[SUBMISSION DATE: NOVEMBER 30, 2025; EDITORIAL ACCEPTANCE DATE: JUNE 1, 2026]
[RUNNING HEAD] **COPY TK**
[RUNNING HEAD] E. C. Corbett et al.

RESEARCH ARTICLE

Phylogenetic relationships, hybridization, and genetic structuring in *"Cassidix"* grackles

Eamon C. Corbett,[1,2]* Robb T. Brumfield,[1] and Brant C. Faircloth[1]

[1] Museum of Natural Science and Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana, USA
[2] Cornell Lab of Ornithology, Cornell University, Ithaca, New York, USA
* Corresponding Author: eamonccorbett@gmail.com

ABSTRACT
The large grackles in the subgenus *Cassidix*—*Quiscalus mexicanus* (Great-tailed Grackle), *Q. major* (Boat-tailed Grackle), and *Q. palustris* (Slender-billed Grackle)—form a clade with a complex evolutionary history of recent divergence and secondary contact. *Quiscalus mexicanus* is widespread from the western U.S. to Peru and encompasses 2 divergent mitochondrial lineages. *Quiscalus major* is found on the Atlantic and Gulf Coasts of the U.S. and shows striking geographic variation in eye color. *Quiscalus palustris* was endemic to central Mexico but went extinct during the early 20th century. Previous research has called into question the monophyly of *Q. mexicanus* and has suggested that *Q. major* and *Q. mexicanus* may hybridize in Texas and Louisiana. Additionally, the patterns of genetic structuring among subspecies within *Q. mexicanus* and *Q. major* are unclear. We resolved many of these questions by sequencing and analyzing data from nuclear ultraconserved elements (UCEs) and mitochondrial DNA collected from samples representing all 13 subspecies in the *Cassidix* clade. Our results show no evidence of admixture between *Q. major* and *Q. mexicanus*, suggesting that if hybrids occur in the contact zone, they are infertile or rare. Our results also support the monophyly of *Q. major* and *Q. mexicanus*, with *Q. palustris* likely sister to a unified *Q. mexicanus*, contrary to evidence from mitochondrial DNA. Finally, these data resolve intraspecific relationships in the *Cassidix* complex, several of which differ from current subspecies-level taxonomy, and show that geographic breaks, not eye color differences, structure genetic variation within *Q. major*.

Keywords: Boat-tailed Grackle, *Cassidix*, Great-tailed Grackle, eye color, Icteridae, *Quiscalus*, Slender-billed Grackle, ultraconserved elements, UCEs

How to Cite
Corbett, E. C., R. T. Brumfield, and B. C. Faircloth (2026). Phylogenetic relationships, hybridization, and genetic structuring in *"Cassidix"* grackles. *Ornithology* 143:ukag000.

LAY SUMMARY
- *Quiscalus major* (Boat-tailed Grackle), *Q. mexicanus* (Great-tailed Grackle), and *Q. palustris* (Slender-billed Grackle) are three closely related species of blackbirds, the latter of which is

extinct.
- Genetic relationships among and within the species have not been previously studied using genomic data.
- We found no evidence of gene flow between *Q. major* and *Q. mexicanus*, despite reports of apparent hybrids.
- We showed that the western and eastern lineages of *Q. mexicanus* are each other's closest relatives, contrary to previous findings using mitochondrial genes.
- There is variation in eye color within *Q. major*, but we found that genetic structure within that species corresponds to geography, not to eye color.

Relaciones filogenéticas, hibridación y estructuración genética en "*Cassidix*"

RESUMEN

Las aves del subgénero *Cassidix*—*Quiscalus mexicanus*, *Q. major* y *Q. palustris*—forman un clado con una compleja historia evolutiva de divergencia reciente y contacto secundario. *Quiscalus mexicanus* se distribuye ampliamente desde el oeste de EUA hasta Perú e incluye dos linajes mitocondriales divergentes. *Quiscalus major* se encuentra en las costas atlántica y del golfo de EUA y presenta una marcada variación geográfica en el color de los ojos. *Quiscalus palustris* era endémico del centro de México, pero se extinguió a comienzos del siglo XX. Investigaciones previas han puesto en duda la monofilia de *Q. mexicanus* y han sugerido que *Q. major* y *Q. mexicanus* podrían hibridarse en Texas y Luisiana. Además, los patrones de estructuración genética entre las subespecies de *Q. mexicanus* y *Q. major* no están claros. Resolvimos muchas de estas cuestiones mediante la secuenciación y el análisis de datos provenientes de elementos ultraconservados nucleares y ADN mitocondrial obtenidos de muestras representativas de las 13 subespecies del clado *Cassidix*. Nuestros resultados no muestran evidencia de mezcla genética entre *Q. major* y *Q. mexicanus*, lo que sugiere que, si existen híbridos en la zona de contacto, son infértiles o raros. Nuestros resultados también respaldan la monofilia de *Q. major* y *Q. mexicanus*, siendo *Q. palustris* probablemente el taxón hermano de un *Q. mexicanus* unificado, en contraste con la evidencia derivada del ADN mitocondrial. Finalmente, estos datos resuelven las relaciones intraespecíficas dentro del complejo *Cassidix*, varias de las cuales difieren de la taxonomía actual a nivel de subespecies, y muestran que las discontinuidades geográficas, y no las diferencias en el color de los ojos, estructuran la variación genética dentro de *Q. major*.

Palabras clave: *Cassidix*, color de ojos, elementos ultraconservados, Icteridae, *Quiscalus*

INTRODUCTION

Careful study of recently diverged taxa can help us understand biogeography, the dynamics of speciation, and the patterns by which reproductive barriers develop. Subspecies descriptions serve as phenotype-based hypotheses regarding the relationships among populations of a species across the landscape. These are often supplemented by analyses using mitochondrial genes, which partially reconstruct the relationships among populations and species through the lens of a single, generally non-recombining locus. However, phenotypic differences and mitochondrial analyses do not necessarily provide a complete picture of the evolutionary history of a clade (Edwards and Bensch 2009, Toews and Brelsford 2012), and range-wide phylogeographic studies using genome-wide markers can provide greater insight into historical patterns of

population divergences, the occurrence of introgression between taxa, and the correspondence of genetic differentiation with phenotypic differences (Firneno et al. 2020, Andersen et al. 2021, Del-Rio et al. 2021, DeRaad et al. 2023).

Quiscalus major (Boat-tailed Grackle) and *Q. mexicanus* (Great-tailed Grackle) are closely related species of New World Blackbirds (family Icteridae) that diverged <1 mya (Powell et al. 2008). Their combined range spans the Americas: *Q. major* is found in coastal marshes along the Atlantic and Gulf Coasts of the United States, and *Q. mexicanus* is widespread across the southwestern U.S., Mexico, Central America, and occurs locally in South America in coastal Colombia, Venezuela, and Peru (Post et al. 2020, Johnson and Peer 2022; Figure 1). A third member of the species complex, *Q. palustris* (Slender-billed Grackle), inhabited marshes in central Mexico before its extinction around 1910, and it is the only species of Icteridae to go extinct during modern times (Dickerman 1965, Peterson 1998, Jaramillo and Burke 1999, Haemig 2009, Muñoz González and Carrillo Martínez 2025). These 3 species were once considered part of the genus *Cassidix*, along with *Q. nicaraguensis* (Nicaraguan Grackle; Selander and Giller 1961). Recent work has shown that the complex is embedded within *Quiscalus*, and that *Q. nicaraguensis* is not closely related (Powell et al. 2008). For simplicity, we refer to the clade comprising *Q. major*, *Q. mexicanus*, and *Q. palustris* (but not *Q. nicaraguensis*) as the *Cassidix* grackles, with *Cassidix* serving as a subgenus of *Quiscalus*.

Phylogenetic relationships among members of the *Cassidix* grackles have long been a subject of study (Brooks 1928, Selander and Giller 1961, Avise and Zink 1988, DaCosta et al. 2008, Powell et al. 2008). *Quiscalus major* and *Q. mexicanus* were once considered a single species, but they were split based on evidence of reproductive isolation in their narrow but expanding zone of sympatry in coastal Texas and southwest Louisiana (Selander and Giller 1961). That contact zone is relatively recent: *Q. mexicanus* has expanded its range northward since the start of the 20th century, coming into contact with *Q. major* in coastal Texas during the 1920s or 1930s (Selander and Giller 1961, Wehtje 2003). Despite evidence for reproductive barriers between the two species, morphological hybrids were reported from Louisiana and Texas (Pratt 1991). However, neither the presence of admixture nor the occurrence of early generation hybrids have been tested with genetic data.

Relationships among populations of *Q. mexicanus* (Figure 1), which comprises 8 named subspecies, are also notable (DaCosta et al. 2008, Powell et al. 2008). Based on trees built using cytochrome *b* (CYTB) and NADH dehydrogenase subunit 2 (ND2) genes, two deeply divergent lineages within *Q. mexicanus* may exist: a "western" lineage that inhabits areas of California, Arizona, Nevada, and northwest Mexico (containing the subspecies *Q. m. nelsoni* in the Southwest U.S. and Sonora and *Q. m. graysoni* from Sinaloa; Sclater 1884, Ridgway 1901, Phillips 1950), and an "eastern" lineage that inhabits the rest of the species' range, from Arizona east to Louisiana and south to Peru (DaCosta et al. 2008, Powell et al. 2008). This eastern lineage comprises the remaining 6 subspecies (Johnson and Peer 2022): *Q. m. monsoni* (eastern Arizona, New Mexico, and western Texas, south into Chihuahua; Phillips 1950), *Q. m. prosopidicola* (Central Texas east to Louisiana and south to Tamaulipas; Lowery 1938), *Q. m. mexicanus* (central and east Mexico to Nicaragua; Hellmayr 1937, Phillips 1950, Johnson and Peer 2022), *Q. m. loweryi* (Yucatán Peninsula; Dickerman and Phillips 1966), *Q. m. obscurus* (coastal west Mexico from Nayarit to Guerrero; Nelson 1900), and *Q. m. peruvianus* (Costa Rica to Peru; Chapman 1917, Todd and Carriker 1922, Hellmayr 1937). Surprisingly, these mitochondrial DNA studies (DaCosta et al. 2008, Powell et al. 2008) have suggested that the western lineage is sister to the extinct *Q. palustris*, whereas the eastern lineage is sister to *Q.*

major, calling into question the monophyly of *Q. mexicanus*. Furthermore, the western and eastern lineages were once allopatric, but range expansion has brought them into contact—they are now broadly sympatric in the southwest U.S. and appear to interbreed, based on the presence of intermediate phenotypes and discordance between mitochondrial haplotypes and phenotypes (Phillips 1950, Wehtje 2004). However, because mitochondria do not generally recombine (Kraytsberg et al. 2004), mitochondrial markers do not provide a complete picture regarding gene flow between lineages, and whether genome-wide nuclear markers show a similar pattern of non-monophyly for *Q. mexicanus* or suggest admixture between eastern and western lineages is unknown.

Relationships among the four subspecies of *Q. major* (Figure 1), found along the Gulf and Atlantic Coasts of the United States, are also interesting. The boundaries between subspecies are unclear because their relationships could not be resolved using mtDNA (DaCosta et al. 2008), *Q. m. alabamensis* individuals have not been included in previous genetic studies, and contact zones may exist between taxa in Mississippi (*major-alabamensis*) and northeast Florida (*westoni-torreyi*) (Stevenson 1978). *Quiscalus major* subspecies also show marked geographic variation in eye color (Stevenson 1978, Corbett et al. 2024, 2025). This variation is thought to follow an alternating, "leapfrog" pattern where, as adults, Atlantic coast *Q. m. torreyi* have yellow eyes (Harper 1935), Florida *Q. m. westoni* have brown eyes (Sprunt 1934), *Q. m. alabamensis* from Alabama and Mississippi have yellow eyes (Stevenson 1978), and nominate *Q. m. major* from Louisiana and Texas generally have brown eyes (Selander and Giller 1961, Stevenson 1978). However, uncertainty remains: eye color of *Q. m. major* is variable (Selander and Giller 1961, Pratt 1974) and light-eyed birds have been observed in populations of this generally dark-eyed subspecies (McIlhenny 1934, Pratt 1974).

Three primary scenarios could explain the relationship between geography, eye color, and genetic variation in *Q. major*. The first is that the difference between light and dark eyes reflects the primary genetic division within the species, with *Q. m. torrey*i and *Q. m. alabamensis* forming a light-eyed clade (as suggested by Clements et al. 2024) and *Q. m. major* and *Q. m. westoni* a dark-eyed clade. The second is that the species is divided into 2 geographically structured clades, with *Q. m. major* (dark eyes) and *Q. m. alabamensis* (light eyes) united in a western clade that is sister to an eastern clade uniting *Q. m. westoni* (dark eyes) and *Q. m. torreyi* (light eyes). Finally, it is possible that eye color does not correspond to population structure and that genetic breaks between populations do not correspond to named subspecies or eye color phenotypes.

Here, we investigate the evolutionary relationships among *Cassidix* grackles by sequencing ultraconserved elements (UCEs) (Faircloth et al. 2012) to genotype thousands of SNPs from across the nuclear genomes of 115 individuals spanning the ranges of extant *Q. mexicanus* and *Q. major* populations, as well as 1 individual representing the extinct *Q. palustris*. Our sampling consists of modern (tissue) and historical (toepad) samples, including individuals from all 3 species and all 13 currently recognized subspecies (Clements et al. 2024). We use these data to assess the (1) monophyly of *Q. mexicanus* and *Q. major* and the presence of admixture between them; (2) relationships among lineages within *Q. mexicanus*; and (3) relationships among the 4 described subspecies of *Q. major*, including whether they cluster by geography or eye color.

METHODS

Sampling

We used searches of the VertNet specimen database to identify museum specimens of grackles from across the range of *Q. mexicanus* and *Q. major*, and we requested frozen or ethanol-preserved tissue or toepad subsamples of these specimens from 9 institutions: the American Museum of Natural History, the Natural History Museum of Los Angeles County, the Louisiana State University Museum of Natural Science, the Museum of Comparative Zoology, the Museum of Southwestern Biology, the Museum of Vertebrate Zoology, the Florida Museum of Natural History, the National Museum of Natural History, and the Burke Museum of Natural History and Culture. We collected additional grackle specimens in Louisiana, Mississippi, and Alabama between 2020 and 2023 following IACUC protocols 18-054 and 21-042, and we deposited study skins and multiple tissues in the collection of the LSU Museum of Natural Science. We photographed and recorded eye color for all specimens we collected in the field. We designed the sampling scheme to include all described taxa within the complex and multiple individuals spanning the range of each subspecies when that was possible, as well as to sample across interspecific and intraspecific contact zones in Louisiana and the Southwest U.S.

DNA Extraction, Library Preparation, Enrichment, and Sequencing

We extracted total DNA from tissues using DNeasy blood and tissue kits (QIAGEN N.V.) and from toepad samples using a phenol-chloroform protocol (Tsai et al. 2020). After quantifying extracts with a Qubit fluorometer (ThermoFisher Scientific, Inc.) and standardizing each to 10 ng μL^{-1}, we sheared extracts from tissues using a sonicator (Qsonica L.L.C.) to a target size range of 400–600 base pairs (bp). We did not shear toepad extracts because these were already fragmented due to DNA degradation (McCormack et al. 2016). We prepared dual-indexed whole genome sequencing libraries from 200 to 250 ng of each extract using KAPA Hyper Prep kits (F. Hoffmann-La Roche AG) at half the standard reaction volumes that integrated a SPRI bead-based size selection step (Rohland and Reich 2012) when used with tissues, and we amplified libraries in 50 µL reactions using 6–16 polymerase chain reaction (PCR) cycles (6–12 for tissues, 12–16 for toepads) and custom indices (Glenn et al. 2019). Following amplification, we performed a round of SPRI-based cleanup for all libraries, we quantified each library using a Qubit fluorometer, and we combined libraries into pools of 7–9 samples (55.5–71.5 ng each library), keeping libraries from tissues and toepads in separate pools. We enriched library pools following the Arbor Daicel Biosciences MyBaits Expert protocol v4.1 using 2, improved sets of UCE baits derived from Faircloth et al. (2012). These improved bait sets were specifically designed for libraries generated from avian tissues or toepads by: dropping underperforming UCEs targeted for enrichment (n = 626 loci) from the original set of 5,060 UCEs, including baits designed from *Gallus gallus* (domestic chicken) (GCF_000002315.6) and *Taeniopygia guttata* (Zebra Finch) (GCA_009859065.2), and reducing the bait length from 120-mer to 80-mer in the toepad-specific bait set. The new bait sets also standardized the number of baits targeting each UCE locus in tissue libraries to 2 (n = 200 loci) or 4 (n = 4,234 loci) and in toepad libraries to 10 (n = 200 loci) or 20 (n = 4,234 loci). These bait sets are available from Arbor Daicel Biosciences as MyBaits Expert UCE kits 5Kv2A (tissues) or 5Kv2B (toepads) and the bait design files are available from https://github.com/openwings-project/baits.

After enrichment, we performed 16 cycles of PCR amplification, and we cleaned the enriched, amplified pools with a SPRI bead cleanup or a Qiagen GeneRead Size Selection Kit. We ran enriched pools on a Bioanalyzer (Agilent, Inc.) to ensure appropriate fragment size distributions, and we quantified pools using a KAPA qPCR kit (F. Hoffmann-La Roche AG). We

combined enriched pools at different ratios to achieve ~4 M reads for each tissue library and ~6 M reads for each toepad library, and we generated data from sequencing pools using 2 partial lanes of paired-end, 150 bp Illumina sequencing on either a NovaSeq 6000 (Novogene, Inc) or a Novaseq X (Azenta, Inc.). Along with the reads representing the ingroup individuals, we downloaded publicly available sequence reads from an enriched UCE library (SRR8236597; Oliveros et al. 2019) of an *Icterus cucullatus* (Hooded Oriole) individual to serve as the outgroup. We used illumiprocessor (Faircloth 2011), which is a wrapper around trimmomatic (Bolger et al. 2014), to trim adapters and low-quality bases from the raw reads in each library, and we assembled the trimmed reads using SPAdes (v3.14.1; Prjibelski et al. 2020) within PHYLUCE (v1.7.0; Faircloth 2016). Because UCE libraries often contain off-target reads from the mitochondrial genome (Meiklejohn et al. 2014, Raposo Do Amaral et al. 2015), we screened assembled libraries for contamination using the COI locus from a *Q. mexicanus* individual in the BOLD database (STRIBC1879: BSPBB418-12; Ratnasingham et al. 2024) using the mitochondrial barcode screening functions in PHYLUCE.

SNP Calling

Because we collected sequencing data across recently diverged (1–2 mya; Powell et al. 2008) species and populations, we chose to follow a workflow for phylogenetic and phylogeographic inference that used variant (single nucleotide polymorphism [SNP]) calls rather than depending on the assembly and alignment of consensus sequences. This approach offered two benefits: (1) it allowed us to perform all analyses with a set/subset of the same data (e.g., the SNP calls) versus one set of consensus contig alignments for phylogenetic analyses and another set of SNP calls for phylogeographic analyses, and (2) it allowed finer-grained and more accurate filtering of missing and incorrectly called variant sites that often exist in toepad libraries and cause problems (see Salter et al. [2022] for a more complete description of these issues).

To call SNPs, we used PHYLUCE (v1.7.3; Faircloth 2016) to extract reference UCE sequences (± 1000 bp flank) from haplotype 2 of the *Q. major* genome assembly (GCA_047301305.1; Corbett et al. 2025) following Tutorial III of the PHYLUCE documentation (Faircloth 2024). Then, we used *snpArcher* (commit 8a0921d; Mirchandani et al. 2024) with the low coverage configuration to call SNPs against the reference UCE sequences by integrating *fastp* (Chen et al. 2018) to trim sequencing adapters; SAMtools (Li et al. 2009) and BWA-MEM (Li 2013) to map trimmed reads to the reference contigs; *sambamba* (Tarasov et al. 2015) for PCR duplicate identification and removal; *genmap* (Pockrandt et al. 2020), d4tools (Hou et al. 2021), *mosdepth* (Pedersen and Quinlan 2018) and *bedtools* (Quinlan and Hall 2010) to compute mappability and callable regions of the reference UCEs, GATK (McKenna et al. 2010) for SNP calling, and VCFtools (Danecek et al. 2011) to concatenate GATK results. We used several custom Python (Python Software Foundation 2020) scripts to summarize output from *fastp* and SAMtools, and we used R (4.5.1) to compare differences (alpha = 0.05) between certain results for the ingroup taxa (after removing the *I. cucullatus* individual) using Welch's two sample *t*-test or the Wilcoxon rank sum test after testing the data for normality using the Shapiro-Wilk test.

SNP Filtering

SnpArcher applies a default set of filters to variant calls using GATK VariantFiltration and following the GATK best practices (Van Der Auwera et al. 2013), including the annotation of variants with: low-quality score (QUAL), low quality-by-depth (QD), low mapping quality (MQ), and high strand bias (FS). To generate our primary dataset from this file, we removed

variants annotated with these filters and used VCFtools to apply additional filters that excluded variants having a coverage depth <10, indels and non-biallelic SNPs, and sites with a minor allele count (MAC) < 3 (Linck and Battey 2019). We also used VCFtools to exclude sites that appeared to violate Hardy-Weinberg equilibrium due to an excess of heterozygotes, and we removed sites that were missing data for >25% of individuals. Although the 25% threshold we used is somewhat arbitrary and choices for this parameter vary by study and analytical purpose (Hemstrom et al. 2024), it strikes a reasonable balance between lenient missingness filters that include a large number of SNPs lacking data for most individuals and more stringent filters that can exclude informative variants and truncate representation of the mutational spectrum (Huang and Knowles 2016).

Because we were interested in understanding the structure of *Q. mexicanus* and *Q. major* populations, we created 2 additional SNP datasets suitable for these types of analyses. First, for principal components analyses (PCA) of population structure, we removed the outgroup individual and subsampled the primary dataset to a single SNP per locus using the VCFtools "-*thin*" command, because PCA results are sensitive to linkage disequilibrium between sites. Second, for maximum likelihood-based clustering analysis, we filtered the raw variant calls output by *snpArcher* to include individuals from 7 well-sampled subspecies (*Q. mexicanus*: *nelsoni*, *monsoni*, *prosopidicola* – including *nelsoni/monsoni* and *prosopidicola/monsoni*; *Q. major*: *major*, *alabamensis*, *westoni*, *torreyi*) that are in contact or parapatric across the southern U.S., and we excluded populations represented by low sample sizes (<6 individuals). We also excluded individuals represented by toepads to reduce the amount of missing data. Then, we used VCFtools to apply the same filters that we used for the primary dataset, except that we thinned the dataset to a single SNP per UCE locus to minimize the effects of linkage disequilibrium, and we removed sites that were missing data for >5% of individuals to minimize the likelihood our results would be biased by missing data and to help ensure that our remaining SNPs were of high quality. The 5% cutoff that we used is somewhat arbitrary, although it strikes a reasonable balance between matrix completeness and site retention, particularly when the data included in the matrix come from only high-quality sources, like well-preserved tissues. Although the program we used for subsequent clustering analysis (PopCluster) can account for uneven sampling across populations (Wang 2017, 2024), small samples can pose problems for similar implementations of population structure analyses (Puechmaille 2016, Wang 2017, 2024), and the minimum sample size required by PopCluster to represent each population is not clear. As a result, we wanted to focus this analysis on well-sampled populations and contact zones, and we did not include several subspecies from Mexico southward that were represented by few (1–5) individuals with sparse geographic sampling (e.g., *Q. mexicanus graysoni, Q. m. mexicanus, Q. m. loweryi, Q. m. obscurus, Q. m. peruvianus, Q. palustris*).

We generated summary statistics for the VCF file representing each data set using a combination of VCFtools and bcftools (v1.21, Danecek et al. 2021), and we compared differences (alpha = 0.05) between certain results from tissues and toepads in R using Welch's two sample *t*-test or the Wilcoxon rank sum test after testing the data for normality using the Shapiro-Wilk test.

Mitochondrial Gene Extraction

Previous studies of *Cassidix* used mitochondrial DNA loci to examine relationships across the group. We were interested in examining how a phylogeny inferred from nuclear DNA (UCE SNPs) differed from phylogenies inferred from mitochondrial gene sequences for many of the

same individuals. We used the MitoFinder container (v1.4.2; Allio et al. 2020) along with an annotated mitochondrial genome assembly for *Q. major* (Corbett et al. 2025) to identify and annotate mitochondrial contigs in the UCE assemblies that we generated. Because most of the resulting mtDNA assemblies were fragmented, we wrote a custom Python script to extract the sequences for the 2 mitochondrial loci, ND2 and CYTB, that were used in previous studies of the clade (DaCosta et al. 2008, Powell et al. 2008). We removed fragmentary sequences to ensure that each gene in every individual was represented by a single contig. We also downloaded a CYTB sequence of *Q. palustris* from Genbank (FJ389557.1; Powell et al. 2008) because we did not obtain many off-target mtDNA reads from our *Q. palustris* library. We aligned ND2 and CYTB sequences using MAFFT (Katoh et al. 2002). We did not concatenate loci prior to phylogenetic analysis (see below), because we wanted to replicate previous analyses to the greatest degree possible. We visually inspected alignments and calculated the average length of each locus using Jalview (v2.11.4.1; Waterhouse et al. 2009).

Phylogenetic Analyses

Prior to phylogenetic analysis, we concatenated SNPs in the primary data set using vcf2phylip (v2.8; Ortiz 2021), and we inferred a maximum likelihood (ML) phylogeny from concatenated SNP data using a combined analysis in IQTree (v2.2.2.6; Minh et al. 2020) that identified the best site-rate-substitution model using model-finder plus (Kalyaanamoorthy et al. 2017) with the ascertainment bias correction flag (MFP+ASC), inferred the best-fitting ML tree given the data, generated 100 standard bootstrap replicates (Felsenstein 1985) from the alignment, and reconciled the bootstrap replicates with the best ML tree. We rooted the resulting tree on *I. cucullatus*, and we collapsed branches having bootstrap support <50% using the *Interactive Tree of Life* web application (iTOL; v7; Letunic and Bork 2024). We refer to this as the maximum likelihood analysis.

We also inferred a phylogeny from the primary dataset using SVDQuartets (Chifman and Kubatko 2014), a coalescent approach which is implemented in Paup* (Swofford 2003). We ran the coalescent analysis by evaluating quartets exhaustively (*evalq = all*), performing 100 bootstrap replicates, and reconciling the bootstrap replicates with the coalescent tree. We rooted the tree on *I. cucullatus*, and we used iTOL to collapse branches having bootstrap support <50%. We refer to this as the coalescent analysis.

Finally, we inferred a ML phylogeny for each mitochondrial locus, independently, using the same general IQTree approach that evaluated each alignment for the best-fitting site-rate-substitution model using ModelFinder Plus and generated 100 standard bootstrap replicates for each locus. We rooted the resulting mitochondrial trees on *I. cucullatus* and collapsed branches having bootstrap support <50% using iTOL. We refer to these as the mitochondrial maximum likelihood analyses.

Principal Component Analyses of Population Structure

We used the *--pca* function of PLINK (v1.9.0-b.7.6; Chang et al. 2015) along with the SNP dataset formatted for PCA analyses to conduct a set of hierarchical PCAs that examined the population structure of the *Cassidix* complex. Specifically, we performed an initial PCA of SNP data that included all individuals, to examine how the species clustered within PC space, and then we conducted 2 independent PCAs for each species group within the complex: one analysis that included all *Q. major* individuals and a second analysis that included all *Q. mexicanus* individuals and the *Q. palustris* individual. We visualized the results for each analysis by

inputting the eigenvalue and eigenvector files created by PLINK to ggplot2 (v3.5.1; Wickham 2016) with R (v4.3.3; R Core Team 2024) in RStudio (v.2023.12.1.402; Posit Team 2023), and set the *x*-axis and *y*-axis scales to be equal with the command *coord_fixed*(1).

Maximum-Likelihood Analysis of Population Structure

We performed clustering analyses of population structure using PopCluster (v1.5; Wang 2022, 2024) which is a computationally efficient program using the mixture model of Pritchard et al. (2000) to conduct ML analyses by simulated annealing (Kirkpatrick et al. 1983). Specifically, we input variant sites selected for clustering analyses to PopCluster to examine the degree of admixture among the 7 well-sampled subspecies of *Q. mexicanus* and *Q. major*. Prior to analysis, we converted the variant call format (VCF) file representing these SNPs to pedigree (PED) format using PLINK to recode the variant calls, set a constant family ID (--const-fid 1), and allow unrecognized chromosome codes (--allow-extra-chr). Then, we analyzed the resulting PED file in PopCluster using an admixture model with unequal allele frequencies, no scaling, and 10 replicate analyses for each level of *K* between 1 and 8. We selected this range of *K* values to evaluate different population substructuring scenarios ranging from panmixia ($K = 1$) to slightly more structuring than the number of named subspecies included in the dataset (i.e., named subspecies = 7, so we evaluated $K = 8$). We selected the best run (by log-likelihood value) for each *K* and plotted these results using the *FSTruct* (Morrison et al. 2022) package and *ggplot2* in RStudio.

RESULTS

Sampling, DNA Extraction, Library Enrichment, and Sequencing

We received tissue loans for 73 specimens and toepad loans for 19 specimens (collection dates ranging from 1904 to 1983), and we collected 37 additional *Q. major* specimens (Supplemental Table 1). Illumina sequencing yielded an average of 3.4 M read pairs (SD: 0.9 M) for tissue libraries and 4.4 M read pairs (SD: 3.0 M) for toepad libraries. Thirteen toepad samples did not produce libraries of sufficient quality for sequencing or received an insufficient number of sequencing reads (< 100 K) for analysis, and our contamination screen suggested one tissue sample that we received was from a bird that was not a grackle. We removed these samples from the study. Post-hoc, general linear mixed-effect model analysis of toepad library performance (success or failure) using *lme4* (Bates et al. 2015) in R with species as a random effect and sex, specimen age, and source (museum) as fixed effects did not suggest ($P > 0.05$) these parameters affected library success. The same was true when we included interactions among these fixed effects.

Our final sample included 115 ingroup individuals representing the 3 species (*Q. major*, $n = 64$; *Q. mexicanus*, $n = 50$; and *Q. palustris*, $n = 1$) and 13 currently recognized subspecies (Clements et al. 2024) within the *Cassidix* grackle clade ($n = 1$–26 per subspecies; Figure 1) as well as 1 outgroup sample representing *I. cucullatus*.

UCE SNP Calling and Filtering

During the SNPArcher workflow, *fastp* removed an average of 5.4 K (SD: 3.7 K) low-quality sequences from each library. After adapter-trimming, the average read length was 146 bp (SD: 2) for tissues and 91 bp (SD: 6) for toepads. BWA-MEM mapped a higher proportion of reads to the reference sequences for tissues (72%, SD: 11) than for toepads (65%, SD: 5), although the

difference was not statistically significant ($p = 0.11$). Because we used a reference file containing only UCE loci and flanking regions, these lower mapping rates likely reflect the failure to map off-target reads from parts of the genome that are not UCEs. The estimated duplication rate was similar ($p = 0.42$) between tissues (mean = 33%, SD: 12) and toepads (mean = 38%, SD: 3). The number of UCE loci recovered was also similar ($p = 0.34$) across tissues and toepads (mean = 4,241), but the length of the loci recovered ("covered bases") was longer ($p = 0.0005$) in tissues (mean = 1,190 bp, SD:133.2) and shorter in toepads (mean = 820 bp, SD:122). The mean sequencing depth across all loci was high (>20x), although there were differences ($p = 0.04$) between tissues (42x, SD: 12) and toepads (22x, SD: 17).

SNPArcher identified 187,294 variants in the "raw" output file for 116 individuals, which we reduced to a total of 19,213 SNPs in our primary data set by applying additional filters (>10x depth, biallelic, no HWE excess, etc.). These SNPs spanned a total of 4,019 UCE loci (mean = 5 SNPs/locus, SD: 3) and their average depth of coverage was 49x (SD: 12). Coverage of SNPs in tissues and toepads was similar ($p = 0.3$). The average number of missing SNPs (per individual) across the primary data set was 897 (4.7%; SD: 1,936), and the count of missing SNPs differed ($p = 6.2e^{-05}$) between toepads (mean = 7,177, SD: 3,887) and tissues (mean = 555, SD: 986).

The dataset that we prepared for PCA included 4,019 SNPs (one per UCE locus) and 115 individuals (it removed the outgroup individual), the average depth of coverage per individual in this data set was 37x (SD: 10), and the average number of missing SNPs (per individual) was 298 (SD: 528). The dataset that we prepared for clustering analysis included 3,794 SNPs (one per UCE locus) and included 99 individuals representing populations of: (1) *Q. mexicanus nelsoni*, *Q. mexicanus monsoni*, and *Q. mexicanus prosopidicola* from California, Arizona, Nevada, New Mexico, Texas, Louisiana, as well as Sonora and Tamaulipas, Mexico, and (2) all 4 subspecies of *Q. major*, ranging from Texas to New York. The average depth of coverage per individual in the clustering data set was 46x (SD: 9), and the average number of missing SNPs (per individual) was low (mean = 24, SD: 71).

Mitochondrial Gene Extraction

Using MitoFinder, we were able to assemble fragmentary ($n = 45$) or complete ($n = 44$) mitogenomes from the 116 enriched libraries. We were unable to identify components of the mitochondrial genome in the remaining libraries, which included all libraries prepared from toepads. From the successful assemblies, we were able to extract 64 ND2 sequences and 81 CYTB sequences having average lengths of 948 bp for ND2 and 1,065 bp for CYTB. Mafft produced an alignment for (1) ND2 that was 1,041 characters in length and included 38 parsimony informative sites, and (2) CYTB that was 1,251 characters in length and included 61 parsimony informative sites.

Phylogenetic Analyses

[THIRD-LEVEL HEADING] Maximum likelihood analysis

IQTree identified the site-rate substitution model TVM+F+ASC+R4 (a transversion model with AG = CT, unequal empirical base frequencies, ascertainment bias correction, and a FreeRate model with 4 categories of rate heterogeneity across sites) as the best fitting given the data. Upon initial examination of the phylogeny, the results indicated 2 specimens collected in Louisiana were misidentified: a putative hybrid *Q. mexicanus* x *Q. major* (LSUMNS B-73355) and a putative female *Q. mexicanus* (LSUMNS B-96004). After examining the specimen tags and

skins of each, we concluded that these individuals should be assigned to *Q. major major*, and we corrected their labels in all results (e.g. see Supplementary Material Table S1).

The ML phylogeny we inferred (Figure 2; Supplementary Material Figures S1–S2) divides the *Cassidix* grackle complex into a monophyletic *Q. major* and a clade containing *Q. mexicanus* and *Q. palustris*. Within the latter clade, we resolve *Q. palustris* as sister to a monophyletic *Q. mexicanus*. Therefore, the deepest splits within the complex correspond to currently recognized species boundaries, and we find no evidence for non-monophyly of either *Q. mexicanus* or *Q. major*.

Within species, bootstrap support values are generally low and well-supported clades are relatively few, but those that we resolved generally correspond to geography or currently recognized subspecies limits. For example, within *Q. mexicanus*, we resolved 4 moderately supported clades, although the relationships among them are less clear (Figure 2). The largest, Clade A, includes individuals from western Mexico (Sinaloa and Sonora) and the southwestern U.S. (California, Nevada, Arizona, New Mexico, and West Texas). Nested within Clade A is a smaller subclade (Supplementary Material Figure S1) that corresponds closely to the western lineage previously identified using mitochondrial data and encompassing the distinctively small *Q. m. nelsoni* and *Q. m. graysoni* subspecies. Clade A also includes samples from farther east (e.g., New Mexico and Texas), spanning the range of *Q. m. monsoni*. This pattern seems to be the result of contact and gene flow between *Q. mexicanus* clades in the southwestern U.S., which we discuss below. Clade B unites individuals from Louisiana and Texas (except for far west Texas) corresponding to *Q. m. prosopidicola*. Clade C includes Mesoamerican birds that inhabit Central and Southern Mexico, Guatemala, Costa Rica, and Panama: *Q. m. mexicanus*, *Q. m. loweryi* from the Yucatan, and *Q. m. peruvianus* individuals from Central America. Finally, 2 individuals representing *Q. m. obscurus* from Guerrero form a well-supported Clade D. Two individual samples had uncertain relationships with these clades. The first was a bird collected from southern Tamaulipas, Mexico, near the contact zone between *Q. m. prosopidicola* (Clade B) and *Q. m. mexicanus* (Clade C), that was resolved as sister to Clade B with weak support. The second was the southernmost sample of *Q. m. peruvianus* from Tumbes, Peru. It was resolved (also with low support) as sister to all remaining *Q. mexicanus*, rather than as a member of Clade C, which included the other individuals of *Q. m. peruvianus* (from Costa Rica and Panama).

Within *Q. major*, we resolved a well-supported "western" clade that includes all samples from Texas, Louisiana, Mississippi, and Alabama, representing *Q. m. major* and *Q. m. alabamensis* (Figure 2). Interestingly, within that clade there are 2 well-supported subclades: one that includes all samples from Alabama and one that includes all samples from Texas, Louisiana, and Mississippi. All Mississippi samples were collected from the described range of *Q. m. alabamensis*, yet these Mississippi birds were united with *Q. m. major* individuals collected from Louisiana and Texas rather than with *Q. m. alabamensis* individuals collected from Mobile Bay, Alabama (the type locality), calling into question the described range of *Q. m. alabamensis*. Atlantic Coast birds, from South Carolina to New York, formed a well-supported clade corresponding to *Q. m. torreyi* (Supplementary Material Figure S1), nested within a broader "eastern" clade that also included *Q. m. westoni* individuals from Florida (Figure 2).

[THIRD-LEVEL HEADING] Coalescent analysis

The results of the coalescent analysis (Supplementary Material Figures S3–S5) were largely congruent with the ML analysis, albeit with varying support for certain nodes. The most notable difference was that, unlike the ML analysis, the coalescent analysis did not resolve the placement

of *Q. palustris* with certainty, beyond placing it in a clade with *Q. mexicanus*. As in the ML analysis, the coalescent analysis also did not resolve the placement of the *Q. m. peruvianus* individual from Peru.

All 4 of the main clades within *Q. mexicanus* identified by the ML analysis were also supported by the coalescent analysis, but in some cases, the presence of intermediate or admixed individuals seems to have complicated the picture. For example, the coalescent analysis showed only weak support (53% bootstrap) for the Clade A resolved by the ML analyses, but strong support (98% bootstrap) for a clade consisting of all Clade A individuals except for one, an apparent *Q. m. monsoni-prosopidicola* intermediate from West Texas. Both analyses resolved a well-supported *Q. m. prosopidicola* (Clade B) and a well-supported Mesoamerican Clade C, although the putative *Q. m. prosopidicola-mexicanus* intermediate from Tamaulipas, Mexico grouped the with former in the ML analysis and the latter in the coalescent analysis. Clade D, comprising two individuals of *Q. m. obscurus*, was supported by both analyses: however, the coalescent analysis placed it sister to the geographically adjacent Clade C, whereas the ML analysis did not resolve its affinities.

Mitochondrial Maximum-Likelihood Analyses

IQTree identified HKY+F+I as the best site-rate substitution model for CYTB and TPM2u+F+G4 as the best model for ND2, and post-hoc analyses of bootstrap replicates for both loci indicated that that they had converged. Many branches in the 2 mitochondrial ML phylogenies (ND2 and CYTB) were not well supported (Supplementary Material S6–S7), but the clades with reasonable support corresponded closely to previous mitochondrial studies of the *Cassidix* complex (Figure 3A–D) although they conflict with the ML and coalescent phylogenies we inferred from nuclear data (Figure 3E, F). For example, in both mitochondrial ML trees (Figure 3C, D), we resolved a relatively deep split between a clade of western *Q. mexicanus* and a clade that includes eastern *Q. mexicanus* and *Q. major*, suggesting non-monophyly of *Q. mexicanus*—similar to results obtained by previous mitochondrial studies (DaCosta et al. 2008, Powell et al. 2008; Figure 3A, B). However, our nuclear UCE phylogenies (Figures 2 and 3E, F) strongly conflict with these results. Similarly, the CYTB analysis in which we were able to include *Q. palustris* (Figure 3C) resolved this taxon as sister to the western *Q. mexicanus* clade, as in the mitochondrial analysis of Powell et al. (2008), although our ML analysis resolved *Q. palustris* as sister to all *Q. mexicanus* individuals (Figure 2). The only notable difference between the 2 mitochondrial ML analyses is that CYTB resolved eastern *Q. mexicanus* as monophyletic and sister to *Q. major* (as in Powell et al. 2008's analysis of CYTB and ND2; Figure 3A, C), whereas the ND2 tree could not resolve relationships within the clade containing *Q. major* and individuals of eastern *Q. mexicanus* (Figure 3D).

In all 5 cases where we sequenced individuals included by Wehtje (2004) or DaCosta et al. (2008), the relationships we resolved concurred with their findings. Using our CYTB and ND2 phylogenies and data from Wehtje (2004) and DaCosta et al. (2008), we assigned 16 individuals to a western *Q. mexicanus* mitochondrial clade, 19 individuals to an eastern Q. mexicanus clade, and 50 individuals to a *Q. major* mitochondrial clade (Supplementary Material Table 1). The western clade included *Q. m. graysoni* individuals from Sinaloa and *Q. m. nelsoni* from Sonora, California, Nevada, and parts of Arizona. All other *Q. mexicanus*, including birds from New Mexico, Texas, and one individual from Arizona, were placed in the eastern clade. Interestingly, the lone "eastern" bird from Arizona (LACM 11135, from Lake Havasu), was collected farther west than 3 of the "western" Arizona individuals (LACM 111355, 111362,

111658).

Principal Component Analyses of Population Structure

Principal component analyses of all 115 individuals representing the 3 Cassidix grackle species (Figure 4A) show a clear separation between *Q. major* and *Q. mexicanus* + *Q. palustris* on PC1, with no intermediate individuals, whereas PC2 separates the *Q. palustris* individual from *Q. mexicanus* while also suggesting a high degree of intraspecific variation among *Q. mexicanus* individuals.

The subset PCA including only *Q. major* individuals (Figure 4B) showed that *Q. m. major*, *Q. m. westoni*, and *Q. m. torreyi* all formed distinct, non-overlapping clusters. *Quiscalus m. alabamensis* formed 2 distinct clusters: one that included most individuals collected in Alabama and another, located much closer to *Q. m. major*, that contained all individuals collected in Mississippi. One *Q. m. alabamensis* individual collected in Alabama (LSUMNS B-100475) appeared intermediate between the Alabama and Mississippi clusters.

The subset PCA that included *Q. mexicanus* and *Q. palustris* individuals (Figure 4C) largely recapitulated the clades found in our phylogenetic analyses. *Quiscalus m. prosopidicola* formed its own cluster in PC space, while *Q. m. nelsoni* and *Q. m. monsoni* formed a cline, with *Q. m. graysoni* individuals positioned most closely to *Q. m. nelsoni* individuals. The Central and South American subspecies *Q. m. mexicanus*, *Q. m. loweryi*, *Q. m. obscurus*, and *Q. m. peruvianus* all clustered near each other. Individuals collected near the contact zones between *Q. m. prosopidicola* and *Q. m. monsoni* as well as between *Q. m. prosopidicola* and *Q. m. mexicanus* were positioned as intermediates between those subspecies. Although the *Q. palustris* individual was slightly separated from all *Q. mexicanus* individuals when plotting PC1 and PC2, it was distinct from all *Q. mexicanus* samples when plotting PC3 and PC4 (Figure 4D).

Maximum Likelihood Analysis of Population Structure

PopCluster uses 2 metrics to determine the "optimal" *K* value: D_{LK2} and F_{ST}/F_{IS}. According to the former, the optimal *K* was 2, while the latter suggested the optimal *K* was 1. However, this is not particularly informative, and increasing the value of *K* can reveal hierarchical patterns of population structure, albeit with the risk of overfitting the model (Pritchard et al. 2000, Janes et al. 2017, Lawson et al. 2018). Therefore, Figure 5 shows the admixture proportions of the best run for all values of *K* from 2 to 8 to give a more complete picture of the nested patterns of population structure in the *Cassidix* grackle complex.

At $K = 2$, individuals were divided into clusters representing *Q. mexicanus* and *Q. major*. Some *Q. mexicanus prosopidicola* individuals from Texas and Louisiana had estimates suggesting ~10% ancestry shared with the *Q. major* cluster. However, this result appears to be an artifact of additional population structure within *Q. mexicanus* rather than representing admixture, because some of these putatively admixed individuals were sampled from well outside the contact zone, admixture is not suggested in any *Q. major* individuals, and signals of potential admixture disappear at values of $K > 3$.

The results for $K = 3$ were similar to $K = 2$, but *Q. major* was divided into a western cluster (*Q. m. major* and *Q. m. alabamensis*) and an eastern cluster (*Q. m. westoni* and *Q. m. torreyi*), mirroring the results from our phylogenetic analyses of UCE data (Figure 2). At $K = 4$, we no longer inferred putative admixture between *Q. mexicanus* and *Q. major*, and *Q. mexicanus* was subdivided into groups generally corresponding to an east–west division: individuals from Sonora, California, Nevada, and Arizona were assigned to one cluster, while birds from

Louisiana, Tamaulipas, and most of Texas were assigned to the other. Three individuals from New Mexico showed some degree of admixture, having 70–86% of their ancestry shared with the western cluster and 14–30% shared with the eastern cluster. The single sample from far west Texas (LSUMNS B-77433) also showed admixture: 46.5% western ancestry and 53.5% eastern ancestry. These results closely match the Clades A and B recovered in our nuclear phylogenies (Figure 2). The admixture results where $K = 5$ were nearly identical to $K = 4$, except that they assigned individuals to a new cluster that coincided with *Q. m. alabamensis* collected in Alabama. Although most individuals collected in Alabama showed no admixture at this value of K (<1%), one sample (LSUMNS B-100475), which also appeared as an intermediate in PCA space (Figure 4B), was estimated to share 28% ancestry with the cluster of *Q. major* individuals collected in Mississippi and Louisiana. Some *Q. m. alabamensis* individuals collected in Mississippi shared a portion of their ancestry (<14%) with this new cluster, but this result also appeared to be an artifact because it disappeared at higher values of K. $K = 6$ suggested further substructure within western *Q. mexicanus*, with individuals in New Mexico, west Texas, and Nevada (broadly mapping to the range of *Q. m. monsoni*) separated from California and Sonora birds (from the range of *Q. m. nelsoni*), and significant admixture inferred among Arizona, Nevada, and some California individuals. At $K = 7$, Mississippi *Q. m. alabamensis* individuals were assigned to their own cluster, distinct from birds collected in Louisiana and Alabama, and at $K = 8$, Florida *Q. m. westoni* and Atlantic coast *Q. m. torreyi* were divided into their own clusters for the first time. More generally, at $K = 8$ each of the 7 named subspecies that we sampled were assigned to their own cluster (as well as the Mississippi subpopulation of *Q. major*), and there was evidence for admixture between populations of *Q. m. nelsoni, Q. m. monsoni,* and *Q. m. prosopidicola.*

We compared admixture results to the mitochondrial clades to which we assigned individuals (Supplementary Material Table S1, Figure 5). All *Q. major* individuals for which we obtained mtDNA had *Q. major* mitochondrial haplotypes. Within *Q. mexicanus,* samples corresponding to the *Q. m. prosopidicola* cluster (from Louisiana, Central and East Texas, and Tamaulipas) belonged to the eastern *Q. mexicanus* mtDNA lineage. Generally, samples belonging to the western *Q. m. nelsoni* cluster had western mtDNA haplotypes, including all California individuals. However, in Nevada and Arizona, eastern and western mtDNA haplotypes of *Q. mexicanus* co-occurred, and there was not perfect correspondence between the cluster to which an individual was assigned based on nuclear UCE data (at $K = 6$–8) and its mitochondrial haplotype.

DISCUSSION

Absence of Gene Flow Between Species

Our phylogenetic results (Figure 2) combined with our PCA (Figure 4A) and analyses of population structure (Figure 5) provide no evidence of admixture between *Q. mexicanus* and *Q. major,* contrary to previous reports of morphological hybrids collected in Texas and Louisiana (Pratt 1991). This result aligns with research showing complete reproductive barriers at the contact zone between *Q. mexicanus* and *Q. major* (Selander and Giller 1961) but differs from hybridization documented in other parapatric or recently sympatric sister species in the region (Maley 2012, Oswald et al. 2019). The potentially hybrid individuals collected by Pratt (1991) were among the historical samples from which we failed to collect adequate sequence data for analysis, so we could not test whether these individuals represent rare, early-generation (F1)

hybrids collected at the forefront of the expanding *Q. mexicanus* range. However, the absence of detectable admixture in individuals of either species collected from the contact zone in Texas and Louisiana suggests that if F1 hybrids occur, they are infertile or rare.

Given that *Q. major* and *Q. mexicanus* are broadly sympatric and sometimes occur in mixed-species colonies (Selander and Giller 1961), the lack of introgression between them is surprising and indicates the presence of strong reproductive barriers. Behavioral studies show that males of both species will attempt to mate with females of both species, as well as dummies of both species and other distantly related blackbird species (Selander and Giller 1961), so male mate choice seems unlikely to pose a significant barrier. Therefore, the two hypotheses that most likely explain the lack of admixture are strong female mate preferences or intrinsic post-mating barriers. There are fixed behavioral differences between the species in elements of their "ruff-out" territorial display: *Q. major* raises and flutters its wings during the display, whereas *Q. mexicanus* does not (Selander and Giller 1961, Pratt 1974). Song also differs between the species, though icterids learn their songs and some individuals have been observed giving characteristic vocalizations of both species (Pratt 1991, Corbett 2021). Perhaps hybrids are very rare due to strong female preferences for conspecific mating displays or vocalizations (Selander and Giller 1961, Pratt 1991). Post-mating and post-zygotic incompatibilities—such as hybrid inviability or sterility—would be a more complete reproductive barrier between the species, but the divergence between *Q. mexicanus* and *Q. major* is much younger (Powell et al. 2008) than most bird species pairs that show post-zygotic incompatibilities (Price and Bouvier 2002). Determining the causes of reproductive barriers between *Q. mexicanus* and *Q. major* will require careful behavioral observations to evaluate the frequency of inter-specific courtship and mating, combined with continued efforts to identify and collect sequence data from individuals with possible hybrid phenotypes.

Monophyly of *Quiscalus mexicanus*

Previous mitochondrial studies of *Q. mexicanus* questioned its monophyly (DaCosta et al. 2008, Powell et al. 2008), suggesting that the eastern clade of *Q. mexicanus* is more closely related to *Q. major* than to the western clade of its own species (Figure 3A, B). Instead, our results from thousands of nuclear loci (Figures 2 and 3E, F; Supplementary Material Figures S1–S5) strongly support the hypothesis that *Q. major* is not embedded within *Q. mexicanus* and that *Q. mexicanus* should remain a single species. However, when we analyzed mitochondrial CYTB and ND2 sequences assembled from off-target reads, we also found that *Q. mexicanus* appeared non-monophyletic (Figure 3C, D), suggesting significant discordance between the phylogenetic signals encoded in the mitochondrial and nuclear genomes of these birds. This discordance also appears when comparing the results from our population structure analyses to the mitochondrial clades to which we assigned individuals (Figure 5). Although mito-nuclear discordance is not rare in birds and other animals (Funk and Omland 2003, McKay and Zink 2010, Toews and Brelsford 2012), this case represents a striking example of conflicting evidence from mtDNA and nuclear DNA regarding species monophyly. This is likely the result of complex patterns of secondary contact—with or without reproductive isolation—in the *Cassidix* clade.

One hypothesis explaining the discordance we observed is that two relatively divergent mitochondrial lineages existed within the ancestral population of the *Cassidix* complex, with one haplogroup becoming fixed in *Q. palustris* and *Q. mexicanus* individuals found along the Pacific coast of Mexico and California, while the other became fixed in eastern *Q. mexicanus* individuals and *Q. major*. DaCosta et al. (2008) and Wehtje (2003) suggested that all 4 lineages

(*Q. palustris*, *Q. mexicanus* "west", *Q. mexicanus* "east", and *Q. major*) were once allopatric but came into secondary contact during the past century because of the *Q. mexicanus* range expansion (except for *Q. palustris*, which went extinct). In the resulting southwestern U.S. contact zone, the eastern and western lineages of *Q. mexicanus* began interbreeding, generating the variation in admixture proportions that we estimated among individuals in California, Arizona, and Nevada (Figure 5). Because mitochondria do not generally recombine, divergent mitochondrial lineages were maintained, producing the mitonuclear discordance we observed (Figure 3). Both western and eastern mitochondrial haplotypes were present among individuals collected in Nevada and Arizona, providing additional evidence that the two mitochondrial clades now broadly overlap geographically in the southwest, as previously reported (Wehtje 2004, DaCosta et al. 2008).

Placement of *Quiscalus palustris*

Even with our genome-wide UCE markers, the relationships of *Q. palustris* could not be resolved with certainty, and conclusions derived from a single toepad sample of one individual should be viewed with caution. The ML analysis showed moderate support (82% bootstrap) for the placement of *Q. palustris* as sister to all *Q. mexicanus* (Figure 2), but the coalescent analysis did not resolve this relationship with confidence. In the PCAs, *Q. palustris* was clearly separated from all *Q. mexicanus* individuals (Figure 4). Considered together, these results suggest that *Q. palustris* and *Q. mexicanus* were sister species (rather than *Q. palustris* being embedded within *Q. mexicanus*), but limited sampling of *Q. palustris*, along with the presence of relatively deep intraspecific structuring within *Q. mexicanus*, suggests additional work is needed to validate this conclusion. Furthermore, care is always warranted when drawing conclusions from taxa represented solely by historical toepad samples (in this study, *Q. palustris*, *Q. m. loweryi*, and *Q. m. obscurus*). However, when we were able to include toepad samples alongside more reliable, modern tissues samples, as with the *Q. m. alabamensis* individuals collected from Mississippi and Alabama, the toepad samples clustered closely with tissue samples from their respective states in all analyses, providing evidence that the inclusion of toepads did not significantly skew our results.

Relationships within *Quiscalus mexicanus*

Our phylogenetic analyses divided *Q. mexicanus* into broad groups that correspond to geography (Figures 3 and 6), most of which are also supported by our PCAs (Figure 4) and our population structure analyses (Figure 5). In general, our results show that the current division of *Q. mexicanus* into 8 subspecies does an imperfect job of describing their genetic variation; while some named subspecies correspond well to genetic groups (e.g., *Q. m. prosopidicola*, *Q. m. obscurus*), others are embedded within or are questionably distinct from neighboring subspecies (e.g., *Q. m. graysoni* relative to *Q. m. nelsoni*; *Q. m. loweryi* relative to *Q. m. mexicanus*), and one appears to encompass distantly related individuals (*Q. m. peruvianus* from Peru compared to those from Central America). There is also evidence of admixture between individuals collected near the boundaries of subspecies ranges (e.g., *nelsoni/monsoni*, *monsoni/prosopidicola*, and *prosopidicola/mexicanus*).

The *Q. mexicanus* subspecies in northwestern Mexico and the southwestern U.S. (*Q. m. graysoni*, *Q. m. nelsoni*, and *Q. m. monsoni*) form a monophyletic group (Clade A, Figure 2, Figure 6). This placement of *Q. m. monsoni* with the western subspecies *Q. m. nelsoni* and *Q. m. graysoni* is surprising because *Q. m. monsoni* is part of the eastern mitochondrial lineage

(DaCosta et al. 2008). At *K* = 6–8, *Q. m. nelsoni* and *Q. m. monsoni* form separate clusters in the population structure analysis (Figure 5), but we find extensive admixture between the two taxa. At these values of *K*, and in our PCA (Figure 4C), we also detected admixture between *Q. m. monsoni* and *Q. m. prosopidicola* in one individual collected in western Texas. The remaining subspecies are resolved as several different clades. Our phylogenetic analyses unite *Q. m. prosopidicola* from Texas and Louisiana as a monophyletic group (Clade B in Figures 2 and 6) that we also identified in the PCA (Figure 4C) and population structure analysis (Figure 5) at most values of *K* (*K* = 3–8). There is also a clear affinity between *Q. m. mexicanus*, *Q. m. loweryi*, and Central American populations of *Q. m. peruvianus* (Clade C in Figures 2 and 6). However, our results suggest that the limits of *Q. m. peruvianus* may need to be re-evaluated: we resolved the sample collected near the type locality in Peru as sister to all remaining *Q. mexicanus* individuals, whereas *Q. m. peruvianus* individuals collected in Panama and Costa Rica were united with birds from Guatemala and southern Mexico (Clade C). We were unable to include *Q. m. peruvianus* individuals from Colombia or Ecuador to evaluate where the genetic break between South and Central America *Q. m. peruvianus* might occur. The two samples of *Q. m. obscurus* from southwest Mexico form their own small Clade D, which our coalescent analysis suggests is sister to nearby Clade C (Supplementary Material Figure 3).

Relationships within *Quiscalus major*

Our phylogenetic, principal component, and population clustering results show that two main clades exist within *Q. major*—one corresponding to a western group (*Q. m. major* and *Q. m. alabamensis*) and one corresponding to an eastern group (*Q. m. westoni* and *Q. m. torreyi*). The break between these groups is associated with the largest disjunction within the range of *Q. major*: a gap of >200 km in the Florida panhandle that spans the distance between the easternmost breeding populations of *Q. m. alabamensis* at the mouth of the Escambia River (Stevenson and Anderson 1994, Duncan and Duncan 2018) and the westernmost breeding populations of *Q. m. westoni* around Apalachicola Bay (Stevenson 1978). This region has relatively little saltmarsh (Davis et al. 2021), although there may be suitable habitat in the area that does not harbor *Q. major* (Stevenson 1978). A nearly identical range gap in the Florida panhandle corresponds to genetic structuring between subspecies of *Ammospiza maritima* (Seaside Sparrow), another marsh-associated songbird (Woltmann et al. 2014, Davis et al. 2021). More broadly, the Apalachicola River region has been identified as an important phylogeographic break for a wide variety of organisms (Avise et al. 1979, Bermingham and Avise 1986, Soltis et al. 2006, Lyman and Edwards 2022).

This division of *Q. major* into eastern and western clades strongly conflicts with the hypothesis that light and dark eye colors correspond to major genetic differences within the species. Instead, both clades contain one dark-eyed subspecies (*Q. m. major* in the west; *Q. m. westoni* in the east) and one light-eyed subspecies (*Q. m. alabamensis* in the west; *Q. m. torreyi* in the east), suggesting multiple evolutionary transitions in eye color within the species. Additionally, there are not always clear relationships between these light-eyed and dark-eyed groups. For example, within the western clade, the relationships among populations are unexpected, particularly given their taxonomy, and do not always correspond to differences in eye color. The description of *Q. m. alabamensis* suggests that the subspecies is found from Horn Island and Gaultier in Mississippi, east throughout coastal Alabama, and locally into northwestern Florida (Stevenson 1978). However, our phylogenetic analyses resolved one clade containing samples from Mobile Bay, Alabama, while *Q. m. alabamensis* individuals collected at

the mouth of the Pascagoula River in Mississippi were resolved as members of a different clade along with all *Q. m. major* individuals collected in Louisiana and Texas. Similarly, population clustering analyses suggested that *Q. m. alabamensis* collected in Mobile Bay comprised their own cluster at lower levels of *K* (*K* = 5–6) while *Q. m. alabamensis* collected in Mississippi were grouped with Louisiana and Texas *Q. m. major* individuals. At higher levels of *K* (*K* = 7–8), *Q. m. alabamensis* individuals collected in Mississippi were assigned to their own group, separate from the *Q. m. alabamensis* individuals collected in Alabama. Our PCA (Figure 4B) recapitulated this finding, dividing *Q. m. alabamensis* into non-overlapping groups representing Alabama individuals and Mississippi individuals, with the Mississippi cluster located closer to *Q. m. major* in PC space. The exception to this trend of differences between populations of *Q. m. alabamensis* in Mississippi and Alabama was a single individual (LSUMNS B-100475) collected from Mobile Bay, Alabama that showed an intermediate position in the PCA (Figure 4B) and admixed ancestry in our ML analyses of population structure (Figure 5, *K* = 5–8).

Relationships within the eastern group of Boat-tailed Grackles (*Q. m. torreyi* and *Q. m. westoni*) were somewhat clearer, although we still did not observe perfect correspondence between our phylogenetic results and eye color. Our phylogenetic analyses resolved a clade of light-eyed *Q. m. torreyi* individuals (collected in South Carolina, North Carolina, and New York) but did not resolve a clade corresponding to dark-eyed *Q. m. westoni*. On the other hand, our PCA clearly separated *Q. m. westoni* and *Q. m. torreyi* individuals, and the population clustering results were similar, with both subspecies assigned to a single cluster at values of *K* < 8, until they were neatly separated into two groups matching the geographic limits of the subspecies at *K* = 8. We did not obtain samples from their contact zone near Jacksonville, Florida, which may provide future insights into the genetic and eye color differences between these two subspecies.

The Curious Case of *Quiscalus m. alabamensis*

Besides questioning the link between eye color and evolutionary relationships, our analyses identified a surprising set of phylogenetic and population clustering results for *Q. m. alabamensis*. The grouping of samples of *Q. m. alabamensis* from Mississippi with *Q. m. major*—both in the phylogeny and the population clustering at low values of *K*—calls into question the distinctness of the Mississippi populations. Conversely, the PCA and population clustering results at higher values of *K* (*K* = 7–8) suggest genetic differences that set Mississippi and Alabama *Q. m. alabamensis* apart, although structure detected at high values of K can sometimes result from overfitting admixture models (Novembre 2016, Lawson et al. 2018). Depending on the taxonomic criteria used, these results alternatively suggest that *Q. m. alabamensis* should either be synonymized with *Q. m. major*, restricted to refer only to populations around Mobile Bay, or that Mississippi populations also deserve subspecies-level recognition. Stevenson's (1978) original description of *Q. m. alabamensis* focused primarily on eye color differences between it and *Q. m. major*, but no single eye color corresponds to a particular subspecies. For example, we collected light-eyed *Q. major* in Louisiana, Mississippi, and Alabama, and dark-eyed individuals in Louisiana and Mississippi (Supplemental Table 1). The only significant morphometric difference between *Q. m. major* and *Q. m. alabamensis* is in male tail length, which also varies clinally within *Q. m. major* – with shorter tails in the east, near the boundary with *Q. m. alabamensis* (Stevenson 1978). Considered together, these results call into question the phenotypic distinctiveness of *Q. m. alabamensis*, while suggesting that there may be previously unrecognized genetic structure associated with geography even within the small range of this subspecies.

Conclusions

Our range-wide study of phylogeography and population structure of the *Cassidix* grackle species complex using thousands of nuclear markers corroborates some previous findings made using mtDNA and phenotypic data but challenges others. These discrepancies illustrate the importance of conducting range-wide sampling and sequencing genome-wide markers to achieve a more complete understanding of the evolutionary history of a widespread species complex. Despite their recent sympatry, we found no evidence for introgressive hybridization between *Q. mexicanus* and *Q. major*, supporting previous reports that the taxa are reproductively isolated. Whether this isolation is due to strong female mate choice or intrinsic barriers warrants further study, and additional sampling would be needed to exclude or confirm the possibility of occasional hybrids, which have been previously reported. Although two non-sister mitochondrial lineages are present within *Q. mexicanus*, the species is monophyletic when considering nuclear markers, and this result further demonstrates the utility of genome-wide sampling because data from a single locus (as in mtDNA) can give misleading results. Our work also provides support for the placement of the extinct *Q. palustris* as sister to *Q. mexicanus* and presents new data on the intraspecific relationships of the 8 described *Q. mexicanus* subspecies, 3 of which are undergoing secondary contact with gene flow in the southwest United States. Finally, within *Q. major*, the most significant genetic break we detected corresponds to a range gap in the Florida panhandle, which has been previously recognized as a barrier for many other species. Both larger clades within *Q. major* include light-eyed and dark-eyed individuals, making the species a promising system for studying the genetic basis of eye color variation in birds.

Supplementary material

Supplementary material is available at *Ornithology* online.

Acknowledgments

We thank the American Museum of Natural History (Paul Sweet, Brian Smith, Joel Cracraft), the Natural History Museum of Los Angeles County (Kimball Garrett, Allison Shultz), the Louisiana State University Museum of Natural Science (Donna Dittmann, Steve Cardiff, Robb Brumfield, Fred Sheldon, Greg Thom), the Museum of Comparative Zoology (Jeremiah Trimble, Scott Edwards, Emily Blank, Cynthia Wang-Claypool), the Museum of Southwestern Biology (Andrew Johnson, Christopher Witt), the Museum of Vertebrate Zoology (Carla Cicero, Rauri Bowie), the Florida Museum of Natural History (Andy Kratter), the National Museum of Natural History (Christopher Milensky, Gary Graves), and the Burke Museum of Natural History and Culture (Sharon Birks, Alejandro Rico-Guevara) for loans of specimen tissues. We also thank the Louisiana Department of Wildlife and Fisheries (Jeffrey Duguay), the Alabama Department of Conservation and Natural Resources (Tereza Ancelet), the Mississippi Department of Marine Resources (Daniel Taylor), and the Mississippi Department of Wildlife, Fisheries, and Parks (Scott Peyton) for their assistance with permitting. Subir Shakya produced the wonderful grackle illustrations used in Figure 2, and Andre Moncrieff, Matt Brady, Ryan Klutts, and Van Remsen assisted with various aspects of fieldwork. Thanks to Michael Hellberg who provided comments on an earlier version of this manuscript and to H. Douglas Pratt for taking the time to discuss his work with and knowledge of grackles. Thank you also to Garth Spellman, Christie Riehl and two anonymous reviewers for their helpful comments and suggestions on improving the manuscript.

Portions of this research were conducted with high performance computational resources provided by Louisiana State University (http://www.hpc.lsu.edu).

Funding statement
This research was supported by funds provided to E.C.C. by the LSU Department of Biological Sciences, the LSU Museum of Natural Science, the Louisiana Ornithological Society, the American Museum of Natural History Chapman Fund, the American Ornithological Society, and the Society of Systematic Biologists. Additional support was provided by NSF DEB-165562 to B.C.F. and R.T.B., NSF IOS- 1754417 to B.C.F, and the LSU College of Science to B.C.F and R.T.B.

Ethics statement
All sample collection was done under IACUC protocols 18-054 and 21-042.

Conflict of interest statement
The authors declare that they have no conflict of interest.

Author contributions
E.C.C., B.C.F., and R.T.B conceived the research idea. E.C.C., R.T.B, and B.C.F. designed the study. E.C.C. and R.T.B. collected the data. E.C.C and B.C.F. analyzed the data and wrote the paper. All authors edited the manuscript. R.T.B. and B.C.F. supervised the research.

Data availability
Analyses reported in this article can be reproduced using the data provided by Corbett et al. (2026).

LITERATURE CITED

Allio, R., A. Schomaker-Bastos, J. Romiguier, F. Prosdocimi, B. Nabholz, and F. Delsuc (2020). MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources* 20:892–905.

Andersen, M. J., J. M. McCullough, E. F. Gyllenhaal, X. M. Mapel, T. Haryoko, K. A. Jønsson, and L. Joseph (2021). Complex histories of gene flow and a mitochondrial capture event in a nonsister pair of birds. *Molecular Ecology* 30:2087–2103.

Avise, J. C., C. Giblin-Davidson, J. Laerm, J. C. Patton, and R. A. Lansman (1979). Mitochondrial DNA clones and matriarchal phylogeny within and among geographic populations of the pocket gopher, *Geomys pinetis*. *Proceedings of the National Academy of Sciences USA* 76:6694–6698.

Avise, J. C., and R. M. Zink (1988). Molecular genetic divergence between avian sibling species: King and Clapper Rails, Long-Billed and Short-Billed Dowitchers, Boat-Tailed and Great-Tailed Grackles, and Tufted and Black-Crested Titmice. *The Auk* 105:516–528.

Bates, D., M. Mächler, B. Bolker, and S. Walker (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* 67.

Bermingham, E., and J. C. Avise (1986). Molecular zoogeography of freshwater fishes in the southeastern United States. *Genetics* 113:939–965.

BirdLife International (2026). IUCN Red List for birds. https://datazone.birdlife.org

Bolger, A. M., M. Lohse, and B. Usadel (2014). Trimmomatic: A flexible trimmer for Illumina

sequence data. *Bioinformatics* 30:2114–2120.
Brooks, A. (1928). Are the Boat-Tailed and Great-Tailed Grackles specifically distinct? *The Auk* 45:506–507.
Chang, C. C., C. C. Chow, L. C. Tellier, S. Vattikuti, S. M. Purcell, and J. J. Lee (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience* 4:7.
Chapman, F. M. (1917). The distribution of bird-life in Colombia: A contribution to a biological survey of South America. *Bulletin of the American Museum of Natural History* 36:1–729.
Chen, S., Y. Zhou, Y. Chen, and J. Gu (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 34:i884–i890.
Chifman, J., and L. Kubatko (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30:3317–3324.
Clements, J. F., P. C. Rasmussen, T. S. Schulenberg, M. J. Iliff, T. A. Fredericks, J. A. Gerbracht, D. Lepage, A. Spencer, S. M. Billerman, B. L. Sullivan, M. Smith, and C. L. Wood (2024). The eBird/Clements checklist of Birds of the World: v2024. https://www.birds.cornell.edu/clementschecklist/download/.
Corbett, E. (2021). eBird Checklist: ebird.org/checklist/S95565741. *eBird: An online database of bird distribution and abundance*. https://ebird.org/checklist/S95565741.
Corbett, E. C., R. T. Brumfield, and B. C. Faircloth (2024). The mechanistic, genetic and evolutionary causes of bird eye colour variation. *Ibis* 166:560–589.
Corbett, E. C., A. E. Moncrieff, R. T. Brumfield, and B. C. Faircloth (2025). A reference genome for Boat-tailed Grackles (*Quiscalus major*). *Journal of Heredity* 116:540–547
Corbett, E. C., R. T. Brumfield, and B.C. Faircloth (2026). Data from: Phylogenetic relationships, hybridization, and genetic structuring in “Cassidix” grackles. *Ornithology* 143:duag000. 10.6084/m9.figshare.29217896 [Dataset].
DaCosta, J., W. Wehtje, and J. Klicka (2008). Historic genetic structuring and paraphyly within the Great-tailed Grackle. *The Condor* 110:170–177.
Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. DePristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. McVean, and R. Durbin (2011). The variant call format and VCFtools. *Bioinformatics* 27:2156–2158.
Danecek, P., J. K. Bonfield, J. Liddle, J. Marshall, V. Ohan, M. O. Pollard, A. Whitwham, T. Keane, S. A. McCarthy, R. M. Davies, and H. Li (2021). Twelve years of SAMtools and BCFtools. *GigaScience* 10:giab008.
Davis, K. E. C., A. E. Settlecowski, M. R. Roeder, C. Enloe, T. Virzi, M. E. Hunter, S. Woltmann, and S. S. Taylor (2021). Range-wide population genetic analysis of Seaside Sparrows (*Ammospiza maritima*) supports at least five distinct population segments that do not align with current subspecies descriptions. *Ornithological Applications* 123:duab019.
Del-Rio, G., M. A. Rego, B. M. Whitney, F. Schunck, L. F. Silveira, B. C. Faircloth, and R. T. Brumfield (2022). Displaced clines in an avian hybrid zone (Thamnophilidae: *Rhegmatorhina*) within an Amazonian interfluve. *Evolution* 76:455–475.
DeRaad, D. A., J. M. McCullough, L. H. DeCicco, P. M. Hime, L. Joseph, M. J. Andersen, and R. G. Moyle (2023). Mitonuclear discordance results from incomplete lineage sorting, with no detectable evidence for gene flow, in a rapid radiation of Todiramphus kingfishers. *Molecular Ecology* 32:4844–4862.
Dickerman, R. W. (1965). The juvenal plumage and distribution of *Cassidix palustris* (Swainson). *The Auk* 82:268–270.

Dickerman, R. W., and A. R. Phillips (1966). A new subspecies of the Boat-tailed Grackle from México. *The Wilson Bulletin* 78:129–131.
Edwards, S., and S. Bensch (2009). Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology* 18:2930–2933; discussion 2934-2936.
Faircloth, B. C. (2011). Illumiprocessor - software for Illumina read quality filtering. https://github.com/faircloth-lab/illumiprocessor.
Faircloth, B. C. (2016). PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32:786–788.
Faircloth, B. C. (2024). Tutorial III: Harvesting UCE Loci From Genomes. *phyluce 1.7.3 documentation*. https://phyluce.readthedocs.io/en/latest/tutorials/tutorial-3.html.
Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn (2012). Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717–726.
Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
Firneno Jr, T. J., J. R. O'Neill, D. M. Portik, A. H. Emery, J. H. Townsend, and M. K. Fujita (2020). Finding complexity in complexes: Assessing the causes of mitonuclear discordance in a problematic species complex of Mesoamerican toads. *Molecular Ecology* 29:3543–3559.
Funk, D. J., and K. E. Omland (2003). Species-level paraphyly and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics* 34:397–423.
Glenn, T. C., R. A. Nilsen, T. J. Kieran, J. G. Sanders, N. J. Bayona-Vásquez, J. W. Finger, T. W. Pierson, K. E. Bentley, S. L. Hoffberg, S. Louha, F. J. Garcia-De Leon, et al. (2019). Adapterama I: Universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* 7:e7755.
Haemig, P. (2009). Ecology and ethnobiology of the Slender-billed Grackle. *Journal of Ornithology* 151:391–399.
Harper, F. (1934). The Boat-Tailed Grackle of the Atlantic Coast. *Proceedings of the Academy of Natural Sciences of Philadelphia* 86:1–2.
Hellmayr, C. E. (1937). *Catalogue of birds of the Americas and the adjacent islands in Field Museum of Natural History -- Part 10: Icteridae*. Field Museum of Natural History, Chicago, IL, USA.
Hemstrom, W., J. A. Grummer, G. Luikart, and M. R. Christie (2024). Next-generation data filtering in the genomics era. *Nature Reviews Genetics* 25:750–767.
Hou, H., B. Pedersen, and A. Quinlan (2021). Balancing efficient analysis and storage of quantitative genomics data with the D4 format and d4tools. *Nature Computational Science* 1:441–447.
Huang, H., and L. L. Knowles (2016). Unforeseen consequences of excluding missing data from next-generation sequences: Simulation study of RAD sequences. *Systematic Biology* 65:357–365.
Janes, J. K., J. M. Miller, J. R. Dupuis, R. M. Malenfant, J. C. Gorrell, C. I. Cullingham, and R. L. Andrew (2017). The K = 2 conundrum. *Molecular Ecology* 26:3594–3602.
Jaramillo, A., and P. Burke (1999). *New World Blackbirds: The Icterids*. Princeton University Press, Princeton, NJ, USA.

Johnson, K., and B. D. Peer (2022). Great-tailed Grackle (*Quiscalus mexicanus*), version 2.0. In *Birds of the World* (P. G. Rodewald and B. K. Keeney, Editors). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.grtgra.02

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. Von Haeseler, and L. S. Jermiin (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 14:587–589.

Katoh, K., K. Misawa, K. Kuma, and T. Miyata (2002). MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* 30:3059–3066.

Kirkpatrick, S., C. D. Gelatt, and M. P. Vecchi (1983). Optimization by simulated annealing. *Science* 220:671–680.

Kraytsberg, Y., M. Schwartz, T. A. Brown, K. Ebralidse, W. S. Kunz, D. A. Clayton, J. Vissing, and K. Khrapko (2004). Recombination of human mitochondrial DNA. *Science* 304:981–981.

Lawson, D. J., L. van Dorp, and D. Falush (2018). A tutorial on how not to over-interpret STRUCTURE and ADMIXTURE bar plots. *Nature Communications* 9:3258.

Letunic, I., and P. Bork (2024). Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Research* 52:W78–W82.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. http://arxiv.org/abs/1303.3997.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078–2079.

Linck, E., and C. J. Battey (2019). Minor allele frequency thresholds strongly affect population structure inference with genomic data sets. *Molecular Ecology Resources* 19:639–647.

Lowery, G. H. J. (1938). A new grackle of the *Cassidix mexicanus* group. *Occasional Papers of the Museum of Zoology, Louisiana State University* 1:11.

Lyman, R. A., and C. E. Edwards (2022). Revisiting the comparative phylogeography of unglaciated eastern North America: 15 years of patterns and progress. *Ecology and Evolution* 12:e8827.

Maley, J. (2012). Ecological speciation of King rails (*Rallus elegans*) and Clapper rails (*Rallus longirostris*). Doctoral dissertation. Louisiana State University, Baton Rouge, LA, USA.

McCormack, J. E., W. L. E. Tsai, and B. C. Faircloth (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources* 16:1189–1203.

McIlhenny, E. A. (1934). Color of iris in the Boat-tailed Grackle (*Cassidix Mexicanus Major*). *The Auk* 51:383–384.

McKay, B. D., and R. M. Zink (2010). The causes of mitochondrial DNA gene tree paraphyly in birds. *Molecular Phylogenetics and Evolution* 54:647–650.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* 20:1297–1303.

Meiklejohn, K. A., M. J. Danielson, B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball (2014). Incongruence among different mitochondrial regions: A case study using complete mitogenomes. *Molecular Phylogenetics and Evolution* 78:314–323.

Minh, B. Q., H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. Von Haeseler,

and R. Lanfear (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* 37:1530–1534.
Mirchandani, C. D., A. J. Shultz, G. W. C. Thomas, S. J. Smith, M. Baylis, B. Arnold, R. Corbett-Detig, E. Enbody, and T. B. Sackton (2024). A fast, reproducible, high-throughput variant calling workflow for population genomics. *Molecular Biology and Evolution* 41:msad270.
Morrison, M. L., N. Alcala, and N. A. Rosenberg (2022). FSTruct: An F_{ST}-based tool for measuring ancestry variation in inference of population structure. *Molecular Ecology Resources* 22:2614–2626.
Muñoz González, Z. A. G. and D. A. Carrillo Martínez (2025). Slender-billed Grackle (*Quiscalus palustris*), version 2.0. In *Birds of the World* (N. D. Sly, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.slbgra1.02
Nelson, E. (1900). Descriptions of thirty new North American birds, in the Biological Survey Collection. *The Auk* 17:253-270.
Novembre, J. (2016). Pritchard, Stephens, and Donnelly on population structure. *Genetics* 204:391–393.
Oliveros, C. H., D. J. Field, D. T. Ksepka, F. K. Barker, A. Aleixo, M. J. Andersen, P. Alström, B. W. Benz, E. L. Braun, M. J. Braun, G. A. Bravo, et al. (2019). Earth history and the passerine superradiation. *Proceedings of the National Academy of Sciences USA* 116:7916–7925.
Ortiz, E. M. (2021). vcf2phylip v2.8: convert a VCF matrix into several matrix formats for phylogenetic analysis. DOI:10.5281/zenodo.2540861.
Oswald, J. A., M. G. Harvey, R. C. Remsen, D. U. Foxworth, D. L. Dittmann, S. W. Cardiff, and R. T. Brumfield (2019). Evolutionary dynamics of hybridization and introgression following the recent colonization of Glossy Ibis (Aves: *Plegadis falcinellus*) into the New World. *Molecular Ecology* 28:1675–1691.
Pedersen, B. S., and A. R. Quinlan (2018). Mosdepth: Quick coverage calculation for genomes and exomes. *Bioinformatics* 34:867–868.
Peterson, A. T. (1998). The distribution and type locality of the extinct Slender-billed Grackle *Quiscalus palustris*. *Bulletin of the British Ornithologists' Club* 118:119–121.
Phillips, A. R. (1950). The Great-tailed Grackles of the Southwest. *The Condor* 52:78–81.
Pockrandt, C., M. Alzamel, C. S. Iliopoulos, and K. Reinert (2020). GenMap: Ultra-fast computation of genome mappability. *Bioinformatics* 36:3687–3692.
Posit Team (2024). *RStudio: Integrated Development Environment for R.* Posit Team, Boston, MA, USA.
Post, W., J. P. Poston, and G. T. Bancroft (2020). Boat-tailed Grackle (*Quiscalus major*), version 1.0. In *Birds of the World* (A. F. Poole, Editor). Cornell Lab of Ornithology, Ithaca, NY, USA. https://doi.org/10.2173/bow.botgra.01
Powell, A. F. L. A., F. K. Barker, and S. M. Lanyon (2008). A complete species-level phylogeny of the grackles (*Quiscalus* spp.), including the extinct Slender-billed Grackle, inferred from mitochondrial DNA. *The Condor* 110:718–728.
Pratt, H. D. (1974). Field identification of Boat-tailed and Great-tailed Grackles in their zone of overlap. *Birding* 6:217–223.
Pratt, H. D. (1991). Hybridization of Great-tailed and Boat-tailed Grackles (*Quiscalus*) in Louisiana. *Journal of Louisiana Ornithology* 2:2–14.
Price, T. D., and M. M. Bouvier (2002). The evolution of F1 postzygotic incompatibilities in

birds. *Evolution* 56:2083–2089.
Pritchard, J. K., M. Stephens, and P. Donnelly (2000). Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
Prjibelski, A., D. Antipov, D. Meleshko, A. Lapidus, and A. Korobeynikov (2020). Using SPAdes De Novo Assembler. *Current Protocols in Bioinformatics* 70:e102.
Puechmaille, S. J. (2016). The program structure does not reliably recover the correct population structure when sampling is uneven: Subsampling and new estimators alleviate the problem. *Molecular Ecology Resources* 16:608–627.
Python Software Foundation (2020). Python Language Reference, version 3.9. www.python.org.
Quinlan, A. R., and I. M. Hall (2010). BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* 26:841–842.
R Core Team (2024). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Raposo Do Amaral, F., L. G. Neves, M. F. R. Resende, F. Mobili, C. Y. Miyaki, K. C. M. Pellegrino, and C. Biondo (2015). Ultraconserved elements sequencing as a low-cost source of complete mitochondrial genomes and microsatellite markers in non-model amniotes. *PLoS One* 10:e0138446.
Ratnasingham, S., C. Wei, D. Chan, J. Agda, J. Agda, L. Ballesteros-Mejia, H. A. Boutou, Z. M. El Bastami, E. Ma, R. Manjunath, D. Rea, et al. (2024). BOLD v4: A centralized bioinformatics platform for DNA-based biodiversity data. *Methods in Molecular Biology (Clifton, N.J.)* 2744:403–441.
Ridgway, R. (1901). New birds of the Families Tanagridae and Icteridae. *Proceedings of the Washington Academy of Sciences* 3:149–155.
Rohland, N., and D. Reich (2012). Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. *Genome Research* 22:939–946.
Salter, J. F., P. A. Hosner, W. L. E. Tsai, J. E. McCormack, E. L. Braun, R. T. Kimball, R. T. Brumfield, and B. C. Faircloth (2022). Historical specimens and the limits of subspecies phylogenomics in the New World quails (Odontophoridae). *Molecular Phylogenetics and Evolution* 175:107559.
Sclater, P. L. (1884). A review of species of the Family Icteridae — Part IV. Quiscalinae. *Ibis* 2:149–166.
Selander, R. K., and D. R. Giller (1961). Analysis of sympatry of Great-tailed and Boat-tailed Grackles. *The Condor* 63:29–86.
Soltis, D. E., A. B. Morris, J. S. McLachlan, P. S. Manos, and P. S. Soltis (2006). Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* 15:4261–4293.
Sprunt, A. (1934). A New Grackle from Florida. *Charleston Museum Leaflet* 6:1–5.
Stevenson, H. M. (1978). The populations of Boat-tailed Grackles in the southeastern United States. *Proceedings of the Biological Society of Washington* 91:27–51.
Swofford, D. L. (2003). *PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods)*, version 4. Sinauer Associates. Sunderland, MA, USA.
Tarasov, A., A. J. Vilella, E. Cuppen, I. J. Nijman, and P. Prins (2015). Sambamba: Fast processing of NGS alignment formats. *Bioinformatics* 31:2032–2034.
Todd, W. E. C., and M. A. Carriker (1922). Birds of the Santa Marta region of Colombia: A study in altitudinal distribution. *Annals of the Carnegie Museum* 14.
Toews, D. P. L., and A. Brelsford (2012). The biogeography of mitochondrial and nuclear

discordance in animals. *Molecular Ecology* 21:3907–3930.
Tsai, W. L. E., M. E. Schedl, J. M. Maley, and J. E. McCormack (2020). More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. *Molecular Ecology Resources* 20:1220–1227.
Van Der Auwera, G. A., M. O. Carneiro, C. Hartl, R. Poplin, G. Del Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, et al. (2013). From FastQ data to high-confidence variant calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics* 43.
Wang, J. (2017). The computer program structure for assigning individuals to populations: Easy to use but easier to misuse. *Molecular Ecology Resources* 17:981–990.
Wang, J. (2022). Fast and accurate population admixture inference from genotype data from a few microsatellites to millions of SNPs. *Heredity* 129:79–92.
Wang, J. (2024). PopCluster: A population genetics model-based toolset for simulating, inferring and visualising individual admixture and population structure. *Molecular Ecology Resources* 25:e14058.
Waterhouse, A. M., J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton (2009). Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25:1189–1191.
Wehtje, W. (2003). The range expansion of the Great-tailed Grackle (*Quiscalus mexicanus* Gmelin) in North America since 1880. *Journal of Biogeography* 30:1593–1607.
Wehtje, W. (2004). The Great-tailed Grackle (*Quiscalus mexicanus Gmelin*) in the western USA: Range expansion and secondary contact between subspecies. Ph.D. dissertation, University of California, Riverside, CA, USA.
Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*, second edition. Springer International Publishing, Cham, Switzerland.
Woltmann, S., P. C. Stouffer, C. M. Bergeon Burns, M. S. Woodrey, M. F. Cashner, and S. S. Taylor (2014). Population genetics of Seaside Sparrow (*Ammodramus maritimus*) subspecies along the Gulf of Mexico. *PLoS One* 9:e112739.

Figure 1. Map of localities for samples included in this study. Purple and teal shading indicate the ranges (Birdlife International 2026) of *Q. mexicanus* and *Q. major*, respectively. Diamonds represent *Q. major* individuals, circles represent *Q. mexicanus* individuals, and the cross represents one *Q. palustris* individual. Divided circles indicate samples that could not be assigned to subspecies with confidence *a priori* based on collecting locality or existing published data. See Supplementary Material Table S1 for full sample details.

Figure 2. UCE phylogeny of *Cassidix* grackles with tips grouped into major clades. Bootstrap support values from the ML analysis are shown above each branch, and branches with <50% bootstrap support are collapsed. Support values below each branch are shown for clades that were also resolved with at least 50% bootstrap support in the coalescent analysis. Dashes (–) indicate that a clade was not well-supported in the coalescent tree. Icons for subspecies correspond to those used in Figure 1. Eastern and western clades within *Q. major* are labeled E and W and notable clades within *Q. mexicanus* are labeled A-D. Full trees showing the relationships between all tips, with low-support branches present or collapsed, are provided in Supplementary Material Figures S1 and S2. An alternative version of this figure, with the coalescent bootstrap values dictating which clades are shown or collapsed, is provided in

Supplementary Material Figure S3.

Figure 3. Variation in the topologies of *Cassidix* grackle phylogenies inferred from data of different types (mitochondrial loci versus nuclear UCE loci).

Figure 4. Principal component analyses of variation (**A**) among all 115 individuals included in this study, (**B**) within *Q. major* individuals, and (**C–D**) within *Q. mexicanus* and *Q. palustris* individuals. Circled groups within *Q. major* (**B**) and *Q. mexicanus* (**C**) correspond to the clades labeled in Figure 2.

Figure 5. Results from maximum likelihood (ML) analyses of population structure at values of $K = 1–8$, which represented slightly more structuring than the number of named subspecies included in the dataset ($n = 7$). Labels across the top denote the two species we included in ML analyses of population structure and abbreviations below species labels indicate the states of the U.S. and Mexico (SO = Sonora, TA = Tamaulipas) in which individuals were collected. Bars represent individuals, and colors filling each bar represent the admixture proportions estimated for each individual at each value of *K*. The bottom row shows the mitochondrial haplotype to which we assigned individuals: *Q. mexicanus* west is pink, *Q. mexicanus* east is purple, and *Q. major* is teal. Gray indicates that we were unable to collect mitochondrial data for these individuals.

Figure 6. Geographic distributions of well-supported clades within *Q. mexicanus* (purple shading) and *Q. major* (teal shading). Labeled clades correspond to those shown in Figures 2 and 4.

Figure 1

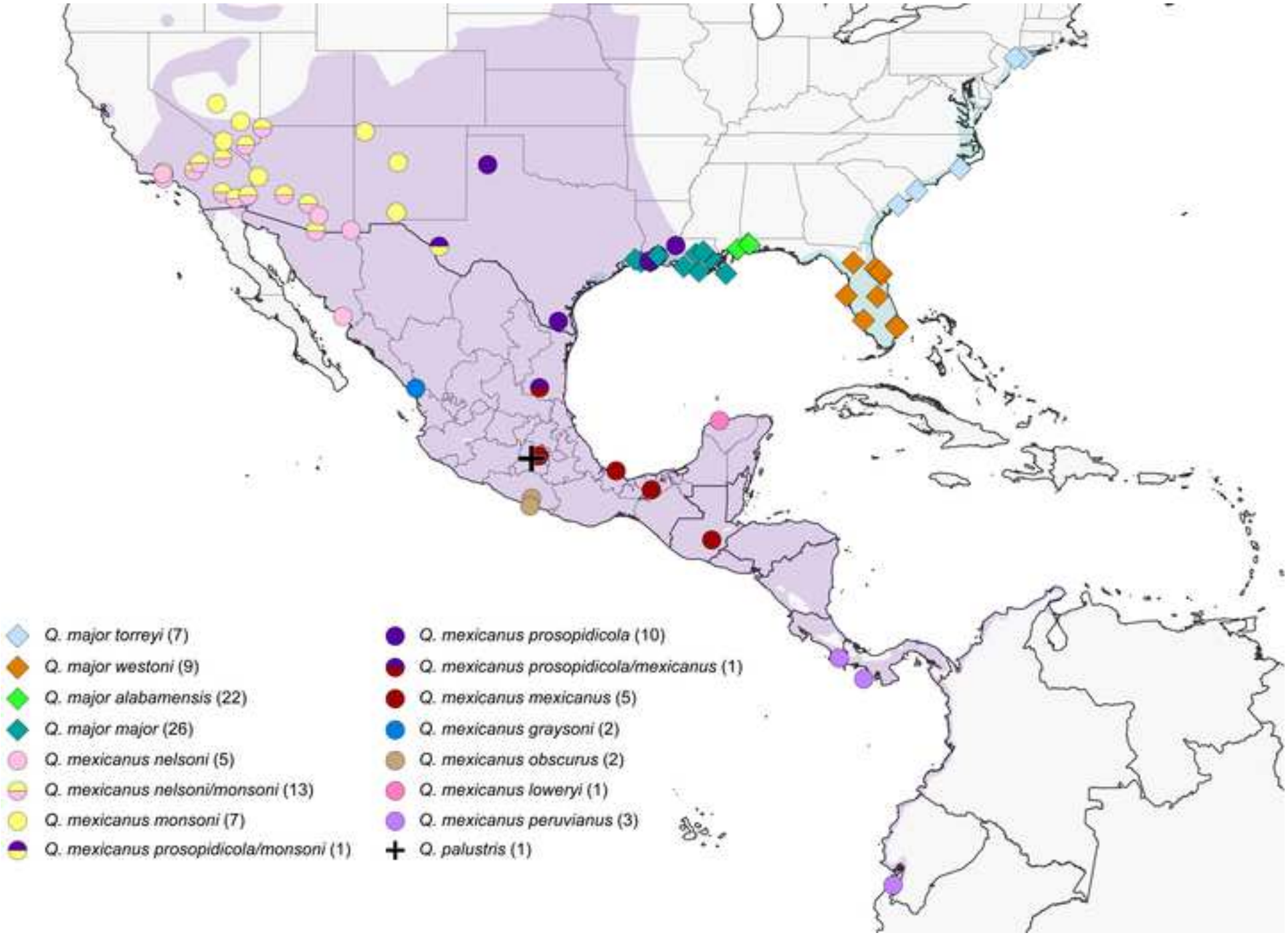

Figure 2

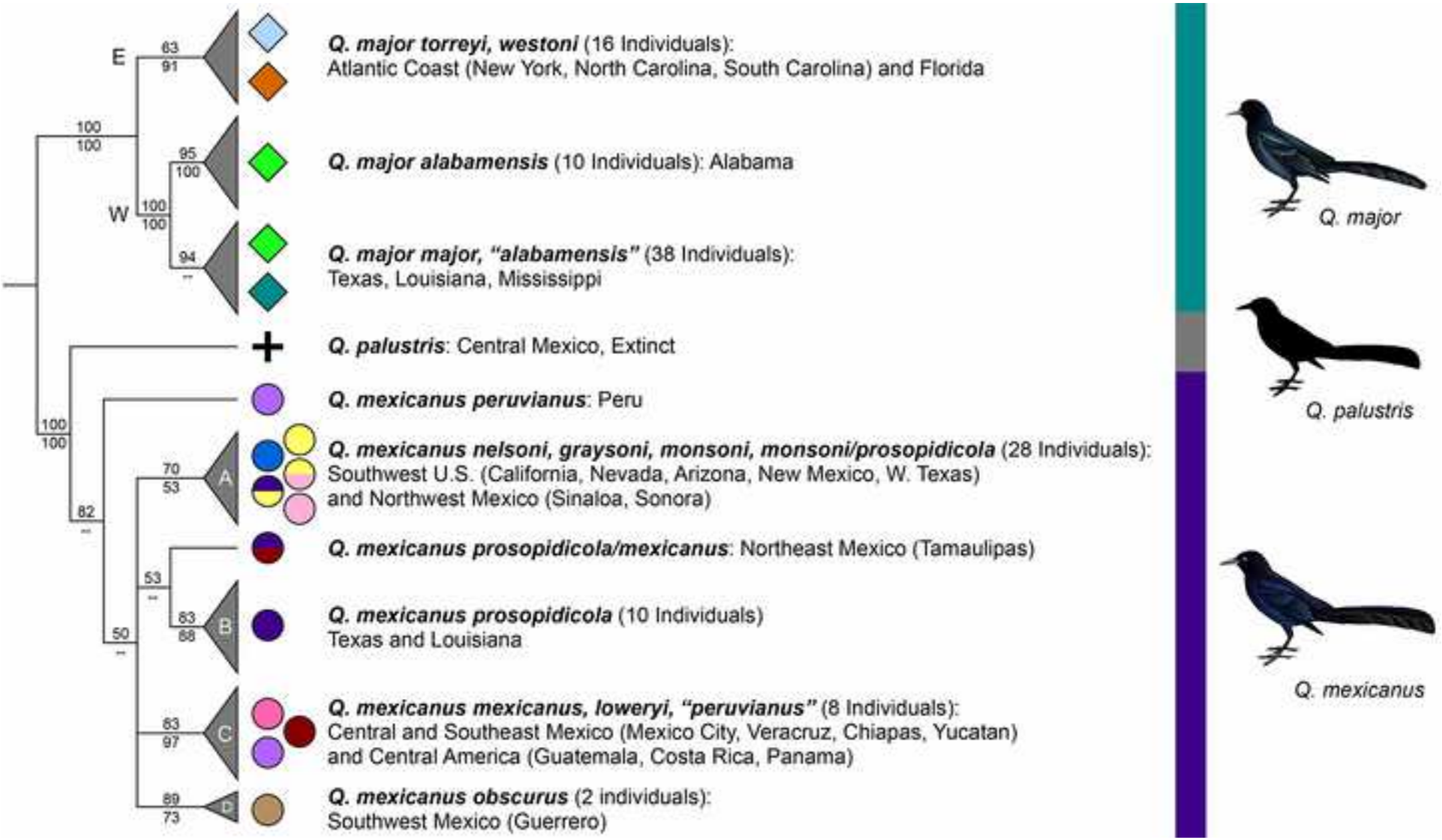

Figure 3

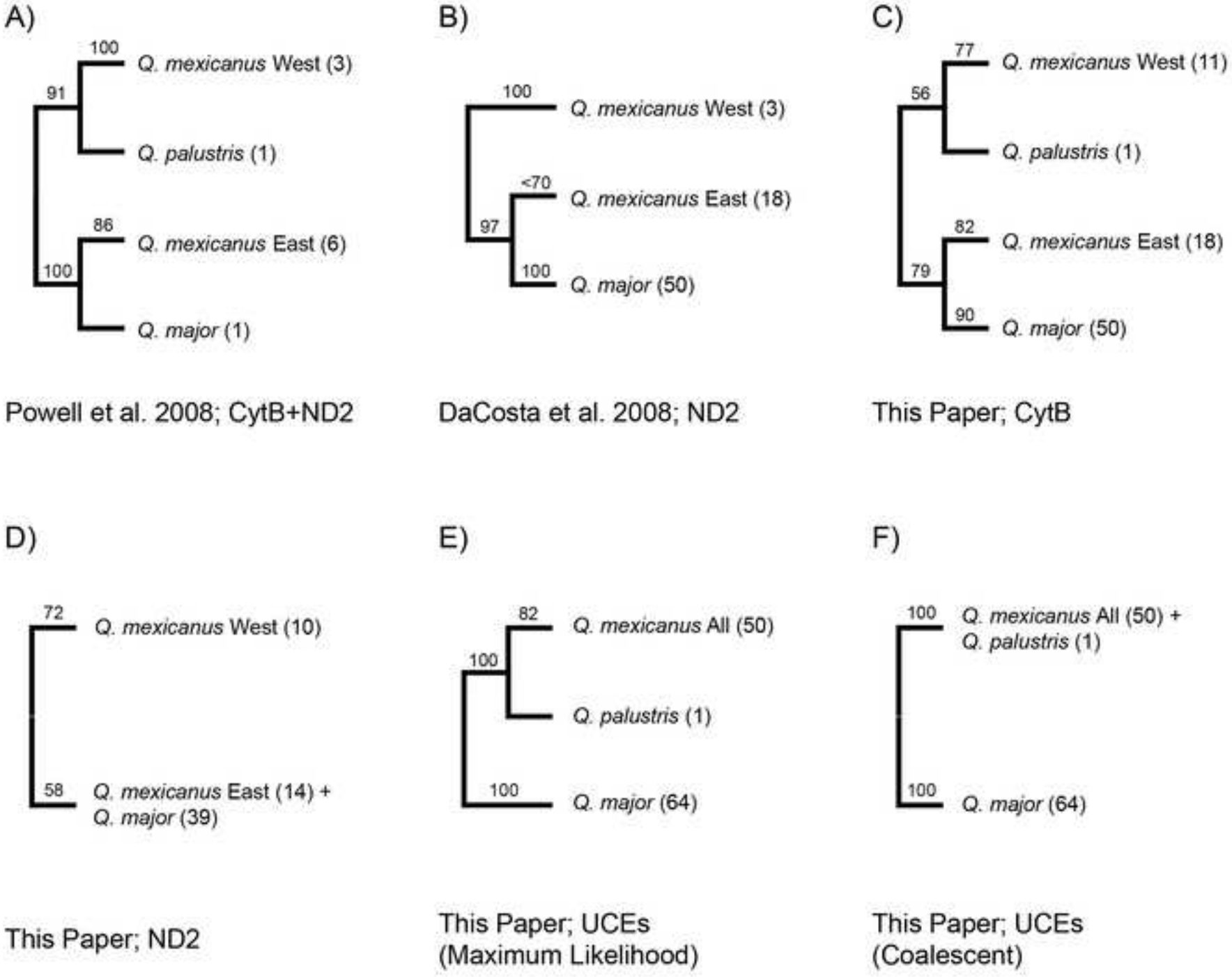

Figure 4

Click here to access/download;Figure;ORNITH-25-091_Figure 4.jpg

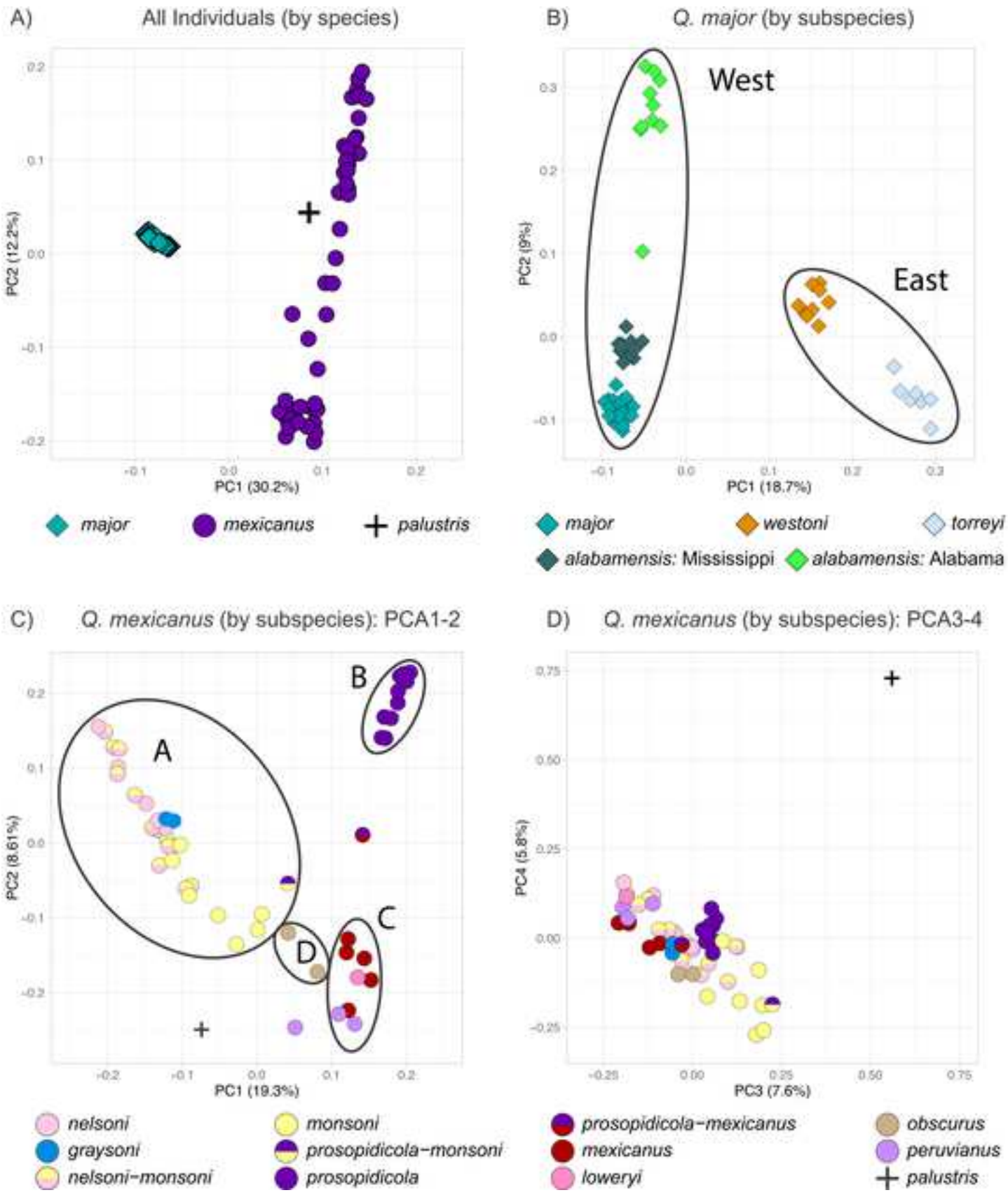

Figure 5

Click here to access/download;Figure;ORNITH-25-091_Figure 5.jpg

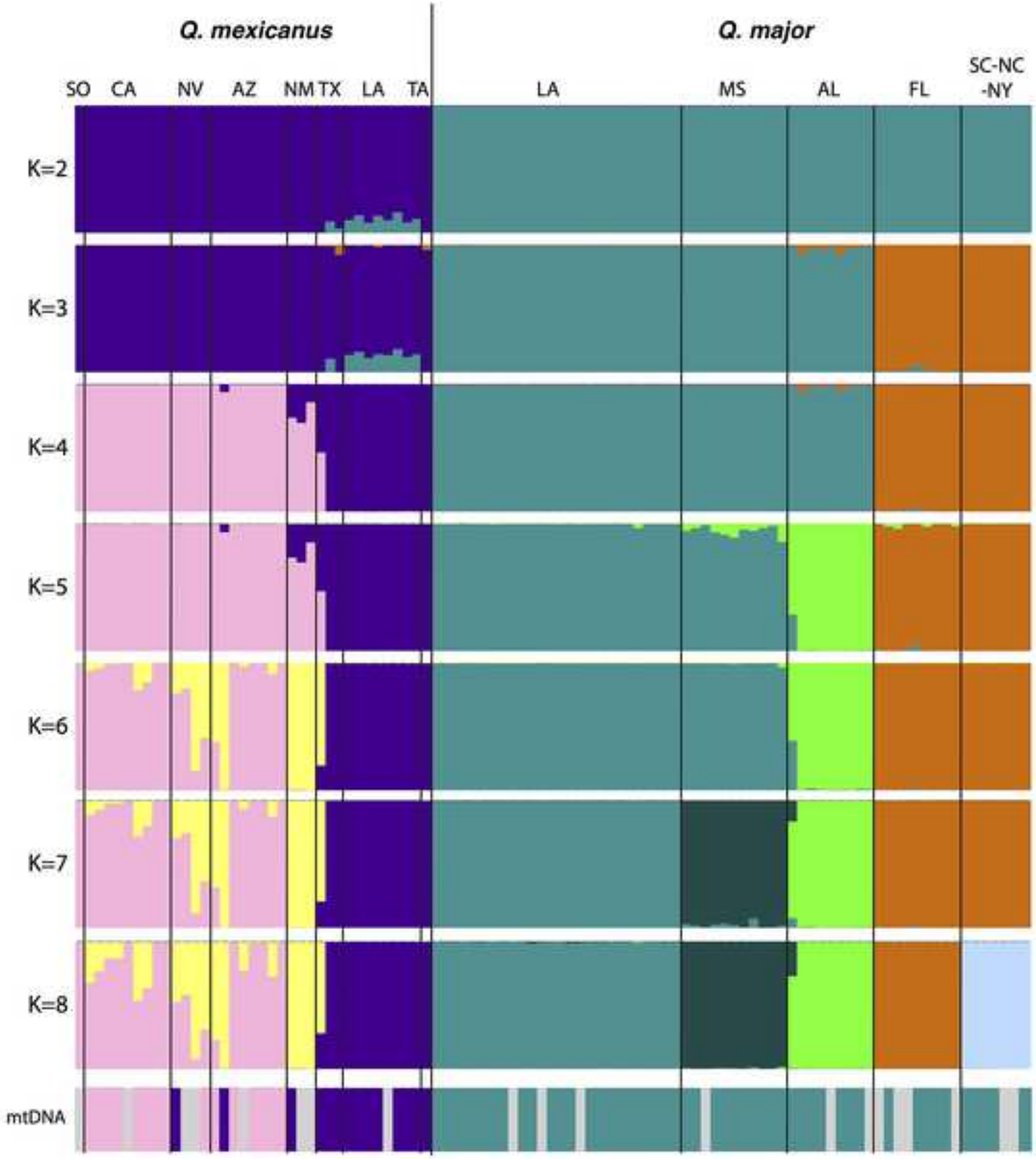

Figure 6