The genetic consequences of historic climate change on the contemporary population structure of a widespread temperate North American songbird

Alison Cloutier [1], David Tsz Chung Chan, Emily Shui Kei Poon [2], Simon Yung Wa Sin [3],[*]

School of Biological Sciences, The University of Hong Kong, Hong Kong, China

ABSTRACT

Studies of widely distributed species can inform our understanding of how past demographic events tied to historic glaciation and ongoing population genetic processes interact to shape contemporaneous patterns of biodiversity at a continental scale. In this study, we used whole-genome resequencing to investigate the current population structure and genetic signatures of past demographic events in the widespread migratory American goldfinch (*Spinus tristis*). Phylogenetic relationships inferred from whole mitochondrial genomes were poorly resolved. In contrast, a genome-wide panel of > 4.5 million single nucleotide polymorphisms (SNPs) strongly supported the existence of eastern and western populations separated by western mountain ranges and additional population structuring within the western clade. Demographic modeling estimated that the eastern and western populations diverged approximately one million years ago, and both populations experienced subsequent population bottlenecks during the last glacial period. Species distribution models showed a severe contraction of suitable habitat for the American goldfinch during this period, with predicted discontinuities that are consistent with multiple, isolated glacial refugia that coincide with present-day population structure. Low overall genetic differentiation between the eastern and western populations ($F_{ST} \sim 0.01$) suggests ongoing gene flow accompanied divergence, and individuals with admixed genomic signatures were sampled along a potential contact zone. Nevertheless, outlier SNPs were identified near genes associated with feather color, song, and migratory behavior and provide strong candidates for further study of the mechanisms underlying reproductive isolation and speciation in birds.

1. Introduction

Understanding how historic demographic events and contemporary population genetic processes contribute to shaping current patterns of biodiversity is receiving renewed focus in view of the threats posed by anthropogenic climate change and landscape-scale alterations to habitat (Cristofari et al. 2018, Dierickx et al. 2020, Gagnaire 2020, McCaslin and Heath 2020). In temperate North America, Quaternary ice ages are recognized as a major historic influence that shaped current species distributions and population structure (Hewitt 2000, 2004). Alternating cycles of glaciation interspersed with periods of interglacial warming during the Pleistocene profoundly reshaped the available habitat for many species and caused populations to contract into refugia located at the southern edges of advancing glaciers before expanding into newly

available habitat upon glacial retreat (Hewitt 2000, 2004; Brunsfeld et al. 2001). These historic responses to a changing climate in many cases left tell-tale signatures in the genomes of current species that can be used to infer past demographic events such as population bottlenecks associated with refugial contraction or population expansions that followed glacial retreat (Milà et al. 2000, Spellman et al. 2007, Nadachowska-Brzyska et al. 2015).

However, current genetic diversity is not merely a static byproduct of deep demographic history but is also a dynamic trait that both reflects and influences ongoing processes of migration, genetic drift, and selection at the population level (Wolf et al. 2010, Miraldo et al. 2016; Sin et al. 2021). Studying these processes, at least at the whole-genome level, has traditionally been the purview of comparative population genomics, but the value of phylogeographic methods that interpret

patterns of genetic variation in a geographically explicit manner has long been recognized (Avise et al. 1987, Avise 2000, Gagnaire 2020, Edwards et al. 2022). A major focus of comparative phylogeographic approaches is to identify commonalities in the population structure and divergence times of co-distributed taxa to understand how landscape features have shaped current species assemblages (Burbrink et al. 2016, Edwards et al. 2022, Huynh et al. 2023). Nevertheless, intraspecific studies, especially of species whose ecology, behavior and life history are well characterized, can not only generate baseline data necessary for future comparative studies but can also yield a fine-grained view of the mechanics of selection underlying processes of local adaptation and reproductive isolation. Although sometimes overlooked relative to species of conservation concern, common and widespread species can offer insight regarding how demographic events have shaped continental-scale biodiversity. Widespread species additionally often harbor geographically structured phenotypic or behavioral variation whose genetic basis can more broadly inform our understanding of speciation processes.

The American goldfinch (*Spinus tristis*, order Passeriformes, family Fringillidae) is a widespread granivorous passerine common to weedy fields, floodplains, and open woodlands across the entire breadth of temperate North America (McGraw and Middleton 2020). With an estimated census size of 44 million breeding individuals (Partners in Flight, 2020), the American goldfinch is a common species throughout much of its range and is categorized as of Least Concern according to the IUCN Red List of Threatened Species (IUCN 2016). American goldfinch are distributed as far north as central Canada during the breeding season, and as far south as central Mexico (in the east) and northern Baja California (in the west) during the winter, with a wide band stretching

across the middle of this distributional range with year-round species occurrence (McGraw and Middleton 2020, Fig. S1). The American goldfinch is a partial migrant, where some individuals migrate southward to overwinter whereas others remain in sedentary year-round resident populations (Prescott and Middleton 1990). Banding data has shown that even in portions of its range where the American goldfinch occurs year-round, some local breeding populations migrate south in winter and are replaced by distinct overwintering populations (Middleton 1978).

There are four described subspecies of American goldfinch that occupy partially overlapping distributional ranges and that are distinguished by clinal variation in body size and color saturation, in addition to differences in the extent of white or pale wing and tail markings (McGraw and Middleton 2020). *S. t. tristis* (Linnaeus 1758) is the most common subspecies and is found throughout eastern North America as far west as Colorado (Fig. 1). *S. t. pallidus* (Mearns 1890) is distributed across the Great Plains, Rocky Mountains, and Intermountain West. The remaining two subspecies occur along the Pacific coast and coastal mountains. *S. t. jewetti* (van Rossem 1943) breeds in the Pacific Northwest and overwinters as far south as California. *S. t. salicamans* (Grinnell 1897) is largely resident along the Pacific coast from southern Oregon through Southern California, but some individuals overwinter as far south as northwestern Baja California and as far east as the Mojave and Colorado Deserts (McGraw and Middleton 2020).

Despite being a common species and the focus of research spanning topics as varied as cold tolerance (e.g. Cheviron and Swanson 2017), carotenoid-based pigmentation (e.g. Kelly et al. 2012) and the effects of herbicide use (Sughrue et al. 2008), there have been no previous efforts to study the evolutionary history and population differentiation of the
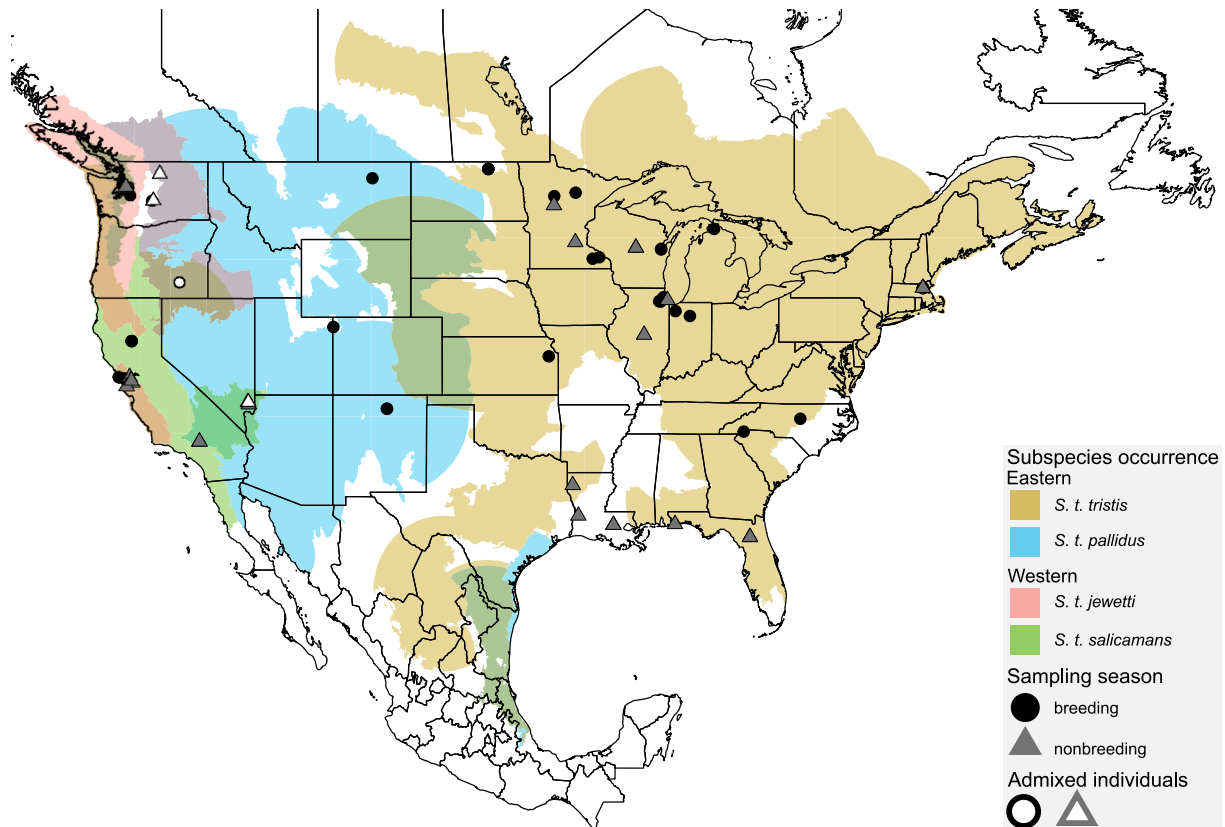

Fig. 1. Subspecies range and sampling map for American goldfinch. Shaded areas indicate distributions of the four described subspecies estimated from occurrence data obtained from GBIF in conjunction with described terrestrial ecoregions. Darker shaded areas indicate regions of overlap between estimated subspecies distributions. Filled symbols indicate the sampling locations of individuals used for whole-genome resequencing, with black circles indicating individuals sampled during the breeding season and gray triangles indicating individuals sampled during the nonbreeding season. Symbols with white centers represent the five individuals with greatest genome admixture from Fig. 2c (note that two breeding season individuals were sampled from the same location in Oregon).

American goldfinch with molecular sequence data. In this study, we generate whole-genome resequencing data for 70 individuals sampled across the distributional range of American goldfinch to investigate population structure using complete mitochondrial DNA (mtDNA) genomes and genome-wide single nucleotide polymorphism (SNP) markers. Our aims were: 1) to assess the extent to which patterns of genetic differentiation coincide with described subspecies; 2) to examine how the demographic history of American goldfinch might have influenced current genetic variation, to what extent population differentiation coincides with other co-distributed taxa and what common biogeographic patterns suggest regarding glacial refugia and post-glacial expansion dynamics in temperate North America. We hypothesized that major landscape features such as western mountain ranges, acting in concert with isolated glacial refugia to either side of the geographic divide, might have produced genetically differentiated populations despite the current widespread and contiguous distribution of American goldfinch across temperate North America; 3) to investigate whether regions of the genome under selection coincide with genes likely to underlie observed phenotypic or behavioral differences among populations and what role these traits might play in reinforcing population divergence. We hypothesized that divergent selection acting upon loci involved in feather coloration, migratory behavior, or song might occur as these traits are associated with sexual selection and reproductive isolation in other passerines (Bearhop et al. 2005, Tietze 2018).

2. Materials and methods

2.1. Whole-genome resequencing

Genomic DNA was isolated from blood or tissue samples from 70 American goldfinch individuals obtained from museum collections or additional field sampling dating between 1983 and 2017 (Fig. 1, Suppl. Data 1). Samples were chosen to represent the distributional range of described subspecies as well as to cover both the breeding and nonbreeding periods. DNA was extracted using the E.Z.N.A. Tissue DNA Kit (Omega Bio-tek Inc., Norcross, Georgia USA) following the manufacturer's protocol and DNA concentration was measured with the Qubit DNA Assay Kit (Thermo Fisher Scientific, Waltham, Massachusetts USA). Libraries of 350 bp insert size were prepared and whole-genome resequencing was conducted by Novogene (Hong Kong). Libraries were sequenced in paired-end mode (2 x 150 bp reads) on an Illumina NovaSeq platform to yield approximately 20 Gb of data per sample. Raw sequencing quality was assessed with FastQC v. 0.11.8 (Andrews 2010).

2.2. Data analysis

2.2.1. Subspecies range mapping

Occurrence data for each of the four described subspecies of American goldfinch were downloaded from the Global Biodiversity Information Facility (GBIF 2024a-d). Distributional range polygons for each subspecies were generated using default parameters of the R package gbif.range (Chauvier et al. 2022) to combine decimal latitude and longitude coordinates from occurrence data with available shapefiles describing terrestrial ecoregions (Olson et al. 2001, The Nature Conservancy 2009).

2.2.2. Read mapping and variant calling

Fastp v. 0.23.2 (Chen et al. 2018) was used to trim adapter sequences and to merge overlapping read pairs. Processed reads were mapped to an existing American goldfinch reference genome assembly (S. Y. W. Sin, manuscript in prep.) with BWA-MEM v. 0.7.17 (Li and Durbin 2009). SAMtools v. 1.14 (Li et al. 2009, Danecek et al. 2021) was used to merge and sort mapped reads prior to marking and removing duplicates with Picard v. 2.26.6 (https://broadinstitute.github.io/picard). Mosdepth v. 0.3.4 (Pedersen & Quinlan 2018) was used to calculate mapping depth of coverage (DoC).

Variants were called using Genome Analysis Toolkit (GATK) HaplotypeCaller v. 4.2.4.0 (Van der Auwera and O'Connor, 2020) followed by GATK GenomicsDBImport and GenotypeGVCFs modules to output variant call format (VCF) files for individual scaffolds. BCFtools v. 1.14 (Danecek et al. 2021) was used to concatenate output into a single VCF and to filter data to retain SNP variants only and to exclude scaffolds that aligned to sex chromosomes of the zebra finch with SatsumaSynteny (see below). VCFTools v. 0.1.17 (Danecek et al. 2011) was used to filter goldfinch scaffolds below 1 Mb in length and to retain biallelic SNPs only. BCFtools and GATK's VariantFiltration module were then used to implement the Genome Analysis Toolkit best practices pipeline for variant calling (Van der Auwera et al. 2013, Van der Auwera and O'Connor, 2020) to filter SNPs with: QualByDepth (QD) < 2.0, FisherStrand (FS) > 60.0, StrandOddsRatio (SOR) > 3.0, RMS Mapping Quality (MQ) < 40.0, Mapping Quality Rank Sum Test (MQRankSum) < -12.5, and ReadPosRankSum < -8.0 or > 8.0. VCFTools was used to filter SNPs at a minor allele frequency (MAF) threshold of 0.05, maximum missing threshold of 0.95 (i.e. allowing up to 5 % missing genotypes per SNP), minDP 3, min-meanDP 3, maxDP 22 and max-meanDP 22 (with the maximum values set at 1.5X the average SNP depth of coverage [DoC] across individuals). PLINK v. 1.90b6.24 (Chang et al. 2015) was used to prune filtered SNPs by linkage disequilibrium with the settings −-indep-pairwise 50 10 0.5 (50 Kb window size, 10 bp step size, and r^2 threshold 0.5). VCFTools was used to output summary metrics for data set quality assessment, including measures of heterozygosity, missingness, and depth of coverage. NGSrelate v. 2 (Hanghøj et al. 2019) was used to calculate pairwise coefficients of relatedness between individuals.

SatsumaSynteny v. 3.0 (Grabherr et al. 2010) was used to align scaffolds of the reference American goldfinch genome assembly to the representative RefSeq genome for the zebra finch (*Taeniopygia guttata* assembly bTaeGut1.4pri, GenBank accession GCF_003957565.2). Goldfinch scaffolds aligning to zebra finch sex chromosomes were excluded from further analyses. Remaining goldfinch scaffolds greater than 1 Mb in length were ordered and oriented into autosomal pseudochromosomes according to their alignment with zebra finch chromosome assemblies and zebra finch coordinates were lifted over to the goldfinch for positions that had unique reciprocal alignments between the two genomes.

2.2.3. Population structure and genetic diversity

Principal components analysis (PCA) of the filtered SNP data set was conducted with PLINK and the results were visualized using the ggplot2 package in R v. 4.1.2 (Wickham 2016, R Core Team 2021). ADMIXTURE v. 1.3.0 (Alexander et al. 2009) was used to infer population structure by assigning individuals to a pre-defined number of genetic clusters (K). Ten replicates were run for each of K = 1 through K = 6, with five-fold cross-validation and 200 bootstrap replicates. The optimal cluster value was determined based on the lowest cross-validation error as well as using the Best K method of Evanno et al. (2005) implemented in CLUMPAK v. 1.1 (Kopelman et al. 2015). ADMIXTURE clustering results were summarized with CLUMPAK and visualized with the R package pophelper (Francis 2017). VCFTools was used to calculate genome-wide estimates of Tajima's D in 20 Kb windows and nucleotide diversity (π) in 20 Kb sliding windows with a 10 Kb step size.

2.2.4. Mitochondrial DNA (mtDNA) assembly and phylogeny

Trimmomatic v. 0.39 (Bolger et al. 2014) was used for adapter removal and read trimming with options ILLUMINACLIP:TruSeq3-PE-2.fa:2:30:10:1:true. 30 million read pairs per individual were subsampled with BBmap v. 39.01 (Bushnell 2014) and were iteratively mapped to the complete mitochondrial genome of the lesser goldfinch (*Spinus psaltria*, GenBank accession NC_025627) with MITObim v. 1.9.1 (Hahn et al. 2013). A consensus American goldfinch mtDNA genome was computed in Geneious Prime v. 2023 (https://www.geneious.com) from the MITObim assemblies for all individuals and the full set of trimmed

reads for each individual was then mapped against this American goldfinch consensus reference with BWA-MEM. Mapped reads were processed with SAMtools to retain reads with minimum mapping and base qualities of 30 and duplicates were marked and removed with Picard. A final consensus mtDNA genome was then output for each individual in Geneious, and MAFFT v. 7.520 (Katoh and Standley 2013) was used to align the American goldfinch mtDNA sequences along with the lesser goldfinch outgroup.

ModelFinder (Kalyaanamoorthy et al. 2017) was used to determine the best-fitting substitution model and partitioning scheme for phylogenetic inference (determined as: partition1 containing the control region [CR], tRNAs, rRNAs, and 1st codon positions of protein-coding genes with a TIM2 + F + I substitution model; partition2 with 2nd codon positions and TN + F + I model; partition3 with 3rd codon positions and TPM3 + F + G4 substitution model). The maximum-likelihood tree was inferred with IQ-TREE v. 2.1.3 (Minh et al. 2020) from 100 searches beginning with random starting trees, and confidence estimates were obtained from 500 nonparametric bootstrap replicates. Bootstrap values were drawn on the maximum-likelihood tree with RAxML v. 8.2.11 (Stamatakis 2014) and the topology was outgroup rooted with lesser goldfinch using ETE3 (Huerta-Cepas et al. 2016).

2.2.5. Nuclear DNA phylogeny

BCFtools was used to output individual SNP genotypes from the final filtered VCF and heterozygous sites within an individual were replaced with their IUPAC code. Sequence for the outgroup zebra finch was compiled using coordinates lifted over from the SatsumaSynteny whole-genome alignment detailed above.

The maximum-likelihood topology was inferred in IQ-TREE from a starting set of 100 trees and a general time reversible (GTR) substitution model. Confidence was assessed with 100 nonparametric bootstrap replicates. Bootstrap replicates were plotted on the maximum-likelihood tree with RAxML and the topology was outgroup rooted with the zebra finch using ETE3.

In addition to the 'base' data alignment matrix detailed above, we also tested the effect of replacing IUPAC codes with 'called' bases for sites where the most frequent allele within an individual occurred at > 2X that of the less common allele, as well as using trimAl v. 1.2rev59 (Capella-Gutiérrez et al. 2009) to trim alignment columns that consisted of > 10 %, >20 %, or > 30 % missing data.

2.2.6. Demographic history

The pairwise sequential Markovian coalescent (PSMC) was used to reconstruct the demographic histories of the eastern and western goldfinch populations. Sequencing reads from the individual used for the reference *de novo* whole-genome assembly (*S. t. tristis*, S. Y. W. Sin, manuscript in prep.), with an average 48X genomic depth of coverage, served as the representative of the eastern population. GOFI_59, which was sampled from Washington state (Suppl. Data 1), was used as the western population representative. A second 350 bp insert sequencing library (138,048,084 read pairs, GenBank accession SRR27604542) was constructed and sequenced in paired-end mode in a partial Illumina NovaSeq lane to increase genome average depth of coverage for this sample to 35X.

Sequencing reads were pre-processed with fastp and mapped to the reference American goldfinch assembly with BWA-MEM as described above. SAMtools was used to filter reads with mapping quality < 30 and Picard was used to mark and remove duplicates. SAMtools was then used to randomly subsample mapped reads for the eastern individual to reduce average DoC to 35X to match that of the western individual, and to retain only autosomal scaffolds greater than 1 Mb in length for both individuals. Consensus calling was done with BCFtools mpileup with minimum base and mapping qualities of 30, followed by BCFtools consensus calling ('call −c' option), and the vcf2fq utility with minimum base quality 30 and sequencing depth thresholds set according to the PSMC developer's recommendations (https://github.com/lh3/psmc,

minimum depth of 1/3 the average [option −d 11] and maximum depth 2X the average [option −D 70]).

PSMC v. 2016.1.21 (Li and Durbin 2011) analyses were run with an atomic time interval parameter found to be appropriate for a wide range of avian species (Nadachowska-Brzyska et. al 2015, parameter −p 4 + 30*2 + 4 + 6 + 10) and confidence was assessed with 100 bootstrap replicates.

EasySFS (https://github.com/isaaccovercast/easySFS, Gutenkunst et al. 2009) was used to compute the folded site frequency spectrum (SFS) for use in two additional methods of demographic inference, Stairway Plot 2 and fastsimcoal2. Site frequency spectra were generated from SNP data with no minor allele frequency (MAF) filtering applied and values were projected down to 86 haplotypes (of 96 total) for the eastern population and 34 haplotypes (of 38 total) for the western population to maximize the total number of segregating sites while retaining a sufficiently large sample size for analysis (Gutenkunst et al. 2009).

Stairway Plot 2 v. 2.1.1 (Liu and Fu 2015, 2020) was run separately for the eastern and western American goldfinch populations using the two-epoch model with the default 67 % of sites used for training, 200 input files created for estimation of pseudo-CIs, and the default four values for the number of random break points for each trial as specified by the authors.

A second SFS-based method, the fast sequential Markov coalescent (fastsimcoal2 v. 2702, Excoffier et al. 2013, 2021) was used to estimate demographic parameters under eight simulated models: 1) no migration between eastern and western goldfinch populations, 2) migration in two different time periods, 3) early migration, 4) ongoing migration, 5) recent migration, 6) early migration followed by independent bottleneck events in each population, 7) independent bottlenecks in each population with no migration, 8) independent bottlenecks in each population followed by migration between populations. Control files specifying fastsimcoal2 models and parameter search ranges are provided in Supplementary File 2. Each model was run 50 times with 500,000 coalescent simulations, the default Brent tolerance level of 0.01, a maximum of 100 ECM optimization cycles, and SFS entries with fewer than 10 SNPs were pooled. The run with the highest estimated likelihood was chosen for each candidate model, and candidate models were compared according to the Akaike information criterion (AIC). Confidence intervals for estimated parameters in the best candidate model were obtained from 10 parametric bootstrap resamplings of the input SFS data.

Results of all demographic analyses were scaled with the estimated mutation rate for the medium ground finch (*Geospiza fortis*, 3.4412 x 10^{-9}, Nadachowska-Brzyska et al. 2015) since no estimates are available for the American goldfinch. A generation time equal to twice the average age of sexual maturity of the American goldfinch according to the Animal Ageing and Longevity Database was used (Tacutu et al. 2018, generation time = 2 years).

2.2.7. Species distribution modeling

Species distribution models (SDMs) were generated using MaxENT v.3.4.4 (Phillips et al. 2017) to investigate the potential effect of paleoclimate on the distribution of American goldfinch. Occurrence data for American goldfinch were downloaded from eBird (eBird 2021). Only data from 1979 to 2013 were retained to match available climate data from PaleoClim (Brown et al. 2018), and occurrence data for nonbreeding (January to February) and breeding (July to August) periods were analyzed separately. Occurrence data were further filtered to retain observations from the middle of each breeding and nonbreeding season to account for individual variation in migration time and to ensure individuals were correctly assigned to their breeding or nonbreeding grounds. Data lacking GPS coordinates were removed and filtered data were rarefied to avoid overfitting through removal of samples located within 40 km of each other with SDMtoolbox (Brown et al. 2017). A total of 26,409 points for breeding season and 22,564 points for nonbreeding season were retained.

Nineteen terrestrial bioclimatic variables were obtained from the PaleoClim database (Brown et al. 2018) for the present time period (1979–2013), Meghalayan (4.2–0.3 thousand years before present, Kbp), Northgrippian (8.326–4.2 Kbp), Greenlandian (11.7–8.326 Kbp), Younger Dryas Stadial (12.9–11.7 Kbp), Bølling-Allerød (14.7–12.9 Kbp), Heinrich Stadial 1 (17.0–14.7 Kbp), Last Glacial Maximum (*c.* 21 Kbp), Last Interglacial (*c.* 130 Kbp), and Marine isotope stage (MIS) 19 (*c.* 787 Kbp). Bioclimatic data were separated into breeding and nonbreeding season based on their description, including climate data of wettest, driest, warmest, and coldest months or quarters since there are no paleoclimatic data per month available from PaleoClim. Global climate data were cropped to the North American continent (15°03′33.6″N to 71°21′55.9″N and 53°59′18.2″W to 164°51′26.3″W). Bioclimatic parameters with a correlation coefficient larger than 0.8 were removed before inferring SDMs for all environmental factors for nonbreeding and breeding seasons, and factors were removed based on jackknife, factor ROC curve, and model AUC. Predictors Bio 1, Bio 8, Bio 12, and Bio 15 were retained for the breeding season, and Bio 1, Bio 8, Bio 15, and Bio 19 were retained for the nonbreeding season. The final SDMs were generated using MaxENT with 50 replicates for each model. Model output and spatial environmental data were visualized in QGIS (QGIS Development Team 2021).

2.2.8. F_{ST} outlier analysis

We compared per-site F_{ST} values between the eastern and western American goldfinch populations and between the two subclades we identified within the western population (see Fig. 3). OutFLANK v. 0.2 (https://github.com/whitlock/OutFLANK) was used to calculate per-site F_{ST} and to identify statistical outliers by fitting a Chi-square distribution to the data and applying a multiple test correction at a q-value threshold of 0.05. BEDTools v. 2.31.0 (Quinlan and Hall 2010) was used to identify annotated genes lying within 1 Kb of outlier SNPs. Statistical overrepresentation of Gene Ontology (GO) terms for biological process, molecular function, and cellular component among outlier SNPs was assessed with the PANTHER classification system v. 18.0 (Mi et al. 2019, Thomas et al. 2022) employing Fisher's exact tests and false discovery rate (FDR) correction.

3. Results

3.1. Whole-genome resequencing and variant calling

Approximately 70 million read pairs were obtained for each of the 70 whole-genome resequencing libraries (Suppl. Data 2, mean = 71.3, range = 63.9–77.3 million read pairs per library), representing on average 11.8X genomic depth of coverage (DoC) when mapped to the reference American goldfinch genome assembly (Suppl. Data 2, range = 8.7–14.0X DoC excluding omitted samples, see below).

No close relatives were identified among sequenced individuals (all pairwise relatedness values < 0.035). However, an initial run-through of the variant calling pipeline and principal components analysis (PCA) identified three outlier individuals: GOFI_07, GOFI_23, and GOFI_57. These samples had the highest proportion of missing SNPs (>25 % of sites) and the lowest average SNP depth of coverage (<5.5X DoC) and were therefore excluded from the data set before re-running the final variant calling pipeline and performing all downstream analyses.

The final SNP dataset comprised 4,783,622 biallelic SNPs with an average 96 % of SNPs called per individual (range = 78.8–98.7 %) and average SNP depth of coverage within an individual of 11.2X (range = 5.4–14.4X, Suppl. Data 2).

3.2. Population structure

Principal components analysis (PCA) did not cluster samples according to sampling season or described subspecies, but instead clearly divided individuals into an eastern group that contained all individuals

sampled throughout the *S. t. tristis* and *S. t. pallidus* distributional ranges and a western group that encompassed all individuals from the *S. t. jewetti* and *S. t. salicamans* ranges (Fig. 2a,b). The second principal component axis (PC2) further subdivided the western group into two clusters that coincided with the described western subspecies when only breeding season individuals were considered (Fig. 2a), but not when samples collected during the nonbreeding season were also included (Fig. 2b). The eastern group lacked a well-defined second cluster in PC2 space and instead separated out two (breeding season only, Fig. 2a) or five individuals (all specimens, Fig. 2b) from the remaining eastern samples (comprising individuals GOFI_36 and 37 in Fig. 2a, and GOFI_36, 37, 49, 61, and 62 in Fig. 2b).

The bestK method of Evanno et al. (2005) supported K = 2 as the optimal cluster value for ADMIXTURE analyses (Fig. 2c), although cross-validation errors and examination of raw likelihood scores would instead indicate that a single population (K = 1) provided the best fit to the data. The K = 2 cluster divided individuals into the same eastern and western groups as were observed from PCA axis PC1. Higher values of K failed to distinguish additional groupings identified from principal component PC2, and replicates with K = 4 clusters and above failed to converge upon a single cluster assignment (Fig. S2). The five individuals listed above that formed a 'diffuse' second PCA group for the eastern samples were clearly assigned to the eastern ADMIXTURE cluster at K = 2; however, these samples displayed the greatest signatures of genomic admixture.

There were eight individuals with no available subspecies metadata and where assignment to a subspecies range was uncertain due to their presence within or very near to a region of predicted distributional overlap between subspecies (specimens designated 'undetermined' in Fig. 2a,b and not assigned to subspecies in Fig. 2c x-axis labels). Of these, GOFI_48, which was sampled from a region of *S. t. tristis/S. t. pallidus* overlap, and GOFI_26, 27, and 28, which were all sampled near the *S. t. jewetti/S. t. salicamans* boundary, can only be classified to the level of eastern and western populations, respectively. GOFI_19 and GOFI_20, sampled during the nonbreeding period from an area of *S. t. salicamans/ S. t. pallidus* overlap, clearly fall within the western group that includes *S. t. salicamans*. GOFI_61 and GOFI_62, both sampled during the nonbreeding period from an area of *S. t. jewetti/S. t. pallidus* overlap are assigned to the eastern population that includes *S. t. pallidus* although, as mentioned above, the cluster assignment proportions for these individuals are the lowest observed.

3.3. Mitochondrial genome assembly and phylogeny

Complete mitochondrial (mtDNA) genomes of 16,814–16,817 bp in length were assembled for all individuals, with all tissue-derived DNA libraries having average depth of coverage > 500X and DNA libraries prepared from blood with DoC > 16X (Suppl. Data 2). There were 315 variable sites in 16,818 bp of aligned sequence for the ingroup American goldfinch specimens (1.9 % of total), with only 0.8 % of sites being parsimony informative (134 of 16,818 alignment columns).

There was a weak trend for western samples to cluster, with 11 of the 19 individuals from the western group identified from PCA and ADMIXTURE forming a clade with 92 % bootstrap support (Fig. 3b). However, the mtDNA phylogeny is overall very poorly resolved and is characterized by low bootstrap support across most of the deeper internal branches. Correspondingly, there is no robust support from mtDNA for individuals clustering according to sampling season, described subspecies, or the eastern and western populations identified from PCA and ADMIXTURE analyses.

3.4. Nuclear SNP phylogeny

In contrast to the mtDNA phylogeny, phylogenetic inference using nuclear SNP data recovered a western clade with 100 % bootstrap support that was identical to the western population from PCA and
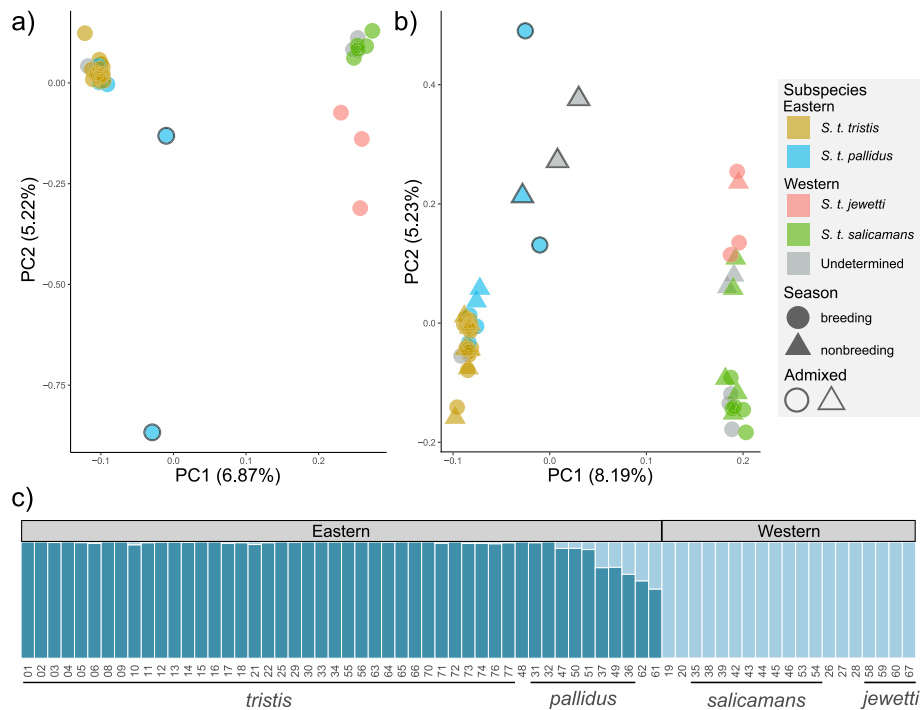
Fig. 2. Principal components analysis (PCA) and ADMIXTURE analysis of population structure in the American goldfinch. a) PCA of individuals sampled during the breeding season. b) PCA of all individuals. Specimens that were identified to subspecies level in museum accession metadata or that fall clearly within a subspecies distributional range are colored whereas specimens not identified to subspecies level and sampled within or immediately adjacent to areas of subspecies distributional overlap are recorded as 'undetermined' and colored in gray. Outlined symbols indicate the five individuals with greatest genomic admixture from part c. c) ADMIXTURE results for cluster size K = 2. The numerical portion of sample IDs are shown along the X-axis. Subspecies designations are shown beneath the x-axis for individuals that were identified to subspecies level or that clearly fell within a predicted subspecies range (corresponding to assignments shown in parts a and b).

ADMIXTURE analyses (Fig. 3a). This western group was further divided into two subclades, each with 100 % support, that were identical to the two western clusters delineated by PCA axis PC2 (Fig. 3a). Additional data sets that assessed the effect of alignment filtering or thresholds for adopting IUPAC degenerate codes produced topologies that were identical in recovering this western clade and two subclades with maximum bootstrap support (results not shown).

Western subclade 1 (Fig. 3c) contains individuals sampled during both the breeding and nonbreeding seasons along the mid-California Pacific coast (with 6 of these 8 individuals identified as *S. t. salicamans* in museum accession metadata), as well as three individuals sampled during the breeding season near a region of predicted *S. t. salicamans/S. t. jewetti* overlap lying to the northeast (none of these samples was identified to subspecies level). Western subclade 2 (Fig. 3c) contains both breeding and nonbreeding season samples from Washington state (none was identified to subspecies level), as well as two individuals sampled from roughly the same mid-California coastal locality as subclade 1 in the nonbreeding season (and recorded as *S. t. salicamans* in accession metadata) and two individuals sampled from a predicted region of *S. t. salicamans/S. t. pallidus* overlap in Southern California during the nonbreeding season (neither was identified to subspecies level).

Unlike the strongly differentiated western specimens, individuals belonging to the eastern population defined by PCA and ADMIXTURE failed to form a well-supported monophyletic clade in the SNP phylogeny. There was also little evidence for lower-level phylogenetic relationships within this larger eastern sample, with almost all internal branches receiving below 50 % bootstrap support (Fig. 3a).

3.5. Demographic analyses

PSMC analysis of single eastern and western representative genomes and Stairway Plot 2 analysis of site frequency spectra (SFS) derived from

all sampled individuals yielded similar overall shapes in the estimated demographic histories of each population through time (Fig. 4a,b). 95 % confidence intervals surrounding estimates of effective population size (Ne) also largely overlap between the two methodologies. PSMC, which is better at inferring more ancient demographic processes (Li and Durbin 2011, Beichman et al. 2018), indicates that the eastern and western populations shared a similar demographic history prior to 1 million years before present (Ybp, Fig. 4a). After this point, the trajectories of the two populations diverge, with the eastern population increasing more sharply and exhibiting a dip in Ne at the beginning of the last glacial period (LGP) before rebounding to an even higher effective population size of *c.* 4 million that holds steady throughout the remainder of the LGP. The western population exhibits a slower rise in effective population size, peaking at a lower maximum estimate and beginning to decline again prior to the onset of the LGP.

Stairway Plot 2, which is better able to reconstruct more recent population size changes (Liu and Fu 2015, 2020; Beichman et al. 2018), is relatively consistent with PSMC throughout except that both eastern and western populations show a pronounced dip in Ne that coincides with the last glacial maximum (LGM) at 20,000–26,000 Ybp rather than declining at the start of the last glacial period (Fig. 4b). Effective sizes of both populations rebound shortly after the LGM, although predicted Ne of the western population remains below what was predicted prior to the LGM, and estimated sizes of both populations remain stable throughout the remaining time interval to the present (Fig. 4b).

In contrast to these results, fastsimcoal2 identified ongoing migration following population divergence as the best candidate model in preference to models incorporating bottleneck episodes (Table S1, Fig. 5a). Despite this difference, the best fastsimcoal2 model estimated current diploid effective population sizes of *c.* 6.86 million and 0.94 million for the eastern and western populations, respectively, and a time to population divergence of *c.* 1 million Ybp that coincided with the results of PSMC and Stairway Plot 2. Although the best candidate model
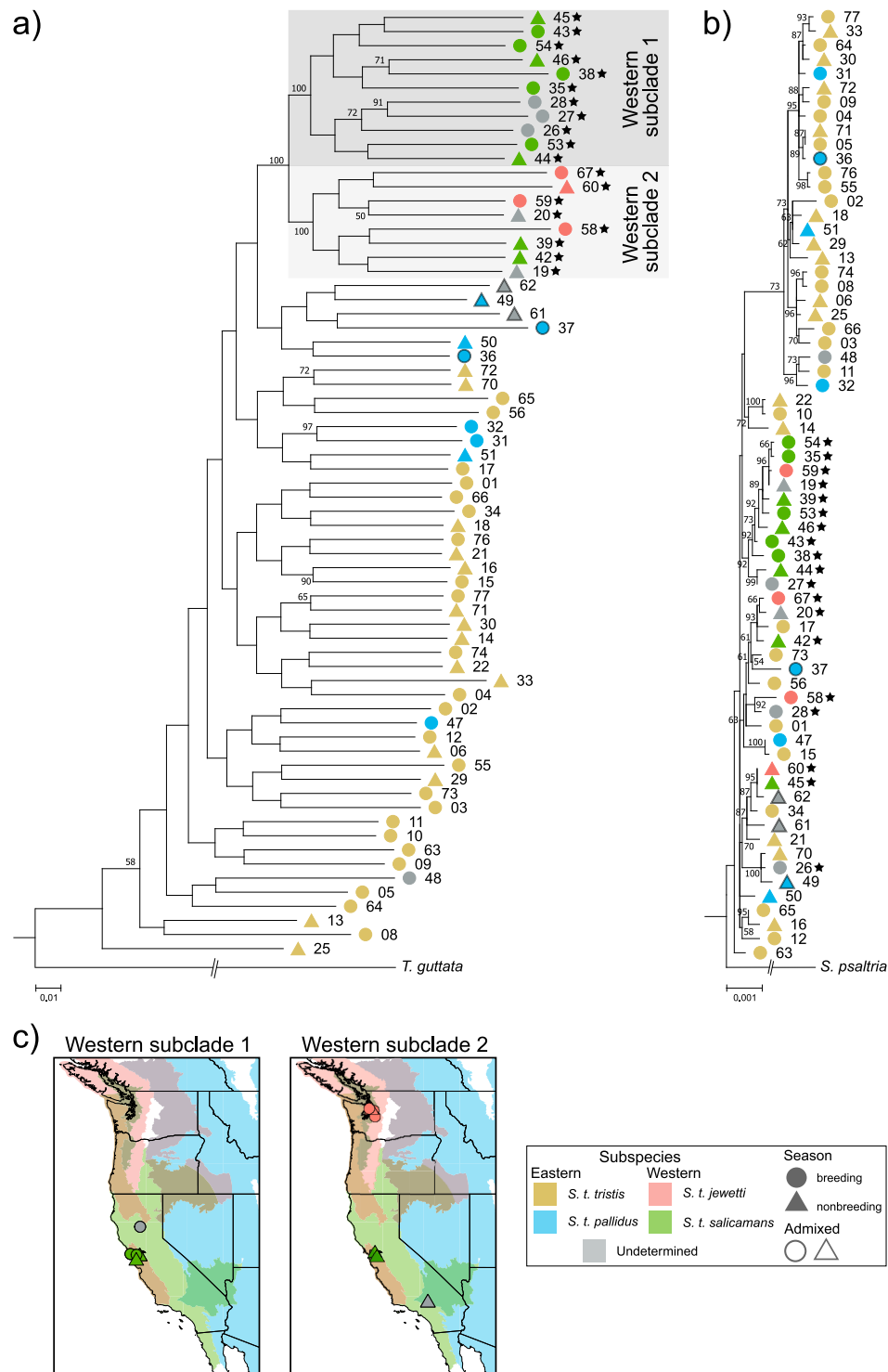
Fig. 3. Phylogenetic relationships among American goldfinch specimens. a) Maximum likelihood phylogeny of American goldfinch and *Taeniopygia guttata* (zebra finch) outgroup inferred from nuclear SNP data with IQ-TREE, with support from 100 bootstrap replicates indicated for relationships with bs > 50 %. Scale bar gives branch length in units of substitutions per site. b) Maximum-likelihood phylogeny of American goldfinch and *Spinus psaltria* (lesser goldfinch) outgroup inferred from complete mtDNA genomes with IQ-TREE. Support from 500 bootstrap replicates is indicated for relationships receiving > 50 % support. Scale bar gives branch length in units of substitutions per site. For parts a and b, the numerical portion of sample IDs from Suppl. Data 1 are indicated for each specimen. Individuals belonging to the western population from PCA and ADMIXTURE analyses (Fig. 2) are indicated with a star symbol, and the two western subclades found at 100 % bootstrap support are labeled and shaded. The five individuals with greatest genomic admixture from Fig. 2c are indicated with outlined symbols. Branches leading to outgroup taxa have been shortened to fit on the page, as denoted by double hash symbols. c) Sampling locations for individuals belonging to the two identified western subclades in part a.
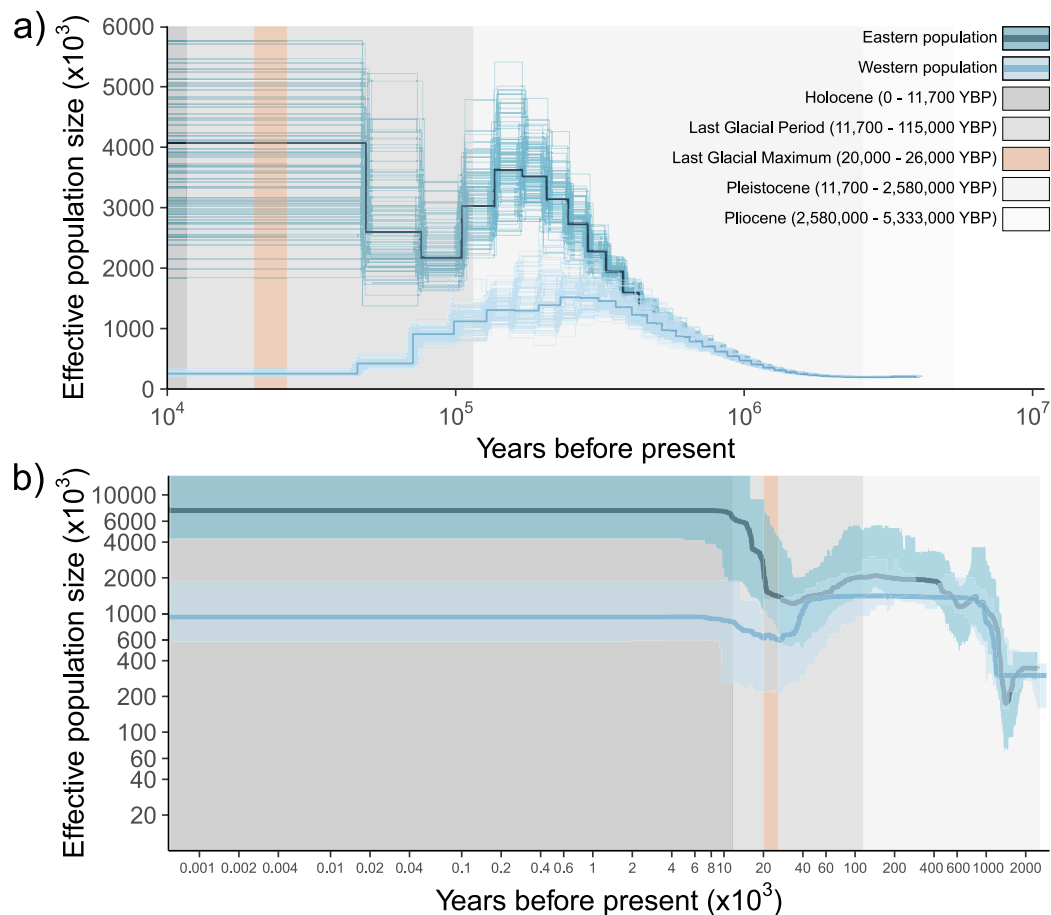
Fig. 4. Inference of the demographic history of American goldfinch. a) PSMC analysis of the genome from a representative individual from the eastern and western populations. b) Stairway Plot 2 analysis of site frequency spectra (SFSs) constructed from individuals used for whole-genome resequencing. Plotted values (light and dark blue lines) show estimated changes in effective population size (Ne) through time, with lighter lines (a) or shaded areas (b) indicating the 95% confidence interval of Ne estimates. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

provides a reasonably good fit overall (Fig. 5b,c), there are some notable areas of divergence between the observed and expected SFS (Fig. 5d), and 95 % confidence intervals for both estimated migration rates encompass zero (Fig. 5a). We attempted to model scenarios that would more closely match the results of PSMC and Stairway Plot 2 to determine if they would provide an improved fit to the observed SFS (for example, modeling independent population expansions after divergence followed by independent bottlenecks and/or population declines). However, most replicates failed to reach convergence for these more parameter-rich models, with parameter estimates failing to stabilize after 100 ECM optimization cycles (results not shown).

3.6. Species distribution modeling

Current habitat suitability predicted from species distribution modeling largely coincides with the described present range of the American goldfinch, except for a more northerly breeding distribution predicted to extend up the Pacific coast and into Alaska and a more southerly breeding range extending into Mexico, albeit both are associated with low probabilities of occurrence (Fig. 6a). Hindcasts of historical habitat suitability predict contractions to breeding and overwintering distributions during cold glacial periods (e.g. Fig. 6b,c) and expansions during warmer interglacial periods (e.g. Fig. 6d,e). The last glacial maximum (LGM, Fig. 6c) is associated with the most severe constriction to predicted habitat suitability, with suitable breeding and overwintering ranges largely restricted to the Pacific coast and extreme southern United States into Mexico. There is a predicted discontinuity in

suitable breeding habitat for American goldfinch between the eastern and western portions of its range during the LGM, as well as a discontinuity between more northern and southern portions of the breeding range along the Pacific coast (Fig. 6c).

3.7. Genomic variation and F_{ST} outlier analysis

Mean genome-wide estimates of nucleotide diversity (π) were identical for the eastern and western populations (0.0011 ± 0.0008 standard deviation (SD) for both, Fig. S3a). Mean estimates of Tajima's D were 0.9254 ± 0.4563 SD for the eastern population and 0.4196 ± 0.4114 SD for the western population (Fig. S3b).

Mean per-site F_{ST} was weak to moderate between the eastern and western populations (0.0132 ± 0.0440 SD, Fig. S4) and very low between the two western subpopulations (0.0026 ± 0.0783 SD, not shown). There were no perfectly segregating SNPs for either the east/west or western subclades comparisons, and there were also no outlier SNPs identified between the western subclades. However, OutFLANK analysis identified 631 outlier SNPs between the eastern and western populations, with 383 of the 631 identified outliers lying within 1 Kb of an annotated gene (291 genes total; some genes contain multiple outlier SNPs, Suppl. Data 3). Statistical overrepresentation tests in PantherDB found no overrepresented Gene Ontology (GO) categories for biological process, molecular function, or cellular component. Closer examination of outlier SNPs falling within annotated genes found that only four SNPs fell within coding regions (one each in *DSCAM*, *FAM199X*, *HKDC1*, and *AHSG*) whereas the remaining outliers localized to intronic or UTR
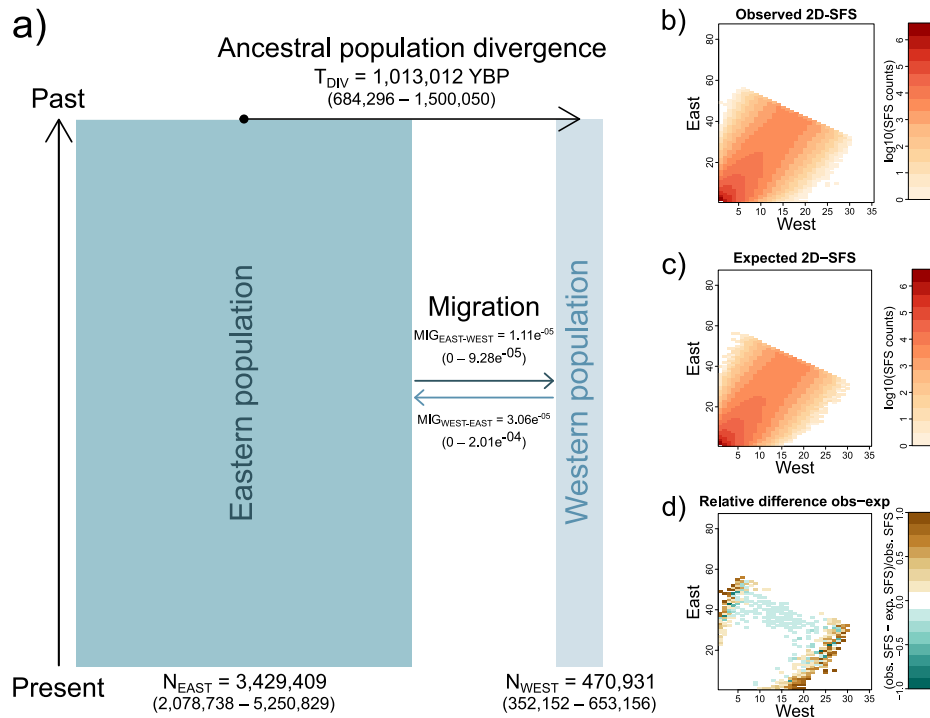
Fig. 5. Parameters of the best-fitting model estimated with fastsimcoal2, involving ongoing migration following divergence of the eastern and western populations of American goldfinch. a) Estimated parameters of the best-fitting candidate model, with 95 % confidence intervals shown in brackets beneath each estimate. Migration events are indicated with blue arrows and represent ongoing events following population divergence. N_{EAST}/N_{WEST} = current effective population sizes of the eastern and western populations, $MIG_{EAST-WEST}/MIG_{WEST-EAST}$ = migration rate from the eastern to western, or western to eastern population, respectively, T_{DIV} = time of ancestral population division into the eastern and western populations. Effective population sizes are given in haploid numbers, and the divergence time is shown as years before present (Ybp). Migration rates are expressed as the probability for any lineage in deme x to move to deme y (i.e. $MIG_{EAST-WEST}$ gives the probability for any lineage in the eastern population to move to the western population, looking backward in time). Panels b-d) Assessment of model fit for the best candidate model shown in part a), showing matrix representations of the observed 2D site frequency spectrum (SFS, b), the 2D SFS expected if the candidate model provided a perfect fit to the data (c), and the relative difference between observed and expected SFS matrices (d). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)
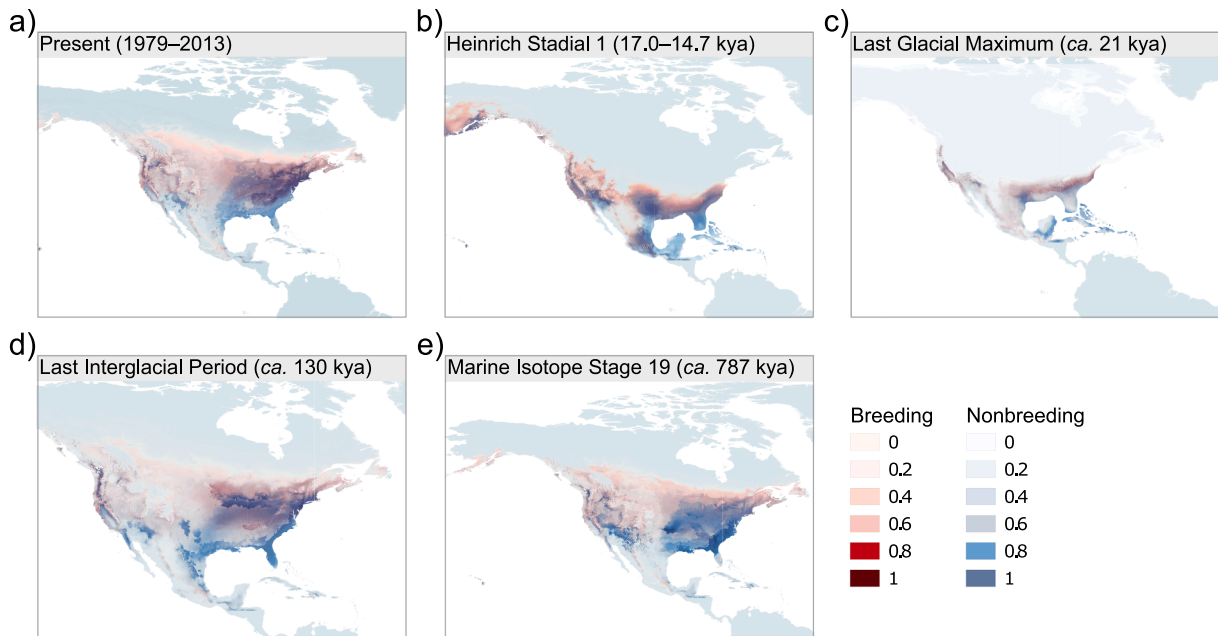

Fig. 6. Species distribution models inferred with MaxENT, showing predicted habitat suitability for the American goldfinch during the present time (a) and historical time periods (b-e). Darker colors indicate higher probabilities of species occurrence for a given geographic area.

sequence. Of the four coding SNPs, only one (in *AHSG*) produced a nonsynonymous substitution, resulting in a semi-conservative replacement of serine by proline.

Among the American goldfinch genes near outlier SNPs, we identified 11 with roles in melanogenesis: *ADCYAP1R1, BMPR1A, BMPR1B, JMJD6, MBP, MYO7A, NAV2, PLCB1, PLCB4, PLCG2,* and *PRKG1* (Suppl. Data 3, Poelstra et al. 2015, Aguillon et al. 2021). An additional seven genes with outlier SNPs have proposed roles in migratory behavior: *DLG2, MBP, NPAS2, PLCB4, PTN, RNF146, ZNF516* (Delmore, 2015; Louder, 2024; Ruegg, 2014; note that myelin basic protein *MBP* and phospholipase C beta 4 *PLCB4* appear as both melanin and migration candidates). Lastly, we identified five outlier genes with proposed roles in vocal learning (*CDH8, FOXP2, KLHL29, SYT1, TENM2*; Enard et al. 2002, Vargha-Khadem et al. 2005, Cahill et al. 2021), as well as *ESR1* (estrogen receptor 1), which is associated with aggression and territorial singing behavior (Maney et al. 2020).

4. Discussion

4.1. Genome-wide nuclear SNPs, but not mtDNA, delineate eastern and western populations of American goldfinch

Although several previous studies used mitochondrial DNA (mtDNA) to investigate phylogenetic relationships and speciation dynamics among goldfinches and siskins (e.g. Arnaiz-Villena et al. 2009, Arnaiz-Villena et al., 2012; Beckman and Witt 2015), to the best of our knowledge this study represents the first use of molecular sequencing data to infer the genetic structure and demographic history within the American goldfinch. In addition to a large panel of more than 4.5 million nuclear SNPs, our whole-genome resequencing approach produced a complete mtDNA genome for each sampled individual. We found low intraspecific sequence variation of mtDNA (only 1.9 % of sites were variable across the entire genome and only 0.8 % were parsimony informative) that provided little support for phylogenetic relationships within American goldfinch. mtDNA was often the marker of choice in past avian phylogeographic studies and provided resolution of population structure at broad geographic scales (reviewed in Avise and Walker 1998). However, it is suggested that mtDNA variation is likely insufficient to resolve relationships at a finer geographic scale for migratory passerines (Lovette et al. 2004), an assessment that would seem to apply here.

In contrast, we find robust support for some intraspecific relationships from nuclear sequence data. Our analysis of genome-wide SNP data does not fully support the division of American goldfinch into its four previously delineated subspecies. Instead, we find evidence for an eastern population that encompasses all individuals sampled from the *S. t. tristis* and *S. t. pallidus* distributional ranges and a western population that includes all individuals sampled from the *S. t. jewetti* and *S. t. salicamans* ranges. The western population extends from central Washington state, Oregon, and east/central California to the Pacific coast, whereas the eastern population covers the entire continental region east of this divide. This pattern strongly suggests that the Cascade and Sierra Nevada Mountain ranges, which extend in a north–south orientation through central Washington and Oregon and east/central California, respectively, represent a geographic barrier between the observed populations. Similar east/west patterns of population differentiation are described for many widespread avian species in temperate North America, with distinct populations often separated by western mountain ranges and/or the Great Plains (e.g. Milot et al. 2000, Kimura et al. 2002, Ruegg and Smith, 2002, Klicka et al. 2011, Lovette et al. 2004, Aguillon et al. 2018, Cicero et al. 2022, DeRaad et al. 2022). East/west discontinuities are also evident in many mammalian species and temperate forest trees (Jaramillo-Correa et al. 2009, Escalante et al. 2010, Holt et al. 2013), indicating that this common phylogeographic break represents a predominant feature of the temperate North American biota across several higher order lineages. In trees, the common east/west

genetic discontinuities coincide with inferred glacial refugia (Jaramillo-Correa et al. 2009), while historic reductions to effective population size in birds that coincide with the onset of the last glacial period (Nadachowska-Brzyska et al. 2015) are also consistent with contraction into isolated refugia. Identifying shared phylogeographic discontinuities across taxa can therefore inform inference of refugial areas and post-glaciation recolonization routes (Avise and Walker 1998, Hewitt 2000, Lovette 2005, Jaramillo-Correa et al. 2009).

Demographic and species distribution modeling for the American goldfinch support the hypothesis that observed east/west genetic differentiation stems from contraction of these populations into isolated refugia to either side of a geographic divide during periods of glacial advance, although the timing of inferred events varies somewhat among methods. All three methods of demographic inference date population divergence to the mid-Pleistocene (*c.* 1 million Ybp), following which PSMC analyses infer reductions to effective population size through the start of the last glacial period and Stairway Plot 2 analyses infer population bottlenecks in both the eastern and western populations concurrent with the last glacial maximum (LGM). Species distribution models predict a contraction in suitable American goldfinch breeding habitat with an east/west discontinuity compatible with isolated glacial refugia during this period, lending further support that historic climate likely contributed to shaping current population structure. The moderate nucleotide diversity observed for both the eastern and western populations ($\pi = 0.0011$) despite large long-term effective population sizes might in part reflect these historic bottlenecks, although similar estimates of nucleotide diversity are reported for other widespread avian species with large census sizes (e.g. hooded crow, *Corvus (corone) cornix*, $\pi = 0.0011$, Dutoit et al. 2017; turkey vulture, *Cathartes aura*, $\pi = 0.0012$, Li et al. 2014, Robinson et al. 2016; great black cormorant, *Phalacrocorax carbo*, $\pi = 0.0014$, Li et al. 2014, Robinson et al. 2016). The relatively low F_{ST} (~ 0.01) and absence of mtDNA haplotype sorting between eastern and western populations of American goldfinch also contrast with a more pronounced divergence reported for some avian taxa across the same Pacific coastal versus interior divide (e.g. Woodhouse's scrub-jay *Aphelocoma woodhouseii* [east] and California scrub-jay *Aphelocoma californica* [west] with pairwise $F_{ST} \sim 0.17$, DeRaad et al. 2022). This difference could imply a much more recent divergence in American goldfinch than is estimated by our demographic models, but could also reflect the much earlier estimated divergence time in scrub-jays (*c.* 2 million Ybp, DeRaad et al. 2022) as well as differences in species ecology or life history influencing the extent of gene flow across repeated paleoclimatic cycles (e.g. both California and Woodhouse's scrub-jays form resident populations).

Despite the clear division into eastern and western populations from PCA, ADMIXTURE, and nuclear phylogenetic analyses, five individuals assigned to the eastern population showed greater admixture proportions than the remaining eastern specimens. These individuals with more admixed genomic signatures were sampled from locales within, or immediately adjacent to, regions of predicted overlap between eastern and western subspecies (between *S. t. pallidus* and *S. t. jewetti* in the north and *S. t. pallidus* and *S. t. salicamans* in the south). Contact zones between the Rocky and Sierra Nevada/Cascade mountains are described for eastern and western populations of several avian taxa (e.g. Barrow-clough et al. 2004, Spellman et al. 2007, Manthey et al. 2011, Cicero et al. 2022, DeRaad et al. 2022). This region is also described as a 'hot spot' of hybrid or suture zones and phylogeographic breaks across mammals, birds, and trees (Swenson and Howard 2005). Our results suggest that introgression could occur along at least some portion of this east/west population divide in the American goldfinch, and this introgression and/or migration across subspecies contact zones could explain why fastsimcoal2 identified ongoing migration as the best candidate demographic model in preference to models that incorporated bottleneck events. Both PSMC and Stairway Plot 2 inferred bottlenecks of relatively small amplitude and short duration, and mean genome-wide estimates of Tajima's D were only weakly positive, indicating that

more recent migration events might be dominating the genomic signal modeled by fastsimcoal2 coalescent simulations (Yang 2014, Burbrink et al. 2016).

4.2. Finer-scaled structuring was not identified within the eastern population of American goldfinch

Analysis of genome-wide SNPs identified no finer-scaled structuring within the eastern population of American goldfinches, and we cannot distinguish between the two eastern subspecies, *S. t. tristis* and *S. t. pallidus*, although we note there was relatively sparse sampling across the *S. t. pallidus* range. Earlier studies of widespread North American passerines employing mtDNA markers similarly found low overall genetic diversity and an absence of genetic structuring among sampling sites in eastern continental North America (Kimura et al. 2002, Ruegg and Smith, 2002, Lovette et al., 2004, Klicka et al. 2011). This absence of genetic differentiation across a broad geographic region could be explained by the loss of variation during bottleneck episodes when populations contract into glacial refugia followed by rapid postglacial expansion (Zink 1996, Hewitt 2000). Weak population structuring, 'star-shaped' mtDNA haplotype networks, and latitudinal gradients in intra-specific genetic variation consistent with post-glacial leading-edge expansion all support a role for historic demographic processes linked to climatic cycles in shaping the current patterns of genetic diversity observed for avian taxa across eastern North America (Zink 1996, Hewitt 2000, Lovette et al. 2004, Klicka et al. 2011, Smith et al. 2017). Gene flow between eastern populations might also be acting to limit genetic differentiation across the east. Biogeographic studies of unglaciated eastern North America identified eight phylogeographic discontinuities that could pose a barrier to gene flow. However, common discontinuities were rarely identified for birds or mammals, suggesting that gene flow across these barriers could be sufficient to prevent genetic structuring in highly vagile species such as migratory birds (Soltis et al. 2006, Lyman and Edwards 2022).

Despite this reported absence of genetic structuring in eastern North American migratory passerines, recent investigations incorporating large numbers of nuclear loci are providing a more nuanced view of breeding population structure in some cases. For instance, Colbeck et al. (2008) distinguished an Atlantic breeding population of the American redstart (*Setophaga ruticilla*) that differed from the remaining mainland continental samples, but neither mtDNA nor amplified fragment length polymorphism (AFLP) markers provided robust support for additional structuring within the mainland. More recently, DeSaix et al. (2023) identified five distinct but weakly differentiated clusters across the American redstart breeding range (including the previously described Atlantic/mainland split) by combining dense population sampling with low-coverage whole-genome sequencing. Our whole-genome resequencing approach yielded ample data in terms of the number of SNP markers generated. However, it is likely that additional sampling within the eastern continental region, particularly throughout the Great Plains and American West, is required to definitively assess whether finer-scale structuring exists within the eastern population of American goldfinch and, if so, to what extent it coincides with the described eastern subspecies.

4.3. The western population of American goldfinch comprises two reciprocally monophyletic subclades

In contrast to the lack of genetic differentiation in the eastern population, principal components analysis and the nuclear SNP phylogeny, but not ADMIXTURE analysis, clearly discriminate between two western subclades of American goldfinch. These groupings might correspond to the described western subspecies, as western subclade 2 includes individuals sampled from the *S. t. jewetti* range in Washington state during the breeding period and subclade 1 contains both breeding and nonbreeding season individuals sampled from California, where *S. t.*

salicamans typically occurs as a year-round resident. Additional sampling across the western distribution of American goldfinch is needed to investigate the full extent of spatial structuring across this region, and to determine if genetic differentiation coincides with the distributional range or morphological and plumage variation of described subspecies.

Greater spatial structuring and/or higher genetic diversity of western populations is noted from mtDNA studies of several North American birds that show east/west population differentiation (e.g. Kimura et al. 2002, Lovette et al. 2004, Klicka et al. 2011). This structure is most apparent for sedentary species or permanent resident populations of migratory species (e.g. Barrowclough et al. 2004, Spellman et al. 2007, Klicka et al. 2011, Manthey et al. 2011), but has also been described for migratory passerines (e.g. Kimura et al. 2002, Ruegg and Smith, 2002, Lovette et al. 2004), leading some authors to suggest that western populations might have undergone less severe reductions in effective population size during Pleistocene glaciation or might have occupied multiple, disjunct glacial refugia (Lovette et al. 2004, Klicka et al. 2011). Species distribution models for the American goldfinch identify a break in suitable breeding habitat during the LGM between northern and southern regions of the Pacific coast, indicating that genetic differentiation between western populations might stem from their occupation of isolated refugia during past glacial cycles. North/south genetic differentiation along the Pacific coast and contact zone 'hot spots' were previously identified from bird, mammal, and tree species distributed west of the Cascade Mountain range, and it is proposed that secondary contact between populations or species expanding from isolated northern and southern glacial refugia is a characteristic biogeographic pattern of western North America (Swenson and Howard 2005). We did not identify admixed individuals or a potential contact zone within the western population of American goldfinch; however, the reciprocally monophyletic western subclades that coincide with a north/south discontinuity when only breeding season individuals are considered echoes the broad phylogeographic trend observed in other taxa.

The topographical complexity of the western mountain ranges and intervening xeric landscapes could also contribute to the current structuring observed among some western populations by limiting post-glaciation gene flow, although it remains unclear to what extent these features might influence differentiation in more vagile species (Calsbeek et al. 2003) and factors such as natal philopatry might play an equally important role (Manthey et al. 2011). We found that all breeding season individuals in western subclade 1 were sampled from California, whereas breeding season individuals in subclade 2 were all sampled from Washington state. However, in the nonbreeding season, some subclade 2 individuals were sampled from the same general region as subclade 1 (and both were recorded as *S. t. salicamans* in museum metadata), or from Southern California, which lies outside the currently described overwintering range for *S. t. jewetti*. It therefore remains unclear if the two subclades we identify here truly correspond to the described western subspecies, although it is possible that some *S. t. jewetti* individuals migrate further south than is currently believed. It should also be noted that subspecies identification for the American goldfinch is sometimes based on geographic location alone since some phenotypic differences are clinal in nature and/or involve more subtle variation in feather color saturation rather than discrete characters such as feather patch coloration. These observations highlight the potential complexity of defining annual cycle connectivity in partially migratory species where populations, or even subspecies, might overlap in distributional range throughout portions of their annual cycle and further work is needed to assign individuals sampled during the nonbreeding period to their source breeding population. With sufficiently dense sampling, population assignment methods can successfully assign individuals to their breeding population even in cases with weak genetic differentiation among reference breeding populations and even when low-coverage whole-genome sequencing data is used (DeSaix et al. 2023, DeSaix et al., 2024). Studying partially migratory species such as the American goldfinch, which has already demonstrated recent shifts in

breeding distributional range (McCaslin and Heath 2020), could prove especially fruitful in understanding the evolutionary cost/benefit tradeoffs in migratory behavior associated with anthropogenic climate change and landscape-scale alterations to habitat.

4.4. F_{ST} outliers identify candidate genes for feather coloration, migration, and vocal learning

In addition to historical demographic events and current gene flow, selection can contribute to current patterns of genetic differentiation by promoting local adaptation or assortative mating. Given the observed population structure in American goldfinch, we hypothesized that divergent selection acting upon loci that are subject to assortative mating or sexual selection in other passerines (Bearhop et al. 2005, Tietze 2018) might underlie some of the described variation in plumage coloration or migratory behavior between American goldfinch subspecies. We found only moderate per-site average F_{ST} in pairwise comparison of the eastern and western populations (F_{ST} = 0.0132) and very weak differentiation between the two western subclades (mean pairwise F_{ST} = 0.0026), with no fixed SNP differences between populations in either comparison. There were also no F_{ST} outlier SNPs when comparing the two western subclades. Outliers were identified between the eastern and western populations, but we found no enrichment for Gene Ontology terms and outlier SNPs almost exclusively fell outside of coding regions or caused only synonymous amino acid replacements.

High-throughput sequencing has been instrumental in identifying genes underpinning migratory behavior and plumage patch coloration in several recent studies (Delmore et al. 2015, Mason and Taylor 2015, Toews et al. 2016, Aguillon et al. 2021, Louder et al. 2014). F_{ST} outliers from comparison of the eastern and western populations of American goldfinch did not coincide with candidate migration genes identified from genomic selection scans (Delmore et al. 2015) or differential gene expression across five brain regions of the Swainson's thrush (*Catharus ustulatus*, Louder et al. 2024), including candidates with known associations to circadian clock regulation in birds. However, seven outlier goldfinch genes did overlap with candidate migration genes identified from *meta*-analysis of whole-transcriptome studies in birds (presented as Supplementary Data 3 in Louder et al. 2024). Eleven American goldfinch genes with east/west F_{ST} outliers also overlapped with genes involved in melanogenesis (Poelstra et al. 2015), including two (*PLCB1* and *PLCB4*) that were significantly associated with plumage patch color differences of Northern flickers (*Colaptes auratus*, Aguillon et al. 2021). In addition to these candidate color and migration genes, we also found six outlier genes with potential roles in vocal learning and song behavior. The transcription factor *FOXP2* (forkhead box P2) is perhaps best known for its involvement in human spoken language but has also shown strong associations with vocal learning in birds (Enard et al. 2002, Vargha-Khadem et al. 2005, Cahill et al. 2021). It, along with four additional outlier genes in the American goldfinch (*CDH8*, *KLHL29*, *SYT1*, and *TENM2*) occur near hotspots for accelerated genomic regions (ARs) in vocal learning birds and are proposed as strong candidate genes for vocal-learning related selection (Cahill et al. 2021). The sixth American goldfinch outlier, *ESR1*, which encodes an estrogen receptor, was previously localized to a supergene complex in the white-throated sparrow (*Zonotrichia albicollis*) where differences in gene expression predicted territorial aggression and singing (Maney et al. 2020).

Feather coloration, song, and migratory behavior are each recognized as traits that can promote speciation via reproductive isolation arising through sexual selection and assortative mating (Bearhop, 2005; Louder, 2024; Ruegg, 2014; Tietze, 2018). Finding multiple genes with outlier SNPs for each of these traits in the American goldfinch suggests their possible relevance to the observed morphological and behavioral differences between eastern and western populations despite the low overall genomic divergence between these populations. However, the underlying basis of traits exhibiting quantitative variation, as is more likely the case for clinal variation in body size and plumage tone or

saturation in the American goldfinch, might be less easily determined (Harrison et al. 2012) and might also be attributable to environmental or dietary factors rather than to individual genetic variation alone. None of the outlier SNPs associated with candidate genes for color, song, or migration fully segregated according to the eastern versus western population of American goldfinch; however, we might not necessarily expect perfect segregation for clinally varying traits. Additionally, none of the outlier SNPs caused nonsynonymous substitutions to protein-coding sequence, indicating that differences might instead arise through noncoding regulatory changes to gene expression. Recent studies have highlighted this potential importance of regulatory sequence difference to each of these traits (e.g. plumage coloration Funk and Taylor 2019, Sin et al. 2024; song Cahill et al. 2021; migration Louder et al. 2024), with Louder et al. (2024) proposing that rapid selection of more labile traits that rely on external cues, such as migration, could stem from *trans*-acting changes to master regulators that coordinate the expression of correlated target genes. Incorporating gene expression data (e.g. Ricchetti et al. 2024) to investigate contrasts in plumage or behavioral traits therefore presents a promising future avenue for research in the American goldfinch.

4.5. Conclusions

We find clear evidence for population structuring in the American goldfinch despite relatively weak differentiation between populations and the potential confounding signatures of introgression or migration. Genome-wide analyses also indicate that past demographic events tied to historic climate cycles likely contributed to shaping current genetic diversity of this species even when accompanied by high long-term effective population size. East versus west divergence in populations of the American goldfinch echoes patterns found in many other temperate North American taxa and bolsters the view that the complex topography of the Pacific West, acting in concert with historic glacial refugia, has heavily influenced the present-day biodiversity of western North America. However, unlike many other east/west pairs of avian taxa that exhibit much greater genetic divergence and are often recognized as 'good species', shallow divergence and a lack of mtDNA lineage sorting in the American goldfinch suggest substantial ongoing migration accompanied population divergence and continues in the form of admixed individuals along a possible east/west contact zone. Notwithstanding this weak genetic differentiation, regions of the genome under divergent selection between the eastern and western populations were located near multiple genes implicated in feather color, song, and migratory behavior in other passerine birds. Further work incorporating gene expression data will help clarify whether these traits serve to reinforce population divergence in the American goldfinch and could contribute more broadly to our understanding of speciation processes in birds.

CRediT authorship contribution statement

Alison Cloutier: Writing – original draft, Visualization, Formal analysis. **David Tsz Chung Chan:** Writing – review & editing, Visualization, Methodology, Investigation, Formal analysis. **Emily Shui Kei Poon:** Writing – review & editing, Resources, Methodology, Investigation. **Simon Yung Wa Sin:** Writing – review & editing, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108216.

References

Aguillon, S.M., Campagna, L., Harrison, R.G., Lovette, I.J., 2018. A flicker of hope: genomic data distinguish Northern Flicker taxa despite low levels of divergence. Auk 135, 748–766.

Aguillon, S.M., Walsh, J., Lovette, I.J., 2021. Extensive hybridization reveals multiple coloration genes underlying a complex plumage phenotype. Proc. R. Soc. B 288, 20201805.

Alexander, D.H., Novembre, J., Lange, K., 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19, 1655–1664.

Andrews, S., 2010. FastQC: A quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Arnaiz-Villena, A., et al., 2012. Three different North American siskin/goldfinch evolutionary radiations (Genus *Carduelis*): pine siskin green morphs and European siskins in America. Open Ornithol. J. 5, 73–81.

Arnaiz-Villena, A., Gomez-Prieto, P., Ruiz-del-Valle, V., 2009. Phylogeography of finches and sparrows. In: Rechi, L.J. (Ed.), Animal Genetics. Nova Science Publishers Inc, pp. 1–54.

Avise, J.C., et al., 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18, 489–522.

Avise, J.C. (Ed.), 2000. Phylogeography: the history and formation of species. Harvard University Press, Cambridge, MA, U.S.A.

Avise, J.C., Walker, D.E., 1998. Pleistocene phylogeographic effects on avian populations and the speciation process. Proc. R. Soc. B Biol. Sci. 265, 457–463.

Barrowclough, G.F., Groth, J.G., Mertz, L.A., Gutiérrez, R.J., 2004. Phylogeographic structure, gene flow and species status in blue grouse (*Dendragapus obscurus*). Mol. Ecol. 13, 1911–1922.

Bearhop, S., et al., 2005. Assortative mating as a mechanism for rapid evolution of a migratory divide. Science 310, 502–504.

Beckman, E.J., Witt, C.C., 2015. Phylogeny and biogeography of the New World siskins and goldfinches: rapid, recent diversification in the Central Andes. Mol. Phylogenet. Evol. 87, 28–45.

Beichman, A.C., Huerta-Sanchez, E., Lohmueller, K.E., 2018. Using genomic data to infer historic population dynamics of nonmodel organisms. Annu. Rev. Ecol. Evol. Syst. 49, 433–456.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120.

Brown, J.L., et al., 2018. PaleoClim, high spatial resolution paleoclimate surfaces for global land areas. Sci. Data 5, 180254.

Brown, J.L., Bennett, J.R., French, C.M., 2017. SDMtoolbox 2.0: the next generation Python-based GIS toolkit for landscape genetic, biogeographic and species distribution model analyses. PeerJ 5, e4095.

Brunsfeld, S.J., Sullivan, J., Soltis, D.E., Soltis, P.S., 2001. Comparative phylogeography of northwestern North America: A synthesis. In: Silvertown, J., Antonovics, J. (Eds.), Integrating Ecology and Evolution in a Spatial Context. Blackwell Science, pp. 319–340.

Burbrink, F.T., et al., 2016. Asynchronous demographic responses to Pleistocene climate change in Eastern Nearctic vertebrates. Ecol. Lett. 19, 1457–1467.

Bushnell, B. (2014). BBMap: A fast, accurate, splice-aware aligner. Lawrence Berkeley National Laboratory. LBNL Report #: LBNL-7065E. Retrieved from: https://escholarship.org/uc/item/1h3515gn.

Cahill, J.A., et al., 2021. Positive selection in noncoding genomic regions of vocal learning birds is associated with genes implicated in vocal learning and speech functions in humans. Genome Res. 31, 2035–2049.

Calsbeek, R., Thompson, J.N., Richardson, J.E., 2003. Patterns of molecular evolution and diversification in a biodiversity hotspot: the California Floristic Province. Mol. Ecol. 12, 1021–1029.

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973.

Chang, C.C., et al., 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, 7.

Chauvier, Y., et al., 2022. gbif.range - an R package to generate species range maps based on ecoregions and a user-friendly GBIF wrapper. EnviDat. https://doi.org/10.16904/envidat.352.

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. FASTP: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890.

Cheviron, Z.A., Swanson, D.L., 2017. Comparative transcriptomics of seasonal phenotypic flexibility in two North American songbirds. Integr. Comp. Biol. 57, 1040–1054.

Cicero, C., et al., 2022. Deep ecomorphological and genetic divergence in Steller's jays (*Cyanocitta stelleri*, Aves: Corvidae). Ecol. Evol. 12, e9517.

Colbeck, G.J., et al., 2008. Phylogeography of a widespread North American migratory songbird (*Setophaga ruticilla*). J. Hered. 99, 453–463.

Cristofari, R., et al., 2018. Climate-driven range shifts of the king penguin in a fragmented ecosystem. Nat. Clim. Chang. 8, 245–251.

Danecek, P., et al., 2011. The variant call format and VCFtools. Bioinformatics 27, 2156–2158.

Danecek, P., et al., 2021. Twelve years of SAMtools and BCFtools. GigaScience 10, 1–4.

Delmore, K.E., et al., 2015. Genomic analysis of a migratory divide reveals candidate genes for migration and implicates selective sweeps in generating islands of differentiation. Mol. Ecol. 24, 1873–1888.

DeRaad, D.A., et al., 2022. Combining species delimitation, species trees, and tests for gene flow clarifies complex speciation in scrub-jays. Syst. Biol. 71, 1453–1470.

DeSaix, M.G., et al., 2023. Low-coverage whole genome sequencing for highly accurate population assignment: mapping migratory connectivity in the American Redstart (*Setophaga ruticilla*). Mol. Ecol. 32, 5528–5540.

DeSaix, M.G., Rodriguez, M.D., Ruegg, K.C., Anderson, E.C., 2024. Population assignment from genotype likelihoods for low-coverage whole-genome sequencing data. Methods Ecol. Evol. 15, 493–510.

Dierickx, E.G., Sin, S.Y.W., van Veelen, H.P.J., Brooke, M.D.L., Liu, Y., Edwards, S.V., Martin, S.H., 2020. Genetic diversity, demographic history and neo-sex chromosomes in the Critically Endangered Raso lark. Proc. R. Soc. B 287, 20192613.

Dutoit, L., et al., 2017. Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. Proc. R. Soc. B Biol. Sci. 284, 20162756.

eBird. 2021. eBird: an online database of bird distribution and abundance [web application]. eBird, Cornell Lab of Ornithology, Ithaca, New York. Available: http://www.ebird.org. (Data Version: 2019, Accessed: June 11, 2020).

Edwards, S.V., Robin, V.V., Ferrand, N., Moritz, C., 2022. The evolution of comparative phylogeography: putting the geography (and more) into comparative population genomics. Genome Biol. Evol. 14, p. evab176.

Enard, W., et al., 2002. Molecular evolution of *FOXP2*, a gene involved in speech and language. Nature 418, 869–872.

Escalante, T., et al., 2010. Delimitation of the Nearctic region according to mammalian distributional patterns. J. Mammal. 91, 1381–1388.

Evanno, G., Regnaut, S., Goudet, J., 2005. Detecting the number of clusters of individuals using the software structure: a simulation study. Mol. Ecol. 14, 2611–2620.

Excoffier, L., et al., 2013. Robust demographic inference from genomic and SNP data. PLoS Genet. 9, e1003905.

Excoffier, L., et al., 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. Bioinformatics 37, 4882–4885.

Francis, R.M., 2017. pophelper: an R package and web app to analyse and visualize population structure. Mol. Ecol. Resour. 17, 27–32.

Funk, E.R., Taylor, S.A., 2019. High-throughput sequencing is revealing genetic associations with avian plumage color. Auk 136, 1–7.

Gagnaire, P., 2020. Comparative genomics approach to evolutionary process connectivity. Evol. Appl. 13, 1320–1334.

GBIF, 2024a. Derived dataset from the Global Biodiversity Information Facility (GBIF.org, accessed 22 August 2024). Filtered export of GBIF occurrence data for *Spinus tristis jewetti*. Data URL: https://doi.org/10.15468/dd.dncqmn.

GBIF 2024b. Derived dataset from the Global Biodiversity Information Facility (GBIF.org, accessed 22 August 2024). Filtered export of GBIF occurrence data for *Spinus tristis pallidus*. Data URL: https://doi.org/10.15468/dd.s7h972.

GBIF, 2024c. Derived dataset from the Global Biodiversity Information Facility (GBIF.org, accessed 22 August 2024). Filtered export of GBIF occurrence data for *Spinus tristis salicamans*. Data URL: https://doi.org/10.15468/dd.b6thx6.

GBIF, 2024d. Derived dataset from the Global Biodiversity Information Facility (GBIF.org, accessed 22 August 2024). Filtered export of GBIF occurrence data for *Spinus tristis tristis*. Data URL: https://doi.org/10.15468/dd.w6xvfd.

Grabherr, M.G., et al., 2010. Genome-wide synteny through highly sensitive sequence alignment: Satsuma. Bioinformatics 26, 1145–1151.

Grinnell, J., 1897. New race of *Spinus tristis* from the Pacific Coast. Auk 14, 397–399.

Gutenkunst, R.N., Hernandez, R.D., Williamson, S.H., Bustamante, C.D., 2009. Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. PLoS Genet. 5, e1000695.

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads–a baiting and iterative mapping approach. Nucleic Acids Res. 41, e129.

Hanghøj, K., et al., 2019. Fast and accurate relatedness estimation from high-throughput sequencing data in the presence of inbreeding. GigaScience 8, 1–8.

Harrison, P.W., Wright, A.E., Mank, J.E., 2012. The evolution of gene expression and the transcriptome–phenotype relationship. Semin. Cell Dev. Biol. 23, 222–229.

Hewitt, G., 2000. The genetic legacy of the Quaternary ice ages. Nature 405, 907–913.

Hewitt, G.M., 2004. The structure of biodiversity - insights from molecular phylogeography. Front. Zool. 1, 4.

Holt, B.G., et al., 2013. An update of Wallace's zoogeographic regions of the world. Science 339, 74–78.

Huerta-Cepas, J., Serra, F., Bork, P., 2016. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. 33, 1635–1638.

Huynh, S., Cloutier, A., Chen, G., Chan, D.T.C., Lam, D.K., Huyvaert, K.P., Sato, F., Edwards, S.V., Sin, S.Y.W., 2023. Whole-genome analyses reveal past population fluctuations and low genetic diversities of the North Pacific albatrosses. Mol. Biol. Evol. 40, msad155.

IUCN. 2016. The IUCN Red List of Threatened Species. Version 2016-3. Available at: www.iucnredlist.org. Accessed on December 10, 2023.

Jaramillo-Correa, J.P., Beaulieu, J., Khasa, D.P., Bousquet, J., 2009. Inferring the past from the present phylogeographic structure of North American forest trees: seeing the forest for the genes. Can. J. for. Res. 39, 286–307.

Kalyaanamoorthy, S., et al., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780.

Kelly, R.J., Murphy, T.G., Tarvin, K.A., Burness, G., 2012. Carotenoid-based ornaments of female and male American Goldfinches (*Spinus tristis*) show sex-specific correlations with immune function and metabolic rate. Physiol. Biochem. Zool. 85, 348–363.

Kimura, M., et al., 2002. Phylogeographical approaches to assessing demographic connectivity between breeding and overwintering regions in a Nearctic−Neotropical warbler (*Wilsonia pusilla*). Mol. Ecol. 11, 1605–1616.

Klicka, J., et al., 2011. A phylogeographic and population genetic analysis of a widespread, sedentary North American bird: the Hairy Woodpecker (*Picoides villosus*). Auk 128, 346–362.

Kopelman, N.M., et al., 2015. CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol. Ecol. Resour. 15, 1179–1191.

Li, H., et al., 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079.

Li, S., et al., 2014. Genomic signatures of near-extinction and rebirth of the crested ibis and other endangered bird species. Genome Biol. 15, 557.

Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.

Li, H., Durbin, R., 2011. Inference of human population history from individual whole-genome sequences. Nature 475, 493–496.

Linnaeus, C., 1758. *Systema Naturae per regna tria naturae, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis.* Editio decima, reformata [10th revised edition].

Liu, X., Fu, Y.-X., 2015. Exploring population size changes using SNP frequency spectra. Nat. Genet. 47, 555–559.

Liu, X., Fu, Y.-X., 2020. Stairway Plot 2: demographic history inference with folded SNP frequency spectra. Genome Biol. 21, 280.

Louder, M.I.M., et al., 2024. Gene regulation and speciation in a migratory divide between songbirds. Nat. Commun. 15, 98.

Lovette, I.J., 2005. Glacial cycles and the tempo of avian speciation. Trends Ecol. Evol. 20, 57–59.

Lovette, I.J., Clegg, S.M., Smith, T.B., 2004. Limited utility of mtDNA markers for determining connectivity among breeding and overwintering locations in three neotropical migrant birds. Conserv. Biol. 18, 156–166.

Lyman, R.A., Edwards, C.E., 2022. Revisiting the comparative phylogeography of unglaciated eastern North America: 15 years of patterns and progress. Ecol. Evol. 12, e8827.

Maney, D.L., et al., 2020. Inside the supergene of the bird with four sexes. Horm. Behav. 126, 104850.

Manthey, J.D., Klicka, J., Spellman, G.M., 2011. Cryptic diversity in a widespread North American songbird: phylogeography of the Brown Creeper (*Certhia americana*). Mol. Phylogenet. Evol. 58, 502–512.

Mason, N.A., Taylor, S.A., 2015. Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. Mol. Ecol. 24, 3009–3025.

McCaslin, H.M., Heath, J.A., 2020. Patterns and mechanisms of heterogeneous breeding distribution shifts of North American migratory birds. J. Avian Biol. 2020, e02237.

McGraw, K. J. & Middleton, A. L., 2020. American Goldfinch (Spinus tristis), version 1.0. In: Birds of the World, P. G. Rodewald, Editor. Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi-org.proxy.lib.sfu.ca/10.2173/bow.amegfi.01.

Mearns, E.A., 1890. Descriptions of a new species and three new subspecies of birds from Arizona. Auk 7, 243–251.

Mi, H., et al., 2019. Protocol update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat. Protoc. 14, 703–721.

Middleton, A.L.A., 1978. The annual cycle of the American Goldfinch. Condor 80, 401–406.

Milà, B., Girman, D.J., Kimura, M., Smith, T.B., 2000. Genetic evidence for the effect of a postglacial population expansion on the phylogeography of a North American songbird. Proc. R. Soc. B Biol. Sci. 267, 1033–1040.

Milot, E., Gibbs, H.L., Hobson, K.A., 2000. Phylogeography and genetic structure of northern populations of the yellow warbler (*Dendroica petechia*). Mol. Ecol. 9, 667–681.

Minh, B.Q., et al., 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534.

Miraldo, A., et al., 2016. An Anthropocene map of genetic diversity. Science 353, 1532–1535.

Nadachowska-Brzyska, K., et al., 2015. Temporal dynamics of avian populations during Pleistocene revealed by whole-genome sequences. Curr. Biol. 25, 1375–1380.

Olson, D.M., et al., 2001. Terrestrial ecoregions of the world: a new map of life on Earth. Bioscience 51, 933–938.

Partners in Flight. 2020. Population Estimates Database, version 3.1. Available at http://pif.birdconservancy.org/PopEstimates. Accessed on December 10, 2023.

Pedersen, B.S., Quinlan, A.R., 2018. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics 34, 867–868.

Phillips, S.J., et al., 2017. Opening the black box: an open-source release of Maxent. Ecography 40, 887–893.

Poelstra, J.W., Vijay, N., Hoeppner, M.P., Wolf, J.B.W., 2015. Transcriptomics of colour patterning and coloration shifts in crows. Mol. Ecol. 24, 4617–4628.

Prescott, D.R.C., Middleton, A.L.A., 1990. Age and sex differences in winter distribution of American Goldfinches in eastern North America. Ornis Scand. 21, 99–104.

QGIS Development Team, 2021. *QGIS Geographic Information System*. Retrieved from: http://qgis.org.

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.

R Core Team, 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Ricchetti, I., et al., 2024. Gene expression variation in geographically diverse populations of two North American songbird species. Front. Bird Sci. 3, 1382657.

Robinson, J., et al., 2016. Genomic flatlining in the endangered Island Fox. Curr. Biol. 26, 1183–1189.

Ruegg, K., et al., 2014. A role for migration-linked genes and genomic islands in divergence of a songbird. Mol. Ecol. 23, 4757–4769.

Ruegg, K.C., Smith, T.B., 2002. Not as the crow flies: a historical explanation for circuitous migration in Swainson's thrush (*Catharus ustulatus*). Proc. R. Soc. B Biol. Sci. 269, 1375–1381.

Sin, S.Y.W., et al., 2024. Genetic basis and evolution of structural color polymorphism in an Australian songbird. Mol. Biol. Evol. 41, msae046.

Sin, S.Y.W., Hoover, B.A., Nevitt, G.A., Edwards, S.V., 2021. Demographic history, not mating system, explains signatures of inbreeding and inbreeding depression in a large outbred population. Am. Nat. 197, 658–676.

Smith, B.T., et al., 2017. A latitudinal phylogeographic diversity gradient in birds. PLoS Biol. 15, e2001073.

Soltis, D.E., et al., 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. 15, 4261–4293.

Spellman, G.M., Riddle, B., Klicka, J., 2007. Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to Quaternary climate change. Mol. Ecol. 16, 1055–1068.

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313.

Sughrue, K.M., Brittingham, M.C., French, J.B., 2008. Endocrine effects of the herbicide linuron on the American Goldfinch (*Carduelis tristis*). Auk 125, 411–419.

Swenson, N.G., Howard, D.J., 2005. Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. Am. Nat. 166, 581–591.

Tacutu, R., et al., 2018. Human ageing genomic resources: new and updated databases. Nucleic Acids Res. 46, D1083–D1090.

The Nature Conservancy, 2009. Global ecoregions, major habitat types, biogeographical realms and The Nature Conservancy terrestrial assessment units. GIS layers developed by The Nature Conservancy with multiple partners, combined from Olson et al. (2001), Bailey 1995 and Wiken 1986. Cambridge (UK): The Nature Conservancy. Data URL: https://geospatial.tnc.org/datasets/b1636d640ede4d6ca8f5e369f2dc368b/about.

Thomas, P.D., et al., 2022. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 31, 8–22.

Tietze, D.T. (Ed.), 2018. Bird species: how they arise, modify and vanish. SpringerOpen. https://doi.org/10.1007/978-3-319-91689-7.

Toews, D.L., et al., 2016. Plumage genes and little else distinguish the genomes of hybridizing warblers. Curr. Biol. 26, 2313–2318.

Van der Auwera, G.A., et al., 2013. From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr. Protoc. Bioinformatics 43, 11.10.1–11.10.33.

Van der Auwera, G.A., O'Connor, B.D., 2020. Genomics in the cloud: using Docker, GATK, and WDL in Terra, First edition. O'Reilly Media.

van Rossem, A.J., 1943. Description of a race of goldfinch from the Pacific Northwest. Condor 45, 158–159.

Vargha-Khadem, F., Gadian, D.G., Copp, A., Mishkin, M., 2005. *FOXP2* and the neuroanatomy of speech and language. Nat. Rev. Neurosci. 6, 131–138.

Wickham, H., 2016. ggplot2: elegant graphics for data analysis, 2nd ed. Springer International Publishing, Cham.

Wolf, J.B.W., Lindell, J., Backström, N., 2010. Speciation genetics: current status and evolving approaches. Philos. Trans. R. Soc. Lond. B Biol. Sci. 365, 1717–1733.

Yang, Z. (Ed.), 2014. Molecular Evolution: A Statistical Approach. Oxford University Press, Oxford, U.K.

Zink, R.M., 1996. Comparative phylogeography in North American birds. Evolution 50, 308–317.