



Divergence, gene flow, and speciation in eight lineages of trans-Beringian birds

Jessica F. McLaughlin^{1,2} | Brant C. Faircloth³ | Travis C. Glenn⁴ | Kevin Winker¹ ¹University of Alaska Museum, Fairbanks, AK, USA²Sam Noble Oklahoma Museum of Natural History, Norman, OK, USA³Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA, USA⁴Department of Environmental Health Science and Institute of Bioinformatics, University of Georgia, Athens, GA, USA**Correspondence**

Kevin Winker, University of Alaska Museum, 907 Yukon Drive, Fairbanks, AK 99775, USA. Email: kevin.winker@alaska.edu

Funding information

National Science Foundation, Grant/Award Number: DEB-1242267-1242241-1242260

Abstract

Determining how genetic diversity is structured between populations that span the divergence continuum from populations to biological species is key to understanding the generation and maintenance of biodiversity. We investigated genetic divergence and gene flow in eight lineages of birds with a trans-Beringian distribution, where Asian and North American populations have likely been split and reunited through multiple Pleistocene glacial cycles. Our study transects the speciation process, including eight pairwise comparisons in three orders (ducks, shorebirds and passerines) at population, subspecies and species levels. Using ultraconserved elements (UCEs), we found that these lineages represent conditions from slightly differentiated populations to full biological species. Although allopatric speciation is considered the predominant mode of divergence in birds, all of our best divergence models included gene flow, supporting speciation with gene flow as the predominant mode in Beringia. In our eight lineages, three were best described by a split-migration model (divergence with gene flow), three best fit a secondary contact scenario (isolation followed by gene flow), and two showed support for both models. The lineages were not evenly distributed across a divergence space defined by gene flow (M) and differentiation (F_{ST}), instead forming two discontinuous groups: one with relatively shallow divergence, no fixed single nucleotide polymorphisms (SNPs), and high rates of gene flow between populations; and the second with relatively deeply divergent lineages, multiple fixed SNPs, and low gene flow. Our results highlight the important role that gene flow plays in avian divergence in Beringia.

KEYWORDS

Beringia, Pleistocene glacial cycles, population genomics, speciation with gene flow, UCEs, ultraconserved elements

1 | INTRODUCTION

The processes of population divergence and speciation are among the most important in generating biodiversity. The genomic differentiation that accompanies population divergence is influenced by multiple mechanisms, including selection, mutation, and drift. However, gene flow might also play an important role, because it can counteract differentiation from selection and drift when it occurs

during divergence or on secondary contact (Dobzhansky, 1937; Price, 2008; Seehausen et al., 2014; Wright, 1931), leaving recognizable signatures across the genome that differ from divergence between strictly allopatric populations (Feder, Egan, & Nosil, 2012; Sousa & Hey, 2013). As a result, understanding gene flow is vital to determining how divergence and speciation operate in populations that may not be strictly isolated. Further, ascertaining whether there are common genomic patterns in how populations diverge,

particularly in a shared geographic system, can reveal how divergence leads to speciation.

Beringia extends from northeastern Asia across the Bering and Chukchi seas and into North America across Alaska into western Canada (Figure 1). Through the Pleistocene (2.6 million - 10,000 years ago), central Beringia experienced multiple cycles of exposure and inundation from sea level changes driven by glacial cycles, causing intermittent periods of terrestrial connectivity between Eurasia and North America, followed by isolation (Figure 1; Hopkins, 1959; Hopkins, McNeil, Merklin, & Petrov, 1965). This episodic cycling, estimated to have occurred at least nine and possibly up to twenty times or more during the Pleistocene (Hopkins, 1967; Pielou, 2008), repeatedly connected and disconnected the biota of Eurasia and North America. This natural, long-term experiment could cause varying estimates of timing and degrees of divergence between populations occurring on both sides of Beringia (trans-Beringian taxa), because different population pairs will likely have split during different flooding events throughout the past 2.6 million years. Mobile taxa distributed across this region, such as birds, are among those most likely to reunite during favourable periods and experience novel, renewed, or increased gene flow. Although allopatric speciation is widely considered to be the predominant and sometimes singular mode of speciation in birds (Mayr, 1963, 2004; Price, 2008), genomic data increasingly show cases among multiple taxa that do not fit this model (Mallet, Besansky, & Hahn, 2016; Morales, Jackson, Dewey, O'Meara, & Carstens, 2017; Peñalba, Joseph, & Moritz, 2019; Rheindt & Edwards, 2011; Zarza et al., 2016). Beringia's Pleistocene biotic history has been revealed through phylogeographic studies in many taxa, but until the advent of large-scale genomic data sets, few studies have been able to

focus on the role of gene flow in regional divergence processes (e.g. DeChaine, 2008; Galbreath, Cook, Eddingsaas, & DeChaine, 2011; Geml, Laursen, O'Neill, Nusbaum, & Taylor, 2005; McLean, Jackson, & Cook, 2016; Peters et al., 2012, 2014). The trans-Beringian region provides a natural system in which to test the hypothesis that allopatric speciation predominates and to study the divergence and speciation process given repeated opportunities for gene flow.

Here, we study the speciation process across Beringia by making pairwise comparisons within diverging lineages at the population, subspecies, and full species levels (as taxonomically designated) in three avian orders: Anseriformes, Charadriiformes, and Passeriformes. Within each order, we examined two or three paired divergence events at the levels of population, subspecies, and species, for a total of eight lineages and divergence events (Figure 2). Our goal was to sample the full speciation process and to make contrasts among lineages. All of these lineages but one undertake seasonal movements and migrations of varying distances, ensuring that they are sufficiently mobile to respond to the intermittent vicariant barrier of the Bering and Chukchi seas. Overall, we ask how divergence and speciation have developed among these lineages across the shared geographic region of Beringia, in which temporal changes in connectivity between Asian and North American populations have likely facilitated both divergence and, possibly, gene flow. We determine the location of each divergence event on the continuum of divergence space (defined by rates of gene flow vs. a measure of population differentiation) and whether there are common patterns in the demographic histories shared among lineages, such as in modes of divergence (e.g. speciation with gene flow, strict isolation, isolation followed by secondary contact, or gene flow with population growth). Our results provide a robust among-lineage contrast

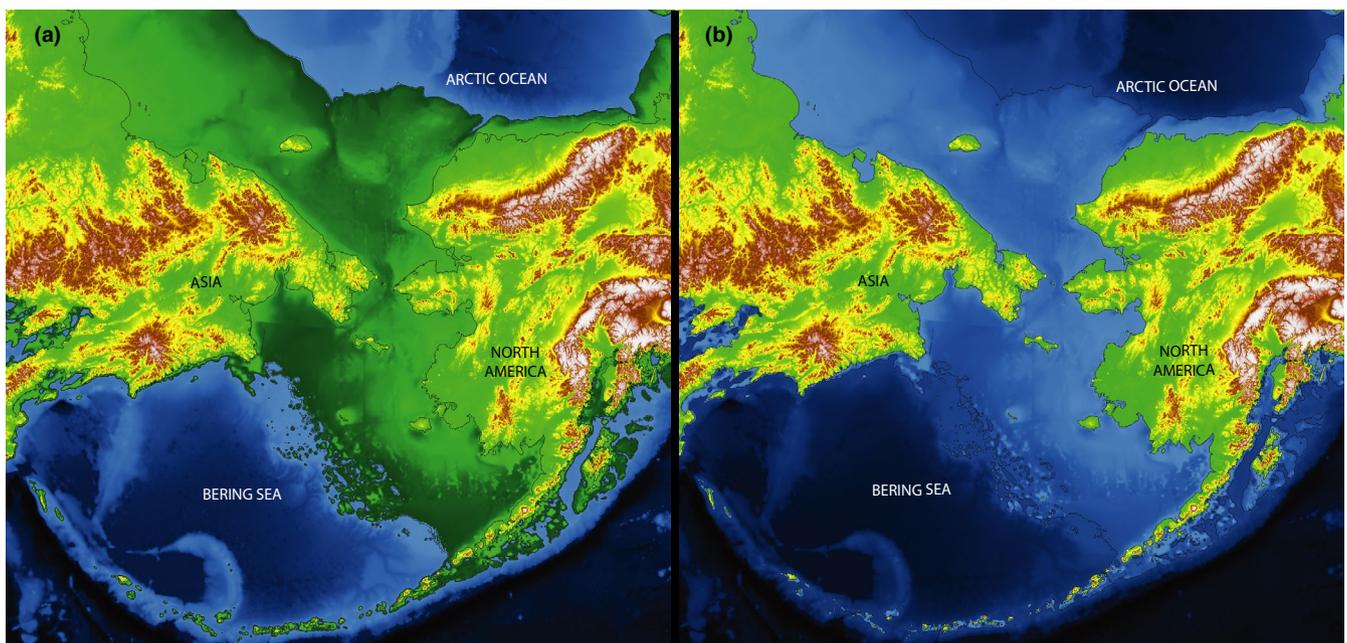


FIGURE 1 Beringia, centred on the Bering Sea, extends into northeastern Asia and northwestern North America (from Manley, 2002). During glacial maxima in the Pleistocene (a), a land bridge existed between the continents (here ~20 Kya), but it was broken by rising seas during interglacials (b) as at present [Colour figure can be viewed at wileyonlinelibrary.com]

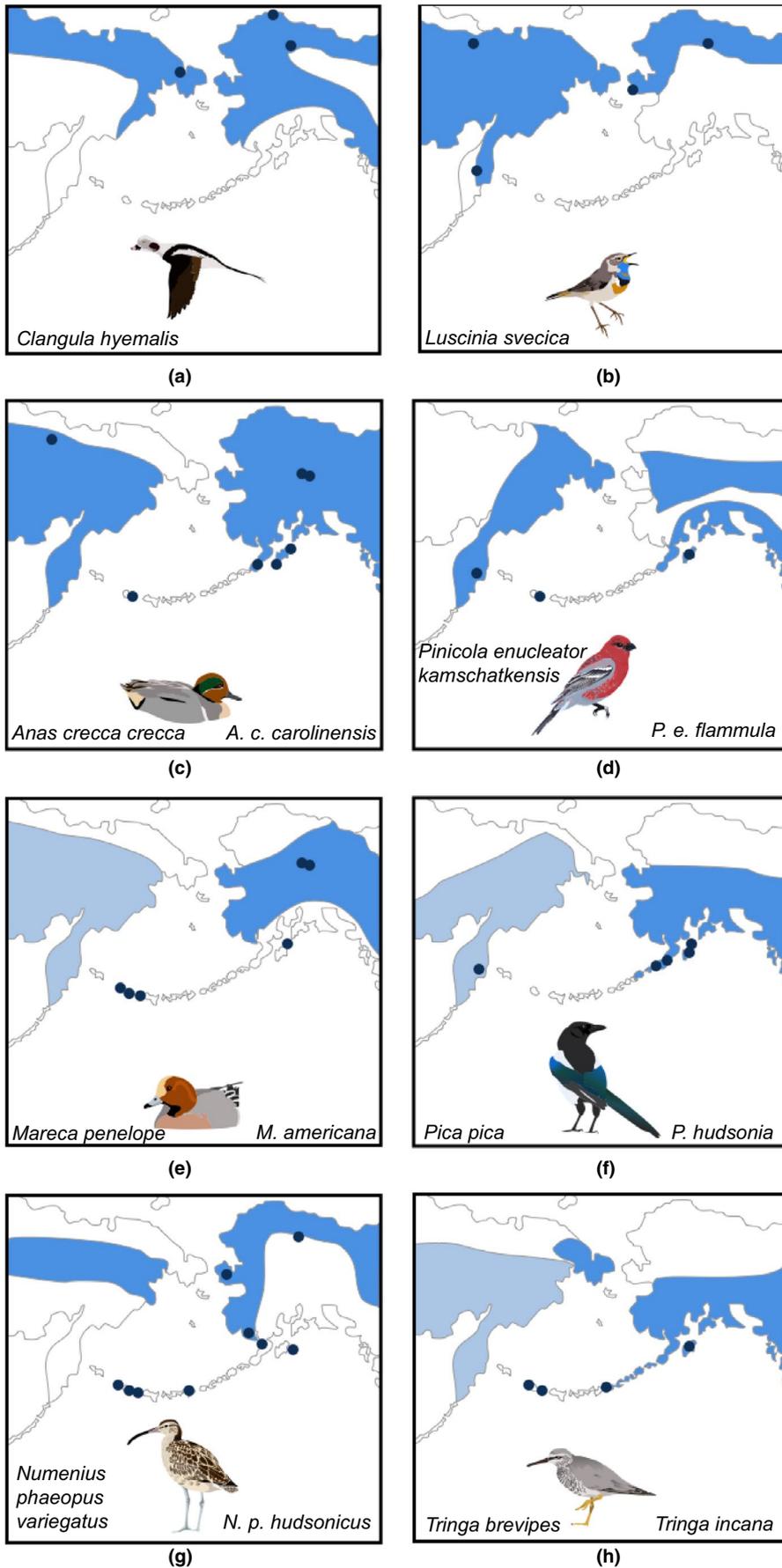


FIGURE 2 The eight lineages in this study, with two at population (a, b), three at subspecies (c, d, g) and three at species (e, f, h) taxonomic levels (different shades of blue represent different species' ranges). The three avian orders represented are Anseriformes (a, c, e), Charadriiformes (g, h), and Passeriformes (b, d, f). Each of the eight comprises a pairwise comparison between Asian and North American representatives (taxonomically named in lower panels except for populations, which are taxonomically identical). Black dots indicate sampling locations; for full details, see Table S1. Specimens taken away from breeding grounds (e.g. Aleutian Islands) were assigned to continental breeding taxon by plumage characteristics [Colour figure can be viewed at wileyonlinelibrary.com]

of divergence patterns in birds across Beringia and show that divergence with gene flow has been the norm across this region, rather than the exception.

2 | METHODS

2.1 | Study design

The eight lineages of birds we studied were selected based on the most closely related trans-Beringian lineages with adequate modern museum specimens (i.e. those with high-quality tissue samples), such that, collectively, we could study what was likely to be the full speciation process between continental populations on each side of the Beringian vicariant barrier. Each lineage represents a pair of Asian and North American populations, subspecies, or species (i.e. trans-Beringian). These continental representatives from northeast Asia and northwestern North America were (Figure 2), by taxonomic level, populations of *Clangula hyemalis* (long-tailed duck; Figure 2a) and *Luscinia svecica* (bluethroat; Figure 2b); subspecies *Anas crecca crecca*/*Anas crecca carolinensis* (green-winged teal; Figure 2c), *Numenius phaeopus variegatus*/*Numenius phaeopus hudsonicus* (whimbrel; Figure 2g), and *Pinicola enucleator kamschatkensis*/*Pinicola enucleator flammula* (pine grosbeak; Figure 2d); and species *Mareca penelope*/*Mareca americana* (Eurasian and American wigeons; Figure 2e), *Tringa brevipes*/*Tringa incana* (grey-tailed and wandering tattlers; Figure 2h), and *Pica pica*/*Pica hudsonia* (Eurasian and black-billed magpies; Figure 2f). Among these lineages, only the magpies (*Pica* spp.) are entirely sedentary, with no regular seasonal movements or migrations. Although prior work suggests (Humphries & Winker, 2011) that these taxonomic levels were not well correlated with genetic divergence in each lineage, these classifications provided reasonable assurance that our study includes the full speciation process in birds in this region. This reasoning follows the observation that although genetic divergence estimates are only loosely correlated with speciation and do not show threshold values signifying its completion (e.g. Winker, 2009), nongenetic data implying that reproductive isolation exists between species pairs result in major avian taxonomies agreeing on the vast majority of species-level taxa (e.g. Clements et al., 2018; Dickinson, Remsen, & Christidis, 2013). Following Humphries and Winker (2011), we used current taxonomic designations as a surrogate for phenotypic divergence.

2.2 | Molecular approach

When comparing the genomes of diverging populations, different portions will have different levels of divergence, depending on a variety of factors, including differing inheritance patterns (Avise, 2004; Funk & Omland, 2003; Toews & Brelsford, 2012), linkage with genes under selection (Casillas & Barbadilla, 2017; Feder et al., 2012; Katzman et al., 2007; Via & West, 2008; Wolf & Ellegren, 2017), the

structure and arrangement of the genome itself, including the effects of recombination (Delmore et al., 2015; Ragland et al., 2017; Vijay et al., 2017; Wolf & Ellegren, 2017), and demographic histories of the populations under consideration (Casillas & Barbadilla, 2017; Sousa & Hey, 2013). Although the general consensus is that a sufficient number of unlinked variable markers sampled from throughout the genome will produce similar parameter estimates (Beerli & Felsenstein, 1999; Carling & Brumfield, 2007; Kuhner, Yamato, & Felsenstein, 1998), divergence estimates made using markers from different subsamples of genomes are frequently discordant (Humphries & Winker, 2011; Peters et al., 2014) and can skew our understanding of the extent and history of gene flow (Cahill et al., 2015; Ellegren et al., 2012; Good, Vanderpool, Keeble, & Bi, 2015; Nosil, Funk, & Ortiz-Barrientos, 2009; Zarza et al., 2016).

We also expect some discord to exist between levels of genetic divergence and taxonomic assignments, because the latter break a continuous process into discrete bins that often rely on subjective attributes such as observed differences in phenotype. In addition, speciation is likely to be strongly influenced by selection, whereas general measures of genetic divergence are likely to be dominated by selectively neutral variation (Price, 2008; Winker, 2009). Striking phenotypic differences can result from a few changes in relatively small genomic regions (Hoekstra, Hirschmann, Bunday, Insel, & Crossland, 2006; Vickrey et al., 2018), and substantial levels of gene flow can occur between phenotypically distinct forms (Toews et al., 2016; Van Belleghem et al., 2017). As a result, estimates of divergence based on a limited subset of the genome might not correlate with, and could be discordant with, taxonomic assignment based on phenotypic variation. Larger-scale genomic data should provide a more holistic understanding of where diverging populations are in the process of speciation.

Comparing divergence histories (e.g. population sizes, timing of separation and whether and to what degree gene flow is occurring) among multiple diverging lineages in a common geographic framework is a powerful approach to understand both the patterns and the underlying processes of divergence (Campbell, Braile, & Winker, 2016; Peñalba et al., 2019; Rincon-Sandoval, Betancur-R, & Maldonado-Ocampo, 2019; Smith, Harvey, Faircloth, Glenn, & Brumfield, 2014). This comparative approach is most effective in a common molecular framework, when a large number of orthologous markers are used, because orthology facilitates comparability among data sets and direct comparisons of parameters among lineages (Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016; Smith et al., 2014). To ensure we were collecting orthologous loci and performing analyses using a common molecular framework, we chose to use ultraconserved elements (UCEs) as a genome-wide molecular marker. UCEs are loci-centred on highly conserved genomic regions distributed throughout animal genomes (Bejerano et al., 2004; Faircloth et al., 2012; Katzman et al., 2007; Siepel et al., 2005; Stephen, Pheasant, Makunin, & Mattick, 2008) and are putatively orthologous across a wide range of taxa, enabling inferences of population history even at relatively shallow levels of divergence (Faircloth et al., 2012; Giarla & Esselstyn, 2015; Harvey et al., 2016;

Smith et al., 2014; Winker, Glenn, & Faircloth, 2018; Winker, Glenn, Withrow, Sealy, & Faircloth, 2019).

This molecular comparative framework should also work well to make comparisons among diverging lineages in terms of their progression across a generically defined divergence space. Here, we use rates of gene flow (M) versus F_{ST} to represent this space, considering that decreased gene flow and increased genetic distance are associated with the speciation process and might be considered key aspects of the divergence and speciation continuum. Although we expect specific genetic changes that accompany or promote divergence and speciation to vary among even closely related lineages (e.g. Delmore et al., 2018; Irwin et al., 2018), theory and empirical evidence indicate that broad similarities also occur among genetic parameters that are widely applicable among populations across eukaryotes (e.g. estimates of gene flow and genetic distance; Price, 2008; Wright, 1931). Other, similar studies have compared lineage divergences across similar divergence space among groups of lineages that are similarly and even more broadly related on evolutionary temporal scales, for example, among *Timema* stick insects (>30 Ma; Riesch et al., 2017) and more widely among animal taxa (across ~525 Ma of evolution; Roux et al., 2016). Among all of our avian lineages, all of our pairwise contrasts together have a most recent common ancestor ~85–88 Ma (Claramunt & Cracraft, 2015; Jarvis et al., 2014).

An ancillary objective of our study was to contrast UCE data with corresponding historic data (mitochondrial DNA [mtDNA] and amplified fragment length polymorphisms [AFLPs], including most of the same individuals) of Humphries and Winker (2011). Placing UCE results in the context of other genetic markers will also enhance our understanding of divergence processes by clarifying how historic studies compare in their inferences.

2.3 | Laboratory procedures

We extracted DNA from muscle tissue of vouchered, archived museum specimens for all taxa (Table S1). We aimed to extract whole genomic DNA from eight individuals per population in each lineage, but fewer specimens were available in some groups, so smaller sample sizes of 6–7 individuals per population were used in those cases (under our bioinformatics pipeline, these sample sizes exceed the suggested optimum for coalescent analyses; see below). We prepared double-indexed DNA libraries for each sample as described in Glenn et al. (2019), which enables a high level of multiplexing, and these were quantified with a Qubit Fluorometer (Invitrogen, Inc.) and combined into equimolar pools. We enriched the pools using the Tetrapods-UCE-5Kv1 kit (MYcroarray) for a set of 5,060 loci (Faircloth et al., 2012), using UCE enrichment protocol version 1.5 and postenrichment amplification protocol version 2.4 (ultraconserved.org) with HiFi HotStart polymerase (Kapa Biosystems) and 14 cycles of postenrichment PCR. We checked the quality of the libraries by quantifying the distribution of fragment size of the enriched pool on a Bioanalyzer (Agilent, Inc), and the

enriched pool was quantified by qPCR with a commercial kit (Kapa Biosystems). Samples from this project were pooled with others and run in four sequencing lanes (8–111 samples from this project per lane) using an Illumina HiSeq 2500 (UCLA Neuroscience Genomics Core).

2.4 | Bioinformatics

After sequencing, we demultiplexed the data with `BCL2FASTQ` (v 1.8.4; Illumina, Inc.) and removed adapters and low-quality bases trimmed using `ILLUMIPROCESSOR` (v 2.0.3; Faircloth, 2013), a parallel wrapper around `Trimmomatic` (v 0.32; Bolger, Lohse, & Usadel, 2014). The singleton and read 1 fastq files for each individual were combined and then, with read 2 files, were assembled with `Trinity` (v 2.0.6; Grabherr et al., 2011) run on `Galaxy` (Afgan et al., 2016). To recover UCEs and provide information on the median number of loci shared and unshared by individuals in each lineage's data set, UCE loci were extracted and a complete matrix was constructed for each lineage using `PHYLUCE` (v 1.5; Faircloth, 2016).

To create a reference, in each of the eight groups the two individuals in each population, subspecies, or species closest to the median number of loci were identified, and the fastq sequence files for the four individuals (per lineage) were combined to produce a single read 1 file and a single read 2 file. These sequences were then assembled with `Trinity` on `Galaxy`, as above, and `PHYLUCE` was used to create a fasta file of UCE loci. Pooling reads from multiple individuals was done to build a high-quality reference against which to call variants for all individuals, reasoning that such a reference will have greater accuracy because it was assembled from more reads and, given the entire data set, would strike a balance between retention and loss of loci due to quality control issues among individuals farther along the pipeline (Winker et al., 2019). This reference was then indexed with `BWA` and `SAMTOOLS` (v. 0.7.7 & v. 0.1.19; Li & Durbin, 2009; Li et al., 2009) for calling single nucleotide polymorphisms (SNPs).

Single nucleotide polymorphisms were called using a modified workflow for population genomics with UCEs developed by Michael Harvey and Brant Faircloth (https://github.com/mgharvey/seqcap_pop). For each lineage's data set (i.e. each pairwise comparison), sequences were aligned to the reference with `BWA-MEM` (Li, 2013) and converted to `BAM` with `SAMTOOLS` (Li et al., 2009). Alignments were checked for `BAM` format violations, read group header information was added, and PCR duplicates were marked for each individual using `PICARD` (v. 1.106; <http://broadinstitute.github.io/picard>). The resulting `BAM` files for each individual in a lineage were merged into a single file with `Picard`, which was then indexed with `SAMTOOLS`. The `GENOME ANALYSIS Toolkit` (`GATK`; v. 3.4-0; McKenna et al., 2010) was then used to locate and realign around indels, which was followed by calling SNPs using the `UnifiedGenotyper` tool in `GATK`. SNPs and indels were then annotated, and indels were masked. We then used `GATK` to restrict our data sets to high-quality SNPs (Q30 filter) and performed read-backed phasing. This process calls and phases both

alleles for each individual, achieving optimal sample sizes of eight haplotypes from just four diploid individuals; this sample size is considered optimal for coalescent-based and population genomic analyses, and in all cases, our sample sizes exceed this optimum (Felsenstein, 2005; Nazareno, Bemmels, Dick, & Lohmann, 2017).

We filtered to biallelic, phased SNPs with `vcftools` (v. 0.0.12; Danecek et al., 2011), reducing our data set to a complete matrix with a minimum genotype quality (GQ) of 10 (i.e. genotyped with 90% confidence or higher). To ensure invariant loci did not result from missing data, we used the GATK function `EMIT_ALL_CONFIDENT_SITES`, followed by filtering with `vcftools` to remove loci with inadequate data. This represents the full data set (i.e. prior to thinning to one SNP per locus or removal of Z-linked loci, as needed for our detailed demographic analyses outlined below).

2.5 | Analyses

`vcftools` was used to calculate coverage depths and both SNP- and locus-specific F_{ST} . To calculate other population statistics, we used the R package `adegenet` (v 3.2.2; Jombart & Ahmed, 2011; R Core Team & Others, 2013) after converting the data sets to the appropriate format with `PGDSPIDER` (v 2.0.9.1; Lischer & Excoffier, 2012). We tested for Hardy–Weinberg equilibrium in each population, calculated observed and expected heterozygosity, calculated F_{ST} (using the G test with 99 bootstraps), and determined assignment probabilities using discriminant analysis of principal components. We included calculated heterozygosity because when observed heterozygosity exceeds expected, it suggests that differentiated (e.g. isolated) populations are undergoing gene flow.

We used diffusion analysis for demographic inference ($\delta a\delta i$, ver. 1.7.0; Gutenkunst, Hernandez, Williamson, & Bustamante, 2009) to estimate parameters of effective population size for both Asian and North American populations (ν_A and ν_{NA} , respectively), migration (m), time since split (T) and θ , defined as $4N_{ref}\mu$, with N_{ref} defined as ancestral population size and μ as substitution rate per generation. From these, biologically meaningful values of effective population size (N_{NA} and N_A), migration in individuals per generation (M), time since split in years (t), and ancestral population size (N_{ref}) were calculated, following a modified version of the methods used in Winker et al. (2018). To prepare our data for $\delta a\delta i$, the phased SNPs were thinned with `vcftools` to 1 SNP per locus (to minimize effects of linkage, as recommended in the $\delta a\delta i$ user manual). Z-linked loci were then removed using a custom `BLASTN` (Zhang, Schwartz, Wagner, & Miller, 2000) script (https://github.com/jfmclaughlin92/beringia_scripts) that aligned loci to the Z chromosome of *Gallus gallus* (for Anseriformes and Charadriiformes; NCBI Annotation Release 103) or *Taeniopygia guttata* (for Passeriformes; NCBI Annotation Release 103). Z-linked loci were excluded from our demographic analyses (although they were included in other analyses) because sample population sex ratios affect allele frequency estimates and these loci have a different inheritance scalar from autosomal loci (Garrigan et al., 2007; Jorde et al., 2000). This is the one-SNP-per-locus data set used for

$\delta a\delta i$ analyses. The remaining SNP data were then converted into the joint site frequency spectrum format using a Perl script by Kun Wang (https://github.com/wk8910/bio_tools/blob/master/01.dadi_fsc/00.convertWithFSC/convert_vcf_to_dadi_input.pl). We then ran two-population models using $\delta a\delta i$.

`Adegenet` F_{ST} calculations (described above) indicated that all eight pairs contained genetically divergent groups (e.g. each lineage at the population, subspecies, and species levels was genetically divergent), and we tested these data for their fit against eight different two-population demographic models (Figure 3). We use the term “population” to refer broadly to each member of the pair of taxa being compared at population, subspecies, and species levels. These models included a “neutral,” no-divergence model (Figure 3a); a split (divergence) model in which one ancestral population splits into two followed by isolation and no migration (Figure 3b); a split-migration model in which one ancestral population splits into two with ongoing gene flow (migration) reflected in a single parameter that assumes gene flow is roughly equal in both directions (Figure 3c); a split-migration model in which two different levels of migration (asymmetric gene flow) are considered between populations (Figure 3d); a secondary contact model in which an ancestral population splits followed by a period of isolation (T_{sc}) and then subsequent migration occurs, estimated using either one migration parameter assuming roughly equal gene flow (Figure 3e; Rougemont et al. 2017) or two migration parameters assuming asymmetric gene flow (Figure 3f); a split model that assumes isolation and population growth (Figure 3g); and a split-migration model that assumes population growth and asymmetric gene flow (Figure 3h). Three of these models are included in the basic $\delta a\delta i$ `Demographics2D.py` module as “neutral” (Figure 3a), “splitmig” (Figure 3c) and “IM” (Figure 3h); scripts are available at https://github.com/jfmclaughlin92/beringia_scripts. Upper and lower bounds were optimized by running each model repeatedly until the highest maximum log composite likelihood value was observed consistently across multiple runs with the same parameter bounds without the parameter estimates obtained approaching the defined bounds. AIC was calculated for each model and, for every model with ΔAIC of ≤ 10 (Burnham & Anderson, 1998), run with 100 bootstrap replicate data sets (constructed with a Python script that resampled individuals with replacement; https://github.com/jfmclaughlin92/beringia_scripts) to estimate the 95% confidence intervals (CI) for each parameter.

To interpret the model parameter estimates in biological terms, we used methods outlined by Winker et al. (2018), obtaining estimates of substitutions per site per generation by `BLASTing` each reference fasta against the most closely related NCBI-available genome and using time since most recent common ancestor estimates from Claramunt and Cracraft (2015; Table S2). Generation time (G) was determined as $G = \alpha + (s/(1 - s))$, where α is the age at first breeding and s is annual adult survival, following Saether et al. (2005) (Table S2). Our estimated substitution rates are applied as constants to some of our $\delta a\delta i$ -derived demographic estimates, and for the formulae that use them, the estimates derived will be positively correlated. Lower substitution rates would cause estimated effective

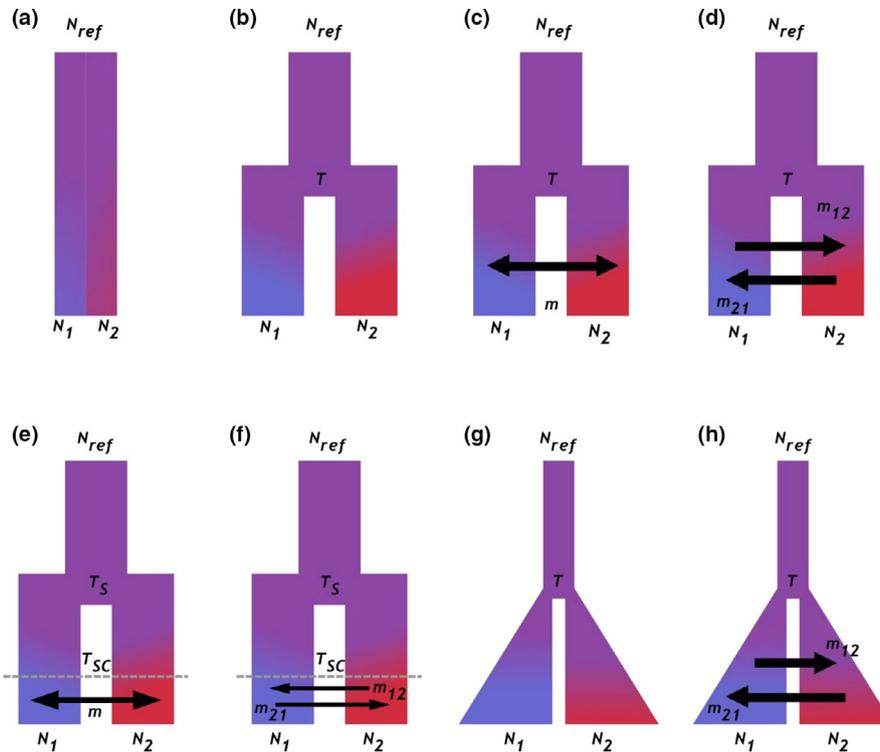


FIGURE 3 Models of divergence tested with $\delta a \delta i$ on two-population splits: (a) neutral model (no divergence), (b) split (divergence) followed by isolation and no migration (gene flow), (c) split-migration (“splitmig”) with unidirectional migration parameter (i.e. roughly similar levels of gene flow in both directions), (d) split-migration with two migration estimates (i.e. asymmetrical gene flow), (e) secondary contact (“SC”) with single migration parameter (i.e. divergence with intermittent gene flow), (f) secondary contact with two migration parameters (i.e. divergence with asymmetric intermittent gene flow), (g) split and isolation with population growth and no gene flow (“island”), and (h) split with migration (gene flow) and population growth (“IM”). Model sequence reflects underlying architecture and then increasing complexity within model family. Single arrows indicate models with gene flow in both directions being roughly equal; double arrows are models with different levels of gene flow in each direction. Models b and g are no-gene-flow models. Rectangles indicate unchanging population sizes; triangles indicate population growth. Colours suggest increasing population differentiation. Those models best fitting the data sets in this study included c, d, e, and f [Colour figure can be viewed at wileyonlinelibrary.com]

population sizes and split times to be lower (N_e : $nu1$, $nu2$, and N_{ref} and T). Gene flow estimates (m) are not affected by substitution rates. UCE-based substitution rates were considered in more depth by Winker et al. (2018), Winker et al. (2019).

We looked for evidence of expected relationships among lineages between divergence (F_{ST}) and key demographic estimates (gene flow and divergence time), using tests of nonlinear correlation for gene flow (which is expected to have a nonlinear relationship; Cabe & Alstad, 1994) and linear correlation for time. We also did a stepwise multiple regression, adding other variables possibly influencing divergence (F_{ST}) in this region, including θ , order, and current taxonomic placement. Taxonomic placement, a surrogate for phenotypic divergence, was converted to a quantitative variable by using 0 to represent population pairs, 0.5 for subspecies and 1 for species, following Humphries and Winker (2011). Because historic data using other marker types exist for these taxa and the population genomics properties of UCes are not yet well understood, we also examined relationships (including linear correlations) among divergence (F_{ST}) estimates made using UCes, mtDNA, and AFLPs (the latter two from Humphries & Winker, 2011).

3 | RESULTS

3.1 | Population genomics and divergence

We obtained >200 million reads, ranging from 379,344 to 4,010,381 (average = 1,450,760) per specimen, of which >99% passed adapter and quality control trimming. Assembly of the reference sequences from four individuals in each lineage pair produced between 130,506 and 657,330 contigs, totalling 47,215,417–254,336,867 bp. All data sets produced more than 1,000 contigs over 1 Kb (range 1,086–9,935; Table S3). We identified 4,040–4,294 UCE loci in each reference data set, with average contig length between 673 and 1,232 bp (Table S4). An average of 54.2% of loci was variable, and averages of 1.99–7.42 SNPs per locus were called. In total, 3,254–13,215 SNPs were called in each data set, and thinning to one SNP per locus left 1,636–2,656 SNPs in each lineage-specific data set (Table 1). Coverage across all SNPs averaged 35.1X, ranging between 30.4 X and 38.6 X (Table 1). Expected heterozygosity (H_e) ranged from 0.079 to 0.160, and observed heterozygosity (H_o) ranged between 0.086 and 0.179 (Table S5). In four lineages (*Numenius phaeopus* subsp.,

TABLE 1 Summary of single nucleotide polymorphisms (SNPs), including number of individuals in each lineage, total number of SNPs, average coverage per SNP, number of SNPs after thinning to one SNP per locus, and average number of SNPs per locus (variable loci only)

	N (Asia/North America)	Total SNPs	Average coverage depth per SNP	SNPs per locus (variable loci only)	Thinned SNPs (total variable loci)
Anseriformes					
<i>Clangula hyemalis</i>	8:8	9,276	30.76	3.798	2,442
<i>Anas crecca</i> subspp.	8:6	9,022	38.64	3.636	2,481
<i>Mareca penelope/Mareca americana</i>	8:8	7,041	37.61	3.041	2,315
Charadriiformes					
<i>Numenius phaeopus</i> subspp.	8:7	6,492	31.15	2.718	2,388
<i>Tringa brevipes/Tringa incana</i>	8:8	3,254	37.99	1.989	1,636
Passeriformes					
<i>Luscinia svecica</i>	8:7	9,379	38.06	3.728	2,516
<i>Pinicola enucleator</i> subspp.	8:7	8,117	36.48	3.056	2,656
<i>Pica pica/Pica hudsonia</i>	7:7	9,276	34.90	4.218	2,199

Pinicola enucleator subspp., *P. pica/Pi. hudsonia*, and *T. brevipes/T. incana*), there were significant differences between H_E and H_O , suggesting hybridization (Table S5).

In all eight lineage pairs, significantly nonzero F_{ST} values were found between Asian and North American populations, ranging between 0.004 and 0.58 (Table 2). Four lineages (*C. hyemalis*, *Anas crecca* subspp., *M. penelope/americana*, and *L. svecica*) had overall between-population F_{ST} values below 0.05 (lower-divergence group), whereas the four other lineages (*N. phaeopus* subspp., *T. brevipes/incana*, *P. enucleator* subspp., and *Pi. pica/hudsonia*) had F_{ST} values above 0.26 (higher-divergence group). In the lower- F_{ST} group (values <0.05; Table 2), there were no fixed SNPs in the full data sets (all SNPs, including Z-linked loci); in the higher- F_{ST} group, the number of fixed SNPs ranged from 12 to 121 in the one-SNP-per-locus data sets and between 31 and 299 in the full data sets (Table 2).

TABLE 2 Between-population F_{ST} values estimated using biallelic one-SNP-per-locus data sets and numbers of fixed SNPs from both full and thinned data sets

	F_{ST}	P-value	Number fixed loci (full data set)	Number fixed loci (1 SNP/locus data set)
Anseriformes				
<i>Clangula hyemalis</i>	0.0039	0.05*	0	0
<i>Anas crecca</i> subspp.	0.0191	0.01*	0	0
<i>Mareca penelope/Mareca americana</i>	0.0439	0.01*	0	0
Charadriiformes				
<i>Numenius phaeopus</i> subspp.	0.269	0.01*	31	12
<i>Tringa brevipes/Tringa incana</i>	0.585	0.01*	299	121
Passeriformes				
<i>Luscinia svecica</i>	0.0138	0.03*	0	0
<i>Pinicola enucleator</i> subspp.	0.442	0.01*	283	91
<i>Pica pica/Pica hudsonia</i>	0.328	0.01*	84	35

Note: P-values are from a G test run with 99 simulations. Asterisks (*) indicate significant differences.

3.2 | Demographic histories

The best-fit population demographic models varied among the eight lineages. None of our no-gene-flow models were supported (i.e. not Figure 3b or 3g). Five lineages had a single, unambiguously best-fitting gene-flow-present model (Table 3). In three of these—*N. phaeopus* subspp., *P. enucleator* subspp., and *Pi. pica/hudsonia*—split-migration with a single migration parameter was the model that best fit the data (Figure 3c; divergence with ongoing gene flow of roughly similar levels in both directions). In the other two lineages, secondary contact with a single migration parameter (*C. hyemalis*; Figure 3e; divergence with intermittent gene flow similar in both directions) or two migration parameters (*M. americana/penelope* Figure 3f; divergence with asymmetric intermittent gene flow) provided the best fit to the data (Table 3). Of the remaining three lineages, either a secondary contact model fit best (*T. brevipes/*

TABLE 3 Two-population model comparisons for each of our eight lineages

	Neutral (3A)	Split-migration, single migration parameter (3C)	Split-migration, two migration parameters (3D)	Secondary contact, single migration parameter (3E)	Secondary contact, two migration parameters (3F)	Isolation, pop. growth, no migration (3G)	Isolation, pop. growth, two migration parameters (3H)
Anseriformes							
<i>Clangula hyemalis</i>	-1,238.01 (2,480.02)	-245.87 (499.74)	-258.89 (527.78)	-216.46 (442.92)	-223.50 (459.00)	-2,841.08 (5,692.16)	n.a.
<i>Anas crecca</i> subspp.	-1,291.50 (2,587.00)	-245.40 (498.8)	-232.73* (475.46)	-253.67 (517.34)	-231.47 (474.94)	-1,593.81 (3,197.62)	n.a.
<i>Mareca penelope</i> / <i>Mareca americana</i>	-1,343.25 (2,690.50)	-281.05 (570.10)	-297.78 (605.56)	-268.91 (547.81)	-259.68 (531.36)	-2,157.45 (4,324.90)	-929.99 (1,869.98)
Charadriiformes							
<i>Numenius phaeopus</i> subspp.	-485.17 (974.34)	-99.12 (206.24)	-286.73 (583.46)	-280.17 (570.34)	-276.76 (565.52)	-207.77 (425.54)	-130.42 (270.84)
<i>Tringa brevipes</i> / <i>Tringa incana</i>	-5,643.88 (11,291.76)	-381.30 (770.60)	-374.01 (758.02)	-371.02* (752.04)	-369.73 (751.46)	-496.01 (1,002.02)	n.a.
Passeriformes							
<i>Luscinia svecica</i>	-1,074.99 (2,153.98)	-200.08 (408.16)	-202.18 (414.36)	-200.09* (410.18)	-228.59 (469.18)	-1,936.60 (3,883.2)	n.a.
<i>Pipicola enucleator</i> subspp.	-2,473.34 (4,950.68)	-126.56 (261.12)	-338.27 (686.54)	-338.92 (687.84)	-336.5 (685.00)	-151.61 (313.22)	n.a.
<i>Pica pica</i> / <i>Pica hudsonia</i>	-5,036.21 (10,076.4)	-167.41 (342.82)	-197.90 (405.8)	-196.73 (403.46)	-197.07 (406.14)	-256.83 (523.66)	n.a.

Note: Maximum log composite likelihood (MLCL) is averaged from the top five runs of the best optimized model presented for each, followed by AIC values in parentheses. Bold numbers indicate best-fit models, and asterisks those with a Δ AIC of <10. "n.a." indicates that we were unable to find a stable configuration of the model and it could not be run to completion. For the split-no-migration model (Figure 3b), all lineages returned n.a., so that model is not included here. The best-fitting models correspond with Figure 3c-f (labelled as 3c-f here).

TABLE 4 Results of two-population $\delta a\delta i$ analyses with all best-fit models (linked to Figure 3c-f by "3x" in parentheses after each model abbreviated in leftmost column)

	N_e Asian	N_e North American	N_{ref}	Time since split (years)	Time of secondary contact (years)	Migration (from ν_A)	Migration (from ν_{NA})
Anseriformes							
<i>Clangula hyemalis</i> SC, 1m (3E)	116,892 ($\pm 7,075$)	426,144 ($\pm 7,410$)	15,054 (± 491)	163,001 ($\pm 32,981$)	33,600 ($\pm 5,564$)	23.835 (± 0.430)	86.892 (± 1.569)
<i>Anas crecca</i> subspp. Splitmig, 2m (3D)	1,438,050 ($\pm 44,714$)	847,179 ($\pm 32,761$)	73,001 ($\pm 1,757$)	289,851 ($\pm 14,076$)		6.737 (± 0.417)	6.272 (± 0.372)
SC, 2m (3F)	1,615,683 ($\pm 39,884$)	585,290 ($\pm 38,181$)	70,305 ($\pm 1,967$)	277,048 ($\pm 12,993$)	21,092 ($\pm 10,979$)	23.305 (± 0.631)	18.350 (± 2.292)
<i>Mareca penelope/Mareca americana</i> SC, 2m (3F)	114,519 ($\pm 5,546$)	585,272 ($\pm 8,365$)	27,138 ($\pm 8,512$)	166,395 ($\pm 15,777$)	8,848 ($\pm 23,974$)	24.241 (± 0.481)	3.173 (± 0.562)
Charadriiformes							
<i>Numenius phaeopus</i> subspp. Splitmig, 1m (3C)	36,417 ($\pm 1,210$)	17,386 (± 602)	5,837 (± 208)	234,124 ($\pm 11,555$)		0.176 (± 0.009)	0.084 (± 0.005)
<i>Tringa brevipes/Tringa incana</i> SC, 1m (3E)	25,917 ($\pm 2,688$)	7,950 (± 830)	5,494 (± 589)	166,514 ($\pm 19,211$)	857 ($\pm 19,471$)	0.886 (± 0.492)	0.272 (± 0.151)
SC, 2m (3F)	26,091 ($\pm 3,032$)	7,855 ($\pm 1,053$)	5,471 (± 578)	164,741 ($\pm 25,939$)	2,224 ($\pm 32,712$)	0.193 (± 0.385)	0.131 (± 0.079)
Passeriformes							
<i>Luscinia svecica</i> Splitmig, 1m (3C)	1,034,459 ($\pm 17,246$)	245,352 ($\pm 8,977$)	47,233 ($\pm 1,826$)	245,352 ($\pm 13,964$)		21.458 (± 0.359)	3.779 (± 0.063)
SC, 1m (3E)	1,018,472 ($\pm 16,643$)	181,380 ($\pm 8,600$)	47,358 ($\pm 1,888$)	378,446 ($\pm 33,984$)	78,443 ($\pm 7,431$)	9.460 (± 0.862)	1.685 (± 0.154)
<i>Pipicola enucleator</i> subspp. Splitmig, 1m (3C)	38,867 ($\pm 1,544$)	54,734 ($\pm 1,736$)	21,617 (± 659)	172,935 ($\pm 5,693$)		0.0071 (± 0.0009)	0.0100 (± 0.0012)
<i>Pica pica/Pica hudsonia</i> Splitmig, 1m (3C)	176,999 (6,588)	67,474 ($\pm 1,989$)	24,928 ($\pm 1,458$)	417,030 ($\pm 14,655$)		0.0488 (± 0.0047)	0.0186 (± 0.0018)

Note: Included are estimates of effective population sizes (N_e) of North American and Asian populations, effective population size of the ancestral population (N_{ref}), time since divergence in years and, for applicable models, time of secondary contact, and migration rates using population sizes of Asian (ν_A) and North American (ν_{NA}) populations. For lineages in which multiple models fit, both models are presented in separate rows, with the best-fit model in bold (splitmig = split migration; SC = secondary contact model; see Figure 3). All estimates include 95% CI calculated with 100 $\delta a\delta i$ bootstraps.

incana; Figure 3e,f; Table 3) or there was similar support between split-migration and secondary contact models (*L. svecica*, Figure 3c,e; *A. crecca* subspp., Figure 3d,f; Table 3). Further calculations of demographic parameters (below) represent the very best-fitting model for each (bold in Table 3). Among the best-fit models, N_{ref} (ancestral population size) was estimated to be from 5,471 (*Tringa* spp.) to 70,305 (*A. crecca*) individuals (Table 4).

Estimates of effective population sizes in Asian populations ranged from 26,091 (*T. brevipes*) to 1,615,683 (*A. c. crecca*) individuals, whereas those of North American populations ranged from 7,855 (*T. incana*) to 585,290 (*A. c. carolinensis*; Table 4) individuals. In six lineages, the Asian population was markedly larger than the North American population: *A. c. carolinensis*, *M. penelope*, *N. phaeopus*, *T. brevipes*, *L. svecica*, and *Pi. pica* (Table 4). Estimates of time since divergence ranged from ~163,000 (*C. hyemalis*) to 417,000 (*Pi. pica/hudsonia*) years, with secondary contact estimated to have occurred between 2,224 (*Tringa* spp.) and 33,600 (*C. hyemalis*) years ago in the lineages for which secondary contact models fit the best (Table 4). Overall, estimates of migration (M) varied from ~0.01 (*P. enucleator* subspp.) to 86.9 (*C. hyemalis*) individuals per generation (Table 4). However, all four of the lower- F_{ST} lineages had gene flow rates of more than 3.17 individuals/generation (*M. penelope/ americana*), whereas all four higher- F_{ST} lineages had far lower gene flow rates (<0.193 birds/generation; Table 4). Significant differences between expected and observed heterozygosities, suggesting hybridization, corresponded exactly with this second group of lineages (contrasting Table 2 lineages having $F_{ST} > 0.26$ with those in Table S5 having significant differences between H_E and H_O).

3.3 | Contrasts among lineages and with historic data

There was a significant correlation between UCE-based estimates of M and F_{ST} following an exponential decay function ($p = .005$, Figure 4). Time since divergence (T) was also correlated with F_{ST} ($p = .01$), and a stepwise multiple regression including these variables and θ , order, and current taxonomic placement created a single-variable model (selecting only T) with equivalent significance and explanatory power. We also examined correlations between marker classes (our UCE data vs. historic mtDNA and AFLP data from Humphries & Winker, 2011), finding no significant relationships between any of the estimates of F_{ST} from UCES, mtDNA, and AFLPS. In general, the highest divergence estimates were found in mtDNA, followed by UCES and then AFLPS, with the lineages that lacked significantly nonzero estimates of F_{ST} in mtDNA and/or AFLPS having significantly nonzero F_{ST} from UCE data (Figure 5).

4 | DISCUSSION

4.1 | Trans-Beringian speciation

Gene flow has been an integral part of the divergence process among the Beringian lineages in our study. Although no single divergence-with-gene-flow model was shared by all lineages, gene flow was present in every best-fit model. The best-fit models included both split-migration and secondary contact models (i.e. divergence

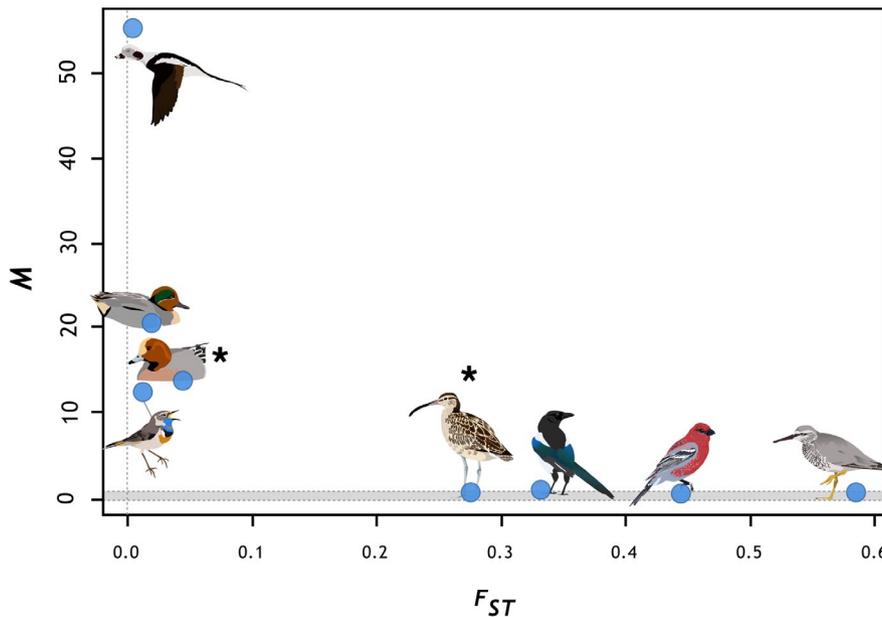


FIGURE 4 Using ultraconserved element-based estimates of F_{ST} versus average migration (gene flow) rate in individuals per generation (M) in our eight two-population lineages, using the best-fit model results. This relationship is significant (as an exponential decay function) and demonstrates a noncontinuous distribution among these lineages in Beringia in this divergence space. Two groups are apparent, one with low divergence and relatively high gene flow on the left, and one of higher divergence and low gene flow on the right. The dotted vertical line indicates $F_{ST} = 0$, and the horizontal grey band indicates M from 0 to 1 individual per generation. Asterisks indicate two lineages that might be taxonomically miscategorized at present, given opportunities for gene flow and the amounts occurring [Colour figure can be viewed at wileyonlinelibrary.com]

with ongoing or intermittent gene flow; Figure 3c–f). We examined eight lineages from three avian orders that have different life histories, seasonal migration behaviours, dispersal abilities, habitat requirements and, possibly (e.g. Table 4), Beringian occupation times. Any of these factors could influence how each lineage responded to the region's glacial–interglacial cycles of connection and isolation. Habitat availability would have varied considerably across Beringia spatially and temporally between glacial–interglacial cycles throughout the Pleistocene (Melles et al., 2012). This would have created a changing mosaic of habitat types during the past 2.6 million years, increasing opportunities for population connectivity in some lineages and likely influencing the observed variations in the levels of gene flow and histories of isolation. In this complex milieu, the detailed histories of avian lineages sharing the landscape are revealed by our UCE data sets to also be diverse. A commonality is that gene flow has apparently been ubiquitous, albeit occurring at low levels in some cases. Thus, in Beringia, avian divergence and speciation do not seem to be happening in a classic allopatric framework (contra Mayr, 1963, 2004).

4.2 | Discontinuous divergence among lineages

These Beringian lineages encompass a range of taxonomically recognized levels of divergence, from slightly divergent populations (*C. hyemalis*, *L. svecica*), to well-diverged populations designated as subspecies (e.g. *N. phaeopus variegatus/hudsonicus* and *P. enucleator kamschatkensis/flammula*), to full biological species (e.g. *T. brevipes/incana* and *Pi. pica/hudsonia*; Table 2). Overall, divergence (F_{ST}) was correlated with time since divergence and diminished gene flow, as expected. However, these lineages were not distributed across a smooth continuum in terms of our measures of gene flow (M) and divergence (F_{ST}). Instead, they clustered into two broad groups: (a) a lower-divergence group ($F_{ST} = 0.004$ – 0.044) with moderate levels of gene flow (mean $M = 25.6$ [range 3.2–86.9] individuals/generation averaging best-fit model estimates from Table 4); and (b) a

higher-divergence group ($F_{ST} = 0.269$ – 0.585) with sharply decreased levels of gene flow (mean $M = 0.08$ [range 0.007–0.19] individuals/generation). These groups did not correspond with the three taxonomic levels represented in the study, nor with the estimates of F_{ST} from other marker types, highlighting the heterogeneous nature of divergence (Figures 4 and 5).

Previous works (Flaxman, Wacholder, Feder, & Nosil, 2014; Hendry, Bolnick, Berner, & Peichel, 2009; Nosil, Feder, Flaxman, & Gompert, 2017; Riesch et al., 2017; Roux et al., 2016; Yamaguchi & Iwasa, 2016) have suggested that the speciation process can be a two-state system, with most diverging populations clustering bimodally near the two ends of the continuum (either showing panmixia/small differences or full reproductive isolation) and few populations existing in the middle ground. Furthermore, periods of gene flow can promote the formation of a dynamic like this (Flaxman et al., 2014; Nosil et al., 2017; Riesch et al., 2017). When gene flow occurs, the feedback process of divergent selection and linkage disequilibrium on the background of genomic architecture can return populations that have begun to diverge to a single well-mixed population (reticulation), unless a critical level of differentiation has already been achieved (Flaxman et al., 2014). Given the cyclical nature of population isolation and connectivity in Beringia, such a bimodal pattern might be more likely to develop, rather than forming stable states near the middle of the divergence continuum (Flaxman et al., 2014), because each round of connectivity following a period of isolation has the potential to reset the divergence clock in population pairs that have not differentiated enough to prevent gene flow sufficient to erode any divergence that has accrued. Interestingly, both split-migration and secondary contact models were strongly supported by data from lineages in each of our two groups of lower-divergence and higher-divergence taxa (Figure 3c–f, Table 3). Our results are concordant with the suggestion of two steady states of divergence (e.g. Flaxman et al., 2014), with no taxa observed in the intermediate region of gene flow and genetic divergence (Figure 4).

We are hesitant to examine our UCE-based divergence time estimates in detail for the following reasons. The most important

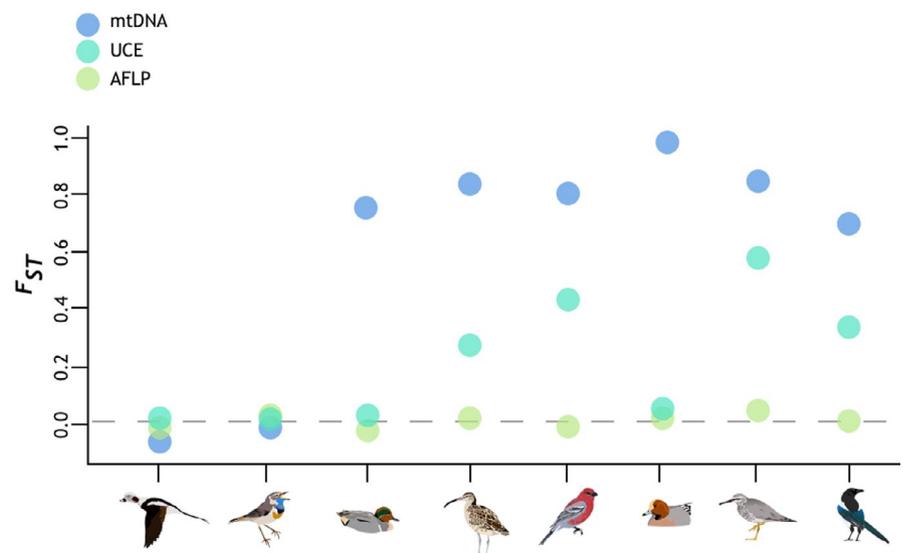


FIGURE 5 F_{ST} comparisons between this study and Humphries and Winker (2011) for all eight lineages, from population to species. Images represent the study taxa detailed in Figure 2. The sequence is populations (2), subspecies (3) and species (3). Although in general there is an increase in estimates of genetic divergence as the speciation process progresses from populations to species (as recognized taxonomically), heterogeneity among marker types and lineages is pronounced [Colour figure can be viewed at wileyonlinelibrary.com]

is that gene flow will affect these estimates, causing them to become younger. That influence is pronounced and variable in our data sets. For example, in the *A. crecca* subspecies, Peters et al. (2012) suggested that continental populations began to split ~2.6 Ma (mtDNA data), while our UCE nuDNA estimate of ~290 Ka (Table 4) suggests that these populations exchanged nuclear genes at a substantial rate during this period, higher than mitochondrial gene flow. Peters et al. (2012) also showed elevated nuDNA gene flow relative to mtDNA, an expected result given male-biased dispersal in this lineage. Even if we accept that our split time estimates for these eight lineages are conservative and represent absolute minimum estimates, our results suggest that all of these lineages experienced more than one glacial–interglacial cycle in this region (such cycles have a periodicity of ~100 Kyr back to 0.74 Ma; Lisiecki & Raymo, 2005) and that some of these lineages have probably experienced more of these cycles than others (Table 4).

4.3 | Contrasts with historic data

Earlier work in these lineages found a remarkable degree of discordance between nuclear and mitochondrial estimates of divergence (Humphries & Winker, 2011). As a new marker class for population genomics, we did not expect UCEs to resolve such discord. Indeed, UCE-based divergence estimates (F_{ST}) lacked significant correlation with divergence estimates from both AFLPs and mtDNA (Figure 5). Some of this discord is likely due to the different effective population sizes between the marker types, with mtDNA having the smallest N_e and highest divergence (fitting nonrecombining, matrilineal inheritance), UCEs being intermediate, and AFLPs having the highest N_e and lowest F_{ST} . Our estimates of divergence from UCEs were concordant with those from mtDNA and AFLPs in some lineages when these general, marker-specific patterns were taken into account; for example, *T. brevipes/incana* and *Pi. pica/hudsonia* have relatively high estimates for each type of marker, and our population-level comparisons (*C. hyemalis*, *L. svecica*) had low divergence (Figure 5). However, we found small but significant levels of divergence in both *C. hyemalis* and *L. svecica* that were not evident from earlier work (Figure 5). Additionally, some lineages that showed discordant divergence signals between mtDNA and AFLPs had significantly nonzero UCE-based F_{ST} estimates despite insignificant estimates from AFLPs (*N. phaeopus*, *P. enucleator*). This reinforces the hypothesis that a strong degree of heterogeneity in divergence between different parts of the genome exists during divergence and speciation (e.g. Harrison & Larson, 2016; Ravinet et al., 2017).

Current taxonomy also does not reflect the genomic patterns we observed. In particular, *N. phaeopus* subspecies have the opportunity for contemporary gene flow, yet gene flow is nearly zero, whereas *M. penelope* and *M. americana* have higher rates of gene flow than would be expected for species that should be reproductively isolated given their taxonomy (see also Peters et al., 2014). UCE data are proving insightful in determining species limits in other avian

lineages (Oswald et al., 2016; Winker et al., 2018, 2019; Zarza et al., 2016), and taxonomic revisions might be warranted in both of these cases. *Tringa brevipes/incana* also have overlapping individuals and breeding ranges in proximity in Beringia (Figure 2), and this overlap might contribute to their slightly elevated gene flow relative to their level of divergence (Table 2 vs. Table 4). However, the levels of gene flow between these *Tringa* taxa that we estimated (Table 4) are well below 1 individual per generation, an important inflection point in the highly nonlinear relationship between F_{ST} and gene flow (Cabe & Alstad, 1994). Given that we are >10 Kyr into the current interglacial, evolutionary reticulation between these two *Tringa* lineages seems highly unlikely.

At present, there is nothing obvious to us in the species' natural history to explain the patterns we have observed other than that their mobility (e.g. through seasonal movements) is sufficient to have shown an expected long-term response to historic, cyclic, connect–disconnect range shifts in Beringia (i.e. in showing gene flow). In the absence of data on divergent selection, it is not clear how isolating mechanisms might have developed or be developing and how and why isolating mechanisms appear to have been effective in some cases and not so effective in others (e.g. shorebirds vs. ducks in Figure 4). With seasonal migration being a dominant avian life history strategy in this region of transitions and migratory divides, factors such as direction and timing of migration, suitable wintering grounds, and sexual selection could all be operating as isolating mechanisms (Turbek, Scordato, & Safran, 2018; Winker, 2010).

4.4 | Conclusions

The cyclic history of isolation and connection between Asia and North America in Beringia has produced a taxonomically diverse group of avian lineages showing divergence with gene flow, and the region's history seems to favour discontinuous dynamics among these lineages in divergence space (Figure 4). The avian lineages in our study span the spectrum of divergence from populations to species and thus encompass the full speciation process. Our data and analyses show that two speciation-with-gene-flow models dominated: a split-with-migration model and a secondary contact model, with four of the lineages we examined exhibiting each (and two of these lineages fitting both). Future studies of speciation in Beringia should examine more lineages to determine whether there is a broader two-phase dynamic to speciation in this region. Additionally, more of the genome could be sampled to clarify the role of gene flow relative to other factors, particularly selection and drift. Gene flow and selection are tightly linked, with greater selection needed to overcome increasing amounts of gene flow if speciation is to proceed (Coyne & Orr, 2004; Price, 2008; Seehausen et al., 2014; Sousa & Hey, 2013). Together, these approaches would further improve our understanding of how divergence and speciation in Beringia have been influenced by this region's cyclic glacial history.

ACKNOWLEDGEMENTS

This study was supported in part by the National Science Foundation (DEB-1242267-1242241-1242260). The University of Washington Burke Museum provided some specimens for use in this study. We thank Troy Kieran, Natalia Bayona-Vasquez, and Jack Withrow for sample processing and the National Center for Genome Analysis Support and the Research Computing Services of the University of Alaska Fairbanks. Oralee Nudson provided invaluable guidance on supercomputer use, and helpful bioinformatics advice was provided by Matthew Miller and Michael Harvey. Bill Manley gave permission to use his map images. We thank Kevin Hawkins, Phil Lavretsky, Jeff Peters, and Ryan Gutenkunst for help in running *daði*. We also thank Kathryn Everson, Lindsey Klueber, Michael Saccone, Krisangel Lopez, Matthew Miller, Anna Santure, and three anonymous reviewers for comments on earlier drafts.

AUTHOR CONTRIBUTIONS

The study was conceived by K.W., T.C.G. and B.C.F. The data were generated by T.C.G. and B.C.F. The bioinformatics pipeline was implemented, custom Python scripts were written and the data were analysed by J.F.M. The first draft was written by J.F.M. All authors contributed to the final version.

DATA AVAILABILITY STATEMENT

Original sequence data have been deposited in the NCBI Sequence Read Archive (SRA; Table S1; PRJN393740).

Assembled sequences used as references, and our analysed *vcf* files have been archived on Dryad (<https://doi.org/10.5061/dryad.zpc866t6n>).

ORCID

Brant C. Faircloth  <https://orcid.org/0000-0002-1943-0217>

Kevin Winker  <https://orcid.org/0000-0002-8985-8104>

REFERENCES

- Afgan, E., Baker, D., van den Beek, M., Blankenberg, D., Bouvier, D., Čech, M., ... Goecks, J. (2016). The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update. *Nucleic Acids Research*, 44(W1), W3–W10. <https://doi.org/10.1093/nar/gkw343>
- Avise, J. C. (2004). *Molecular markers, natural history, and evolution* (2nd ed.). Sunderland, MA: Sinauer and Associates Inc..
- Beerli, P., & Felsenstein, J. (1999). Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. *Genetics*, 152(2), 763–773.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004). Ultraconserved elements in the human genome. *Science*, 304(5675), 1321–1325.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
- Burnham, K. P., & Anderson, D. R. (1998). *Model selection and multimodel inference: A practical information-theoretical approach*. New York, NY: Springer.
- Cabe, P. R., & Alstad, D. N. (1994). Interpreting population differentiation in terms of drift and selection. *Evolutionary Ecology*, 8(5), 489–492. <https://doi.org/10.1007/BF01238253>
- Cahill, J. A., Stirling, I., Kistler, L., Salamzade, R., Ersmark, E., Fulton, T. L., ... Shapiro, B. (2015). Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, 24(6), 1205–1217. <https://doi.org/10.1111/mec.13038>
- Campbell, K. K., Braile, T., & Winker, K. (2016). Integration of genetic and phenotypic data in 48 lineages of Philippine birds shows heterogeneous divergence processes and numerous cryptic species. *PLoS One*, 11(7), e0159325. <https://doi.org/10.1371/journal.pone.0159325>
- Carling, M. D., & Brumfield, R. T. (2007). Gene sampling strategies for multi-locus population estimates of genetic diversity (θ). *PLoS One*, 2(1), e160.
- Casillas, S., & Barbadilla, A. (2017). Molecular population genetics. *Genetics*, 205(3), 1003–1035. <https://doi.org/10.1534/genetics.116.196493>
- Claramunt, S., & Cracraft, J. (2015). A new time tree reveals Earth history's imprint on the evolution of modern birds. *Science Advances*, 1(11), e1501005. <https://doi.org/10.1126/sciadv.1501005>
- Clements, J. F., Schulenberg, T. S., Iliff, M. J., Roberson, D., Fredericks, T. A., Sullivan, B. L., & Wood, C. L. (2018). *The eBird/Clements checklist of birds of the world: v2018*. Retrieved from <http://www.birds.cornell.edu/clementschecklist/download/>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates Inc.
- 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(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- DeChaine, E. G. (2008). A bridge or a barrier? Beringia's influence on the distribution and diversity of tundra plants. *Plant Ecology and Diversity*, 1(2), 197–207. <https://doi.org/10.1080/17550870802328660>
- Delmore, K. E., Hübner, S., Kane, N. C., Schuster, R., Andrew, R. L., Câmara, F., ... Irwin, D. E. (2015). Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. *Molecular Ecology*, 24(8), 1873–1888. <https://doi.org/10.1111/mec.13150>
- Delmore, K. E., Lugo, J., Van Doren, B. M., Lundberg, M., Bensch, S., Irwin, D. E., & Liedvogel, M. (2018). Comparative analysis examining patterns of genomic differentiation across multiple episodes of population divergence in birds. *Evolution Letters*, 2, 76–87. <https://doi.org/10.1002/evl3.46>
- Dickinson, E. C., Remsen, J. V. Jr & Christidis, L. (Eds.) (2013). *The Howard and Moore complete checklist of the birds of the world* (4th ed.). Eastbourne, UK: Aves Press.
- Dobzhansky, T. (1937). *Genetics and the origin of species*. New York, NY: Columbia University Press.
- Ellegren, H., Smeds, L., Burri, R., Olason, P. I., Backström, N., Kawakami, T., ... Wolf, J. B. W. (2012). The genomic landscape of species divergence in *Ficedula* flycatchers. *Nature*, 491(7426), 756–760.
- Faircloth, B. C. (2013). *IllumiProcessor: a trimmomatic wrapper for parallel adapter and quality trimming*. Retrieved from <https://doi.org/10.6079/J9ILL> (accessed 4 November 2016).
- Faircloth, B. C. (2016). PHYLUCe is a software package for the analysis of conserved genomic loci. *Bioinformatics*, 32(5), 786–788. <https://doi.org/10.1093/bioinformatics/btv646>
- Faircloth, B. C., McCormack, J. E., Crawford, N. G., Harvey, M. G., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology*, 61(5), 717–726. <https://doi.org/10.1093/sysbio/sys004>
- Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. *Trends in Genetics*, 28(7), 342–350. <https://doi.org/10.1016/j.tig.2012.03.009>
- Felsenstein, J. (2005). Accuracy of coalescent likelihood estimates: Do we need more sites, more sequences, or more loci? *Molecular Biology and Evolution*, 23, 691–700. <https://doi.org/10.1093/molbev/msj079>

- Flaxman, S. M., Wacholder, A. C., Feder, J. L., & Nosil, P. (2014). Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Molecular Ecology*, 23(16), 4074–4088. <https://doi.org/10.1111/mec.12750>
- 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(1), 397–423. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132421>
- Galbreath, K. E., Cook, J. A., Eddingsaas, A. A., & DeChaine, E. G. (2011). Diversity and demography in Beringia: Multilocus tests of paleodistribution models reveal the complex history of Arctic ground squirrels. *Evolution*, 65(7), 1879–1896. <https://doi.org/10.1111/j.1558-5646.2011.01287.x>
- Garrigan, D., Kingan, S. B., Pilkington, M. M., Wilder, J. A., Cox, M. P., Soodyall, H., ... Hammer, M. F. (2007). Inferring human population sizes, divergence times and rates of gene flow from mitochondrial, X and Y chromosome resequencing data. *Genetics*, 177, 2195–2207. <https://doi.org/10.1534/genetics.107.077495>
- Geml, J., Laursen, G. A., O'Neill, K., Nusbaum, H. C., & Taylor, D. L. (2005). Beringian origins and cryptic speciation events in the fly agaric (*Amanita muscaria*). *Molecular Ecology*, 15(1), 225–239.
- Giarla, T. C., & Esselstyn, J. A. (2015). The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology*, 64(5), 727–740. <https://doi.org/10.1093/sysbio/syv029>
- 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. <https://doi.org/10.7717/peerj.7755>
- Good, J. M., Vanderpool, D., Keeble, S., & Bi, K. (2015). Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, 69(8), 1961–1972. <https://doi.org/10.1111/evo.12712>
- 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(7), 644–652. <https://doi.org/10.1038/nbt.1883>
- 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(10), e1000695. <https://doi.org/10.1371/journal.pgen.1000695>
- Harrison, R. G., & Larson, E. L. (2016). Heterogeneous genome divergence, differential introgression, and the origin and structure of hybrid zones. *Molecular Ecology*, 25(11), 2454–2466. <https://doi.org/10.1111/mec.13582>
- 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(5), 910–924. <https://doi.org/10.1093/sysbio/syw036>
- Hendry, A. P., Bolnick, D. I., Berner, D., & Peichel, C. L. (2009). Along the speciation continuum in sticklebacks. *Journal of Fish Biology*, 75(8), 2000–2036. <https://doi.org/10.1111/j.1095-8649.2009.02419.x>
- Hoekstra, H. E., Hirschmann, R. J., Bunday, R. A., Insel, P. A., & Crossland, J. P. (2006). A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science*, 313(5783), 101–104.
- Hopkins, D. M. (1959). Cenozoic history of the Bering land bridge: The seaway between the Pacific and Arctic basins has often been a land route between Siberia and Alaska. *Science*, 129(3362), 1519–1528.
- Hopkins, D. M. (1967). *The Bering Land Bridge*. Palo Alto, CA: Stanford University Press.
- Hopkins, D. M., Macneil, F. S., Merklin, R. L., & Petrov, O. M. (1965). Quaternary correlations across Bering Strait: Recent Soviet and American studies cast new light on the history of the Bering land bridge. *Science*, 147(3662), 1107–1114.
- Humphries, E. M., & Winker, K. (2011). Discord reigns among nuclear, mitochondrial and phenotypic estimates of divergence in nine lineages of trans-Beringian birds. *Molecular Ecology*, 20(3), 573–583. <https://doi.org/10.1111/j.1365-294X.2010.04965.x>
- Irwin, D. E., Milá, B., Toews, D. P. L., Brelsford, A., Kenyon, H. L., Porter, A. N., ... Irwin, J. H. (2018). A comparison of genomic islands of differentiation across three young avian species pairs. *Molecular Ecology*, 27(23), 4839–4855. <https://doi.org/10.1111/mec.14858>
- Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., ... Zhang, G. (2014). Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science*, 346, 1320–1331. <https://doi.org/10.1126/science.1253451>
- Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27(21), 3070–3071.
- Jorde, L. B., Watkins, W. S., Bamshad, M. J., Dixon, M. E., Ricker, C. E., Seielstad, M. T., & Batzer, M. A. (2000). The distribution of human genetic diversity: A comparison of mitochondrial, autosomal, and Y-chromosome data. *American Journal of Human Genetics*, 66, 979–988. <https://doi.org/10.1086/302825>
- Katzman, S., Kern, A. D., Bejerano, G., Fewell, G., Fulton, L., Wilson, R. K., ... Haussler, D. (2007). Human genome ultraconserved elements are ultrasselected. *Science*, 317(5840), 915. <https://doi.org/10.1126/science.1142430>
- Kuhner, M. K., Yamato, J., & Felsenstein, J. (1998). Maximum likelihood estimation of population growth rates based on the coalescent. *Genetics*, 149(1), 429–434.
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754–1760. <https://doi.org/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(16), 2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. <https://arxiv.org/abs/1303.3997>
- Lischer, H. E. L., & Excoffier, L. (2012). PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28(2), 298–299. <https://doi.org/10.1093/bioinformatics/btr642>
- Lisiecki, L. E., & Raymo, M. E. (2005). A Pliocene-Pleistocene stack of 57 globally distributed benthic $\delta^{18}\text{O}$ records. *Paleoceanography*, 20, PA1003.
- Mallet, J., Besansky, N., & Hahn, M. W. (2016). How reticulated are species? *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 38(2), 140–149. <https://doi.org/10.1002/bies.201500149>
- Manley, W. F. (2002). *Postglacial flooding of the Bering Land Bridge: A geo-spatial animation*. INSTAAR, University of Colorado, v1. Retrieved from http://instaar.colorado.edu/QGISL/bering_land_bridge
- Mayr, E. (1963). *Animal species and evolution*. Cambridge, MA: Belknap Press.
- Mayr, E. (2004). 80 years of watching the evolutionary scenery. *Science*, 305, 46–47.
- 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(9), 1297–1303. <https://doi.org/10.1101/gr.107524.110>
- McLean, B. S., Jackson, D. J., & Cook, J. A. (2016). Rapid divergence and gene flow at high latitudes shape the history of Holarctic ground squirrels (*Urocitellus*). *Molecular Phylogenetics and Evolution*, 102, 174–188. <https://doi.org/10.1016/j.ympev.2016.05.040>
- Melles, M., Brigham-Grette, J., Minyuk, P. S., Nowaczyk, N. R., Wennrich, V., DeConto, R. M., ... Wagner, B. (2012). 2.8 million years of Arctic

- climate change from Lake El'gygytgyn, NE Russia. *Science*, 337(6092), 315–320.
- 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(3), 440–452.
- 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. <https://doi.org/10.1111/1755-0998.12654>
- Nosil, P., Feder, J. L., Flaxman, S. M., & Gompert, Z. (2017). Tipping points in the dynamics of speciation. *Nature Ecology & Evolution*, 1(2), 1.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Oswald, J. A., Harvey, M. G., Rensen, R. C., Foxworth, D. U., Cardiff, S. W., Dittmann, D. L., ... Brumfield, R. T. (2016). Willet be one species or two? A genomic view of the evolutionary history of *Tringa semipalmata*. *The Auk: Ornithological Advances*, 133(4), 593–614.
- 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(3), 630–643. <https://doi.org/10.1111/mec.14978>
- Peters, J. L., McCracken, K., Pruett, C., Rohwer, S., Drovetski, S. V., Zhuravlev, Y. N., ... Winker, K. (2012). A parapatric propensity for breeding in common teal (*Anas crecca*, sensu lato) precludes the completion of speciation. *Molecular Ecology*, 21(18), 4563–4577. <https://doi.org/10.1111/j.1365-294X.2012.05711>
- Peters, J. L., Winker, K., Millam, K. C., Lavretsky, P., Kulikova, I., Wilson, R. E., ... McCracken, K. G. (2014). Mito-nuclear discord in six congeneric lineages of Holarctic ducks (genus *Anas*). *Molecular Ecology*, 23(12), 2961–2974.
- Pielou, E. C. (2008). *After the ice age: The return of life to glaciated North America*. Chicago, IL: University of Chicago Press.
- Price, T. (2008). *Speciation in birds*. Greenwood Village, CO: Roberts and Company.
- R Core Team & Others (2013). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ragland, G. J., Doellman, M. M., Meyers, P. J., Hood, G. R., Egan, S. P., Powell, T. H. Q., ... Feder, J. L. (2017). A test of genomic modularity among life-history adaptations promoting speciation with gene flow. *Molecular Ecology*, 26(15), 3926–3942. <https://doi.org/10.1111/mec.14178>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477.
- Rheindt, F. E., & Edwards, S. V. (2011). Genetic introgression: An integral but neglected component of speciation in birds. *The Auk*, 128(4), 620–632. <https://doi.org/10.1525/auk.2011.128.4.620>
- Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A. A., Farkas, T. E., ... Nosil, P. (2017). Transitions between phases of genomic differentiation during stick-insect speciation. *Nature Ecology & Evolution*, 1(4), 82. <https://doi.org/10.1038/s41559-017-0082>
- Rincon-Sandoval, M., Betancur-R, R., & Maldonado-Ocampo, J. A. (2019). Comparative phylogeography of trans-Andean freshwater fishes based on genome-wide nuclear and mitochondrial markers. *Molecular Ecology*, 28(5), 1096–1115. <https://doi.org/10.1111/mec.15036>
- 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(1), 142–162.
- Roux, C., Fraïsse, C., Romiguier, J., Anciaux, Y., Galtier, N., & Bierne, N. (2016). Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biology*, 14(12), e2000234. <https://doi.org/10.1371/journal.pbio.2000234>
- Saether, 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.
- Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., ... Widmer, A. (2014). Genomics and the origin of species. *Nature Reviews. Genetics*, 15(3), 176–192. <https://doi.org/10.1038/nrg3644>
- Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., ... Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), 1034–1050. <https://doi.org/10.1101/gr.3715005>
- Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology*, 63(1), 83–95. <https://doi.org/10.1093/sysbio/syt061>
- Sousa, V., & Hey, J. (2013). Understanding the origin of species with genome-scale data: Modelling gene flow. *Nature Reviews. Genetics*, 14(6), 404–414. <https://doi.org/10.1038/nrg3446>
- Stephen, S., Pheasant, M., Makunin, I. V., & Mattick, J. S. (2008). Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Molecular Biology and Evolution*, 25(2), 402–408. <https://doi.org/10.1093/molbev/msm268>
- Toews, D. P. L., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, 21(16), 3907–3930. <https://doi.org/10.1111/j.1365-294X.2012.05664.x>
- Toews, D. P. L., Taylor, S. A., Vallender, R., Brelsford, A., Butcher, B. G., Messer, P. W., & Lovette, I. J. (2016). Plumage genes and little else distinguish the genomes of hybridizing warblers. *Current Biology*, 26(17), 2313–2318. <https://doi.org/10.1016/j.cub.2016.06.034>
- Turbek, S. P., Scordato, E. S. C., & Safran, R. J. (2018). The role of seasonal migration in population divergence and reproductive isolation. *Trends in Ecology and Evolution*, 33(3), 164–175. <https://doi.org/10.1016/j.tree.2017.11.008>
- Van Belleghem, S. M., Rastas, P., Papanicolaou, A., Martin, S. H., Arias, C. F., Supple, M. A., ... Papa, R. (2017). Complex modular architecture around a simple toolkit of wing pattern genes. *Nature Ecology & Evolution*, 1(3), 52. <https://doi.org/10.1038/s41559-016-0052>
- Via, S., & West, J. (2008). The genetic mosaic suggests a new role for hitchhiking in ecological speciation. *Molecular Ecology*, 17(19), 4334–4345. <https://doi.org/10.1111/j.1365-294X.2008.03921.x>
- Vickrey, A. I., Bruders, R., Kronenberg, Z., Mackey, E., Bohlender, R. J., Maclary, E. T., ... Shapiro, M. D. (2018). Introgression of regulatory alleles and a missense coding mutation drive plumage pattern diversity in the rock pigeon. *eLife*, 7, e34803. <https://doi.org/10.7554/eLife.34803>
- Vijay, N., Weissensteiner, M., Burri, R., Kawakami, T., Ellegren, H., & Wolf, J. B. W. (2017). Genomewide patterns of variation in genetic diversity are shared among populations, species and higher-order taxa. *Molecular Ecology*, 26(16), 4284–4295. <https://doi.org/10.1111/mec.14195>
- Winker, K. (2009). Reuniting phenotype and genotype in biodiversity research. *BioScience*, 59(8), 657–665. <https://doi.org/10.1525/bio.2009.59.8.7>
- 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. <https://doi.org/10.1525/om.2010.69.1.1>
- Winker, K., Glenn, T. C., & Faircloth, B. C. (2018). Ultraconserved elements (UCEs) illuminate the population genomics of a recent, high-latitude avian speciation event. *PeerJ*, 6, e5735. <https://doi.org/10.7717/peerj.5735>
- Winker, K., Glenn, T. C., Withrow, J., Sealy, S. G., & Faircloth, B. C. (2019). Speciation despite gene flow in two owls (*Aegolius* spp.): Evidence from 2,517 ultraconserved element loci. *The Auk*, 136(2), ukz012. <https://doi.org/10.1093/auk/ukz012>

- Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews. Genetics*, *18*(2), 87–100. <https://doi.org/10.1038/nrg.2016.133>
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, *16*(2), 97–159.
- Yamaguchi, R., & Iwasa, Y. (2016). Smallness of the number of incompatibility loci can facilitate parapatric speciation. *Journal of Theoretical Biology*, *405*, 36–45. <https://doi.org/10.1016/j.jtbi.2015.10.024>
- Zarza, E., Faircloth, B. C., Tsai, W. L. E., Bryson, R. W. Jr, Klicka, J., & McCormack, J. E. (2016). Hidden histories of gene flow in highland birds revealed with genomic markers. *Molecular Ecology*, *25*(20), 5144–5157. <https://doi.org/10.1111/mec.13813>
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology: A Journal of Computational Molecular Cell Biology*, *7*(1–2), 203–214. <https://doi.org/10.1089/10665270050081478>

SUPPORTING INFORMATION

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

How to cite this article: McLaughlin JF, Faircloth BC, Glenn TC, Winker K. Divergence, gene flow and speciation in eight lineages of trans-Beringian birds. *Mol Ecol*. 2020;29:3526–3542. <https://doi.org/10.1111/mec.15574>