Review of Ecology, Evolution, and Systematics.* 38, 459–487. DOI: 10.1146/annurev.ecolsys.38.091206.095804.
- Botero, J.E., Rusch, D.H. 1988. Recoveries of North American waterfowl in the Neotropics. *Waterfowl in winter*, edited by Milton W. Weller. University of Minnesota Press, Minneapolis, MN, USA.
- Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... Drummond, A.J. 2019. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLOS Computational Biology.* 15, e1006650. DOI: 10.1371/journal.pcbi.1006650.
- Bouckaert, R.R. 2010. DensiTree: Making sense of sets of phylogenetic trees. *Bioinformatics.* 26, 1372–1373. DOI: 10.1093/bioinformatics/btq110.
- Boul, K.E., Funk, W.C., Darst, C.R., Cannatella, D.C., Ryan, M.J. 2007. Sexual selection drives speciation in an Amazonian frog. *Proceedings of the Royal Society B: Biological Sciences.* 274, 399–406. DOI: 10.1098/rspb.2006.3736.
- 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. *Molecular Biology and Evolution.* 29, 1917–1932. DOI: 10.1093/molbev/mss086.
- Buehler, D., Baker, A. 2005. Population divergence times and historical demography in red knots and dunlins. *Condor.* 107, 497–513. DOI: 10.1093/condor/107.3.497.
- Burnham, K.P., Anderson D.R. 2004. Multimodel inference: understanding AIC and BIC in model selection. *Sociological Methods & Research.* 33, 261–304. DOI: 10.1177/0049124104268644.

- Butlin, R.K., Galindo, J., Grahame, J.W. 2008. Sympatric, parapatric or allopatric: the most important way to classify speciation? *Philosophical Transactions of the Royal Society B: Biological Sciences*. 363, 2997–3007. DOI: 10.1098/rstb.2008.0076.
- Castro-Insua, A., Gómez-Rodríguez, C., Wiens, J.J., Baselga, A. 2018. Climatic niche divergence drives patterns of diversification and richness among mammal families. *Scientific Reports*. 8, 8781. DOI: 10.1038/s41598-018-27068-y.
- Claramunt, S., Cracraft, J. 2015. A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*. 1, e1501005. DOI: 10.1126/sciadv.1501005.
- Collins, R.A., Cruickshank, R.H. 2012. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*. 13, 969–975. DOI: 10.1111/1755-0998.12046.
- Cooney, C.R., Seddon, N., Tobias, J.A. 2016. Widespread correlations between climatic niche evolution and species diversification in birds. *Journal of Animal Ecology*. 85, 869–878. DOI: 10.1111/1365-2656.12530.
- Coyne, J.A., Orr, H.A. 2004. *Speciation*. MA: Sinauer Associates.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... 1000 Genomes Project Analysis Group. 2011. The variant call format and VCFtools. *Bioinformatics*. 27, 2156–2158. DOI: 10.1093/bioinformatics/btr330.
- Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., Flouri, T. 2020. ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. *Molecular Biology and Evolution*. 37, 291–294. DOI: 10.1093/molbev/msz189.
- Després, L. 2019. One, two or more species? Mitonuclear discordance and species delimitation. *Molecular Ecology*. 28, 3845–3847. DOI: 10.1111/mec.15211.
- del Hoyo J, Collar NJ, Eds. 2014. *HBW and Birdlife International Illustrated Checklist of the Birds of the World. Part 1*. Barcelona: Lynx Edicions and Birdlife International.
- Drovetski S.V., Raković, M., Semenov, G., Fadeev, I.V., Red'kin, Y.A. 2014. Limited phylogeographic signal in sex-linked and autosomal loci despite geographically, ecologically, and phenotypically concordant structure of mtDNA variation in the holarctic avian genus *Eremophila*. *PLoS ONE*. 9, e87570. DOI: 10.1371/journal.pone.0087570.
- Drovetski, S.V., Zink, R.M., Rohwer, S., Fadeev, I.V., Nesterov, E.V., Karagodin, I., Koblik, E.A., Red'kin, Y.A. 2004. Complex biogeographic history of a Holarctic passerine. *Proceedings of the Royal Society B, Biological Sciences*. 271, 545–551. DOI: 10.1098/rspb.2003.2638.
- eBird. 2021. eBird: An online database of bird distribution and abundance. eBird, Cornell Lab of Ornithology, Ithaca, New York. Available: <http://www.ebird.org>. (Accessed: 11 August 2021).
- Edwards, S., Bensch, S. 2009. Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough. 2008. *Molecular Ecology*. 14, 2930–2933. DOI: 10.1111/j.1365-294X.2009.04270.x.
- Edwards, S.V., Kingan, S.B., Calkins, J.D., Balakrishnan, C.N., Jennings, W.B., Swanson, W.J., Sorenson, M.D. 2005. Speciation in birds: genes, geography, and sexual selection. *Proceedings of the National Academy of Sciences*. 102, 6550–6557. DOI: 10.1073/pnas.0501846102.

- Elias, S., Brigham-Grette, J. 2013. Late Pleistocene glacial events in Beringia. *Encyclopedia of Quaternary Science*. 2, 191–201. DOI: 10.1016/B978-0-444-53643-3.00116-3.
- Edwards, S. V., Xi, Z., Janke, A., Faircloth, B. C., McCormack, J. E., Glenn, T. C., Zhong, B., Wu, S., Lemmon, E. M., Lemmon, A. R., Leaché, A. D., Liu, L., Davis, C. C. 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics, *Molecular Phylogenetics and Evolution* 94, 447–462. DOI: 10.1016/j.ympev.2015.10.027.
- Endler, J.A. 1977. *Geographic variation, speciation, and clines*. NJ: Princeton University Press.
- Everson, K.M., McLaughlin, J.F., Cato, I.A., Evans, M.M., Gastaldi, A.R., Mills, K.K., Shink, K.G., Wilbur, S.M., Winker, K. 2019. Speciation, gene flow, and seasonal migration in *Catharus* thrushes (Aves:Turridae). *Molecular Phylogenetics and Evolution*. 139, 106564. DOI: 10.1016/j.ympev.2019.106564.
- Faircloth, B.C. 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. DOI: 10.6079/J9ILL.
- Faircloth, B.C. 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics*. 32, 786–788. DOI: 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. *Systematic Biology*. 61, 717–726. DOI: 10.1093/sysbio/sys004.
- Feder, J.L., Egan, S.P., Nosil, P. 2012. The genomics of speciation-with-gene-flow. *Trends in Genetics*. 28, 342–350. DOI: 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? *Molecular Biology and Evolution*. 23, 691–700. DOI: 10.1093/molbev/msj079.
- Fitzpatrick, B.M., Fordyce, J.A., Gavrillets, S. 2008. What, if anything, is sympatric speciation? *Journal of Evolutionary Biology*. 21, 1452–1459. DOI: 10.1111/j.1420-9101.2008.01611.x.
- Forbes, A.A., Devine, S.N., Hippee, A.C., Tvedte, E.S., Ward, A.K., Widmayer, H.A., Wilson, C.J. 2017. Revisiting the particular role of host shifts in initiating insect speciation. *Evolution*. 71, 126–1137. DOI: 10.1111/evo.13164.
- Funk, D.J., Omland, K.,E. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*. 34, 397–423. DOI: 10.1146/annurev.ecolsys.34.011802.132421.
- Futuyma, D.J., Mayer G.C. 1980. Non-allopatric speciation in animals. *Systematic Biology*. 29, 254–271. DOI: 10.1093/sysbio/29.3.254.
- Gavrillets, S. 2003. Perspective: models of speciation: what have we learned in 40 years? *Evolution*. 57, 2197–2215. DOI: 10.1111/j.0014-3820.2003.tb00233.x.
- Gavrillets, S. 2004. *Fitness landscapes and the origin of species*. NJ: Princeton University Press.
- Gavrillets, S. 2014. Models of speciation: where are we now? *Journal of Heredity*. 105, 743–755. DOI: 10.1093/jhered/esu045.

- Glenn, T.C., Nilsen, R., Kieran, T.J., Finger, J.W., Pierson, T.W., Bentley, K.E., ... Faircloth, B.C. 2019. Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially indexed Illumina libraries (iTru & iNext). *PeerJ*. 7, e7755. DOI: 10.7717/peerj.7755.
- Gómez-Bahamón, V., Márquez, R., Jahn, A.E. Miyaki, C.Y., Tuero, D.T., Laverde-R, O. Restrepo, S., and Cadena, C.D. 2020. Speciation associated with shifts in migratory behavior in an avian radiation. *Current Biology*. 30, 1312–1321. DOI: 10.1016/j.cub.2020.01.064.
- Gonzalez, J., Düttmann, H., Wink, M. 2009. Phylogenetic relationships based on two mitochondrial genes and hybridization patterns in Anatidae. *Journal of Zoology*. 279, 310–318. DOI: 10.1111/j.1469-7998.2009.00622.x.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., ... Regev, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*. 29, 644–652. DOI: 10.1038/nbt.1883.
- Gray, A. P. 1958. *Bird Hybrids*. Commonwealth Agricultural Bureaux, Farnham Royal, Bucks, England.
- 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 Genetics*. 5, e1000695. DOI: 10.1371/journal.pgen.1000695.
- Harrison, R.G. 2012. The language of speciation. *Evolution*. 66, 3643–3657. DOI: 10.1111/j.1558-5646.2012.01785.x.
- Hartl, D.L., Clark, A.G. 1989. *Principles of Population Genetics*, 2nd edition. Sinauer Associates, Inc., Sunderland, Massachusetts.
- 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. *Systematic Biology*. 65, 910–924. DOI: 10.1093/sysbio/syw036.
- Hernández-Hernández, T., Miller, E.C., Román-Palacios, C., Wiens, J.J. 2021. Speciation across the Tree of Life. *Biological Reviews*. 96, 1205–1242. DOI: 10.1111/brv.12698.
- Hewitt, G. 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society*. 58, 247–276. DOI: 10.1006/bijl.1996.0035.
- Hines, B. 1963. *Ducks at a distance: a waterfowl identification guide*. U.S. Department of the Interior, Fish and Wildlife Service.
- Holland, B.S., Cowie, R.H. 2009. Land snail models in island biogeography: a tale of two snails. *American Malacological Bulletin*. 27, 59–68. DOI: 10.4003/006.027.0205.
- Hoskin, C.J., Higgie, M., McDonald, K.R., Moritz, C. 2005. Reinforcement drives rapid allopatric speciation. *Nature*. 437, 1353–1356. DOI: 10.1038/nature04004.
- Hostert, E.E. 1997. Reinforcement: a new perspective on an old controversy. *Evolution*. 51, 697–702. DOI: 10.2307/2411146.
- Hudson, R., Turelli, M. 2003. Stochasticity overrules the “three-times rule”: genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*. 57, 182–190. DOI: 10.1111/j.0014-3820.2003.tb00229.x.
- Humphries, E.M., Winker, K. 2011. Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds: discord among divergence estimates. *Molecular Ecology*. 20, 573–583. DOI: 10.1111/j.1365-294X.2010.04965.x.

- Irwin, D.E. 2002. Phylogeographic breaks without geographic barriers to gene flow. *Evolution*. 56, 2383–2394. DOI: 10.1111/j.0014-3820.2002.tb00164.x.
- Johnson, D.H., Nichols, J.D., Schwartz, M.D. 1992. Population dynamics of breeding waterfowl. 446–485.
- Johnson, K.P., Sorenson, M.D. 1999. Phylogeny and biogeography of dabbling ducks (genus: *Anas*): a comparison of molecular and morphological evidence. *The Auk*. 116, 792–805. DOI: 10.2307/4089339.
- Jombart, T., Ahmed, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*. 27, 3070–3071. DOI: 10.1093/bioinformatics/btr521.
- Jordaens, K., van Riel, P., Frias Martins, A.M., Backeljau, T. 2009. Speciation on the Azores islands: congruent patterns in shell morphology, genital anatomy, and molecular markers in endemic land snails (Gastropoda, Leptaxinae). *Biological Journal of the Linnean Society*. 97, 166–176. DOI: 10.1111/j.1095-8312.2008.01197.x.
- Kalinowski, S.T. 2005. Do polymorphic loci require large sample sizes to estimate genetic distances? *Heredity*. 94, 33–36. DOI: 10.1038/sj.hdy.6800548.
- Kozak, K.H., Wiens, J.J. 2010. Accelerated rates of climatic-niche evolution underlie rapid species diversification. *Ecology Letters*. 13, 1378–1389. DOI: 10.1111/j.1461-0248.2010.01530.x.
- Kubatko L.S. 2009. Identifying hybridization events in the presence of coalescence via model selection. *Systematic Biology*. 58, 478–488. DOI: 10.1093/sysbio/syp055.
- Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K. 2018. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution*. 35, 1547–1549. DOI: 10.1093/molbev/msy096.
- Lanier, H.C., Knowles, L.L. 2012. Is recombination a problem for species-tree analyses? *Systematic Biology*. 61, 691–701. DOI: 10.1093/sysbio/syr128.
- 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. DOI: 10.1093/bioinformatics/btp324.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. *Bioinformatics*. 25, 2078–2079. DOI: 10.1093/bioinformatics/btp352.
- Lischer, H., Excoffier, L. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*. 28, 298–299. DOI: 10.1093/bioinformatics/btr642.
- Maddison, W.P. 1997. Gene trees in species trees. *Systematic Biology*. 46, 523–536. DOI: 10.1093/sysbio/46.3.523.
- Maddison W.P., Knowles L.L. 2006. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*. 55, 21–30. DOI: 10.1080/10635150500354928.
- Mallet, J., Besansky, N., Hahn, M.W. 2016. How reticulated are species? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*. 38, 140–149. DOI: 10.1002/bies.201500149.
- Mayr, E. 1942. *Systematics and the origin of species from the viewpoint of a zoologist*. NY: Columbia University Press.

- Mayr, E. 1963. *Animal species and evolution*. MA: Belknap Press of Harvard University Press.
- Mazet, O., W. Rodriguez, and L. Chikhi. 2015. Demographic inference using genetic data from a single individual: Separating population size variation from population structure. *Theoretical Population Biology* 104:46–58. <http://dx.doi.org/10.1016/j.tpb.2015.06.003>.
- McCarthy, E. M. 2006. *Handbook of Avian Hybrids of the World*. Oxford University Press, New York, NY.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., ... DePristo, M. A. 2010. The Genome Analysis Toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. 20, 1297–1303. DOI: 10.1101/gr.107524.110.
- McKinnon, J.S., Mori, S., Blackman, B.K., David, L., Kingsley, D.M., Jamieson, L., Chou, J., Schluter, D. 2004. Evidence for ecology's role in speciation. *Nature*. 429, 294–298. DOI: 10.1038/nature02556.
- McLaughlin, J.F., Faircloth, B.C., Glenn, T.C., & Winker, K. 2020. Divergence, gene flow, and speciation in eight lineages of trans-Beringian birds. *Molecular Ecology*. 29, 3526–3542. DOI: 10.1111/mec.15574.
- McLaughlin, J.F., Winker, K. 2020. An empirical examination of sample size effects on population demographic estimates in birds using single nucleotide polymorphism (SNP) data. *PeerJ*. 8, e9939. DOI: 10.7717/peerj.9939.
- Meyer de Schauensee, R. 1966. *The species of birds of South America and their distribution*. Livingston Publishing, Narberth, PA, USA.
- Moen, D.S., Wiens, J.J. 2017. Microhabitat and climatic niche change explain patterns of diversification among frog families. *American Naturalist*. 190, 29–44. DOI: 10.1086/692065.
- Moisan, G., Smith, R.I., Martinson, R.K. 1967. *The green-winged teal: Its distribution, migration, and population dynamics*. US Fish and Wildlife Service Special Report – Wildlife 100.
- Moore, W. 1995. Inferring phylogenies from mtDNA variation mitochondrial-gene trees versus nuclear-gene trees. *Evolution*. 49, 718–726. DOI: 10.1111/j.1558-5646.1995.tb02308.x.
- Morales, A.E., Jackson N.D., Dewey, T.A., O'Meara, B.C., Carstens, B.C. 2017. Speciation with gene flow in North American *Myotis* bats. *Systematic Biology*. 66, 440-452. DOI: 10.1093/sysbio/syw100.
- Morgan-Richards, M., Bulgarella, M., Sivyler, L., Dowle, E.J., Hale, M., McKean, N.E., Trewick, S.A. 2017. Explaining large mitochondrial sequence differences within a population sample. *Royal Society Open Science*. 4, 170730. DOI: 10.1098/rsos.170730.
- Morin, P.A., Martien, K.K., Taylor, B.L. 2009. Assessing statistical power of SNPs for population structure and conservation studies. *Molecular Ecology Resources*. 9, 66–73. DOI: 10.1111/j.1755-0998.2008.02392.x.
- 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. *Molecular Ecology Resources*. 17, 1136–1147. DOI: 10.1111/1755-0998.12654.
- Nosil, P. 2008. Speciation with gene flow could be common. *Molecular Ecology*. 17, 2103–2106. DOI: 10.1111/j.1365-294X.2008.03715.x.

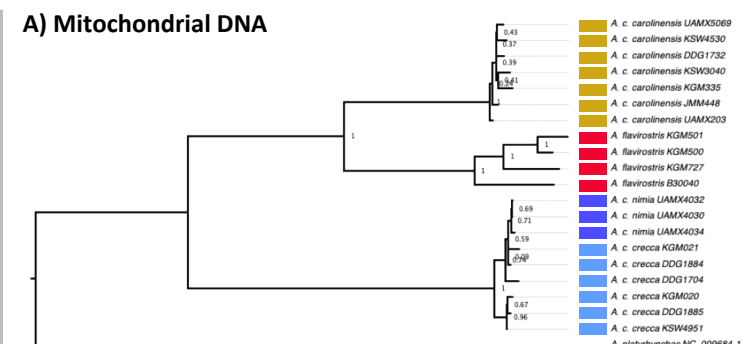
- Nosil, P., Egan, S.P., Funk, D.J. 2008. Heterogeneous Genomic Differentiation Between Walking-Stick Ecotypes. *Evolution*. 62, 316–336. DOI: 10.1111/j.1558-5646.2007.00299.x.
- Peñalba, J.V., Joseph, L., Moritz, C. 2019. Current geography masks dynamic history of gene flow during speciation in northern Australian birds. *Molecular Ecology*. 28, 630–643. DOI: 10.1111/mec.14978.
- Peters, J.L., McCracken, K.G., Pruett, C.L., Rohwer, S., Drovetski, S.V., Zhuravlev, Y.N., Kulikova, I., Gibson, D.D., Winker K. 2012. A parapatric propensity for breeding precludes the completion of speciation in common teal (*Anas crecca*, sensu lato). *Molecular Ecology*. 21, 4563–4577. DOI: 10.1111/j.1365-294X.2012.05711.x.
- Peters, J.L., Winker, K., Millam, K.C., Lavretsky, P., Kulikova, I., Wilson, R.E., Zhuravlev, Y.N., McCracken, K.G. 2014. Mitochondrial discord in six congeneric lineages of Holarctic ducks (genus *Anas*). *Molecular Ecology*. 23, 2961–2974. DOI: 10.1111/mec.12799.
- Peters, J.L., Zhuravlev, Y., Fefelov, I., Logie, A., Omland, K.E. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas spp.*) *Evolution*. 61, 1992–2006. DOI: 10.1111/j.1558-5646.2007.00149.x.
- Picard toolkit. 2019. Broad Institute. <http://broadinstitute.github.io/picard/>
- Price, T. 2008. *Speciation in birds*. CO: Roberts and Company Publishers.
- QIAGEN®. 2006. *DNeasy Blood & Tissue Handbook*. 62.
- Rambaut, A. 2006. *FigTree (v.1.4.2)*. Institute of evolutionary biology, University of Edinburgh. <https://github.com/rambaut/figtree/releases>
- Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*. 67, 901–904. DOI: 10.1093/sysbio/syy032.
- Remsen, J. V., Jr., Areta, J. I., Bonaccorso, E., Claramunt, S., Jaramillo, A., Lane, D. F., Pacheco, J. F., Robbins, M. B., Stiles, F. G., and Zimmer, K. J. 2022. A classification of the bird species of South America. Version 4 April 2022. American Ornithological Society. <http://www.museum.lsu.edu/~Remsen/SACCBaseline.htm>
- Renaudier, A. 2010. Rare birds in French Guiana in 2005–07. *Contiga* 32, 95-104.
- Rheindt, F.E., Edwards, S.V. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *The Auk*. 128, 620–632. DOI: 10.1525/auk.2011.128.4.620.
- Rice, W.R., Hostert, E.E. 1993. Laboratory experiments on speciation: what have we learned in 40 years? *Evolution*. 47, 1637–1653. DOI: 10.1111/j.1558-5646.1993.tb01257.x.
- Rougemont, Q., Gagnaire, P.A., Perrier, C., Genthon, C., Besnard, A.L., Launey, S., Evanno, G. 2017. Inferring the demographic history underlying parallel genomic divergence among pairs of parasitic and nonparasitic lamprey ecotypes. *Molecular Ecology*. 26, 142–162. DOI: 10.1111/mec.13664.
- Rubinoff, D., Holland, B.S. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic Inference. *Systematic Biology*. 54, 952–961. DOI: 10.1080/10635150500234674.

- Ruegg, K., Anderson, E.C., Slabbekoorn, H. 2012. Differences in timing of migration and response to sexual signaling drive asymmetric hybridization across a migratory divide. *Journal of Evolutionary Biology*. 25, 1741–1750. DOI: 10.1111/j.1420-9101.2012.02554.x.
- Rundell, R.J. 2008. Cryptic diversity, molecular phylogeny and biogeography of the rock- and leaf litter-dwelling land snails of Belau (Republic of Palau, Oceania). *Philosophical Transactions of the Royal Society B: Biological Sciences*. 363, 3401–3412. DOI: 10.1098/rstb.2008.0110.
- Rundle, H.D., Nosil, P. 2005. Ecological speciation. *Ecology Letters*. 8, 336–352. DOI: 10.1111/j.1461-0248.2004.00715.x.
- Sæther, B.E., Lande, R., Engen, S., Weimerskirch, H., Lillegård, M., Altwegg, R., ... Visser, M. E. 2005. Generation time and temporal scaling of bird population dynamics. *Nature*. 436, 99–102. DOI: 10.1038/nature03666.
- Schluter, D. 2001. Ecology and the origin of species. *Trends in Ecology & Evolution*. 16, 372–380. DOI: 10.1016/s0169-5347(01)02198-x.
- Schluter, D. 2009. Evidence for ecological speciation and its alternative. *Science*. 323, 737–741. DOI: 10.1126/science.1160006.
- Schluter, D., Nagel, L. 1995. Parallel speciation by natural selection. *The American Naturalist*, 146, 292–301. DOI: 10.1086/285799.
- Scott, D.A., Carbonell, M. 1986. *A Directory of Neotropical Wetlands*. IUCN Conservation Monitoring Centre.
- Silvestro, D., Michalak, I. 2012. raxmlGUI: a graphical front-end for RAxML. *Organisms Diversity & Evolution*. 12, 335–337. DOI: 10.1007/s13127-011-0056-0.
- Smith, H. M. 1955. The perspective of species. *Turtox News*, 33, 4–77.
- Springer, M. S., and J. Gatesy, J. 2018. Delimiting coalescence genes (c-genes) in phylgenomic data sets. *Genes*. 9, 123. DOI: 10.3390/genes9030123
- Uy, J.A., Irwin, D.E., Webster, M.S. 2018. Behavioral isolation and incipient speciation in birds. *Annual Review of Ecology Evolution and Systematics*. 49, 1–24. DOI: 10.1146/annurev-ecolsys-110617-062646.
- Van Dam, M.H., Henderson, J.B., Esposito, L., Trautwein, M. 2021. Genomic characterization and curation of UCEs improves species tree reconstruction. *Systematic Biology*. 70, 307–321. DOI: 10.1093/sysbio/syaa063.
- Verzijden, M.N., ten Cate, C., Servedio, M.R., Kozak, G.M., Boughman, J.W., Svensson, E.I. 2012. The impact of learning on sexual selection and speciation. *Trends in Ecology & Evolution*. 27, 511–519. DOI: 10.1016/j.tree.2012.05.007.
- Winker, K. 2010. On the origin of species through heteropatric differentiation: a review and a model of speciation in migratory animals. *Ornithological Monographs*. 69, 1–30. DOI: 10.1525/om.2010.69.1.1.
- Winker, K. 2021. An overview of speciation and species limits in birds. *Ornithology*. 138, 1–27. DOI: 10.1093/ornithology/ukab006.
- 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. DOI: 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* spp.): evidence from 2,517 ultraconserved element loci. *The Auk*. 136, ukz012. DOI: 10.1093/auk/ukz012.

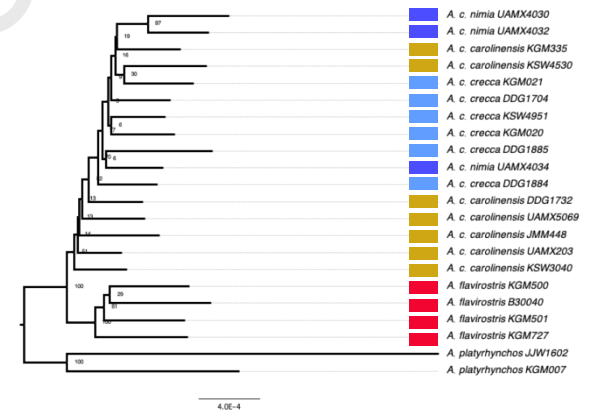
- Winker, K., McCracken, K.G., Gibson, D.D., Peters, J.L. 2013. Heteropatric speciation in a duck, *Anas crecca*. *Molecular Ecology*. 22, 5922–5935. DOI: 10.1111/mec.12525.
- Winkler, D.W., Gandoy, F.A. Areta, J.I., Iliff, M.J., Rakhimberdiev, E., Kardynal, K.J., Hobson, K.A. 2017. Long-distance range expansion and rapid adjustment of migration in a newly established population of barn swallows breeding in Argentina. *Current Biology*. 27, 1080–1084. DOI: 10.1016/j.cub.2017.03.006.
- Withrow, J.J., Sealy, S.G, Winker, K. 2014. Genetics of divergence in the northern saw-whet owl (*Aegolius acadicus*). *The Auk*. 131, 73–85. DOI: 10.1642/AUK-13-187.1.
- 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. *Molecular Ecology*. 25, 5144–5157. DOI: 10.1111/mec.13813.
- Zink, R., Rohwer, S., Andreev, A., Dittmann, D. 1995. Trans-Beringia comparisons of mitochondrial-DNA differentiation in birds. *Condor*. 97, 639–649. DOI: 10.2307/1369173.
- Zink, R.M., Barrowclough, G.F. 2008. Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*. 17, 2107–2121. DOI: 10.1111/j.1365-294X.2008.03737.

Author Statement

Fern Spaulding: Formal analysis, Investigation, Writing - Original Draft, Writing – Review & Editing; **Jessica F. McLaughlin:** Writing - Review & Editing; **Rebecca G. Cheek:** Investigation, Writing - Review & Editing; **Kevin G. McCracken:** Conceptualization, Writing - Review & Editing; **Travis C. Glenn:** Conceptualization, Resources; **Kevin Winker:** Conceptualization, Writing - Review & Editing, Supervision



Journal Pre-proofs



Highlights

- Resolved uncertainty in the green-winged teal and the yellow-billed teal complex.
- Ultraconserved elements were used to resolve phylogenetic relationships.
- Mitogenomic and nuclear DNA showed discordant phylogenetic relationships.
- Divergence with gene flow was present in each pairwise demographic analyses.
- Heteropatric, parapatric, and allopatric modes of speciation are likely to be involved.