Molecular phylogenetics of the New World blackbirds (*Icteridae*)

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Alexis Frederick Leo Alvey Powell

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

SCOTT M. LANYON, ADVISOR

DECEMBER 2012

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

CHAPTER 2

CHAPTER 3

LIST OF FIGURES

CHAPTER 1

CHAPTER 2

CHAPTER 3

INTRODUCTION

The New World blackbirds (Icteridae) are among the best known songbirds, both through exemplar species, such as the Red-winged Blackbird (*Agelaius phoeniceus*), and collectively, through service as a model clade in numerous studies of morphological, ecological, and behavioral trait evolution. Knowledge of phylogeny is a prerequisite for reconstructing evolutionary patterns, and it is the basis for systematic classification, but as of yet there has been no comprehensive analysis of blackbird phylogeny. The central concern of my dissertation research was to infer of the pattern of evolutionary relationships among New World blackbird species using molecular phylogenetic methods. Because results of earlier studies informed later ones, the three chapters herein are arranged in chronological sequence, though each may be read separately from the others.

In Chapter 1, I investigate the phylogeny of the grackles (*Quiscalus* spp.), a group of 7 blackbird species that are distributed throughout North America, the Caribbean, and northern South America, and which are common in lawns, gardens, and other anthropogenic habitats. I used gene sequences of cytochrome *b* and NADH dehydrogenase subunit 2 (ND2) to reconstruct relationships within the group. A primary concern was determining the phylogenetic position and genetic distinctiveness of the extinct Slender-billed Grackle (*Q. palustris*) and of the Nicaraguan Grackle (*Q. nicaraguensis*), which is unusual among grackles for its restricted geographic range. The recovered phylogeny reveals Slender-billed Grackle to be most closely related to one of two major haplotype clades of Great-tailed Grackle (*Q. mexicanus*), the other being sister to Boat-tailed Grackle (*Q. major*). Nicaraguan Grackle appears sister to Carib Grackle (*Q. lugubris*). I also found that several species (e.g. *Q. mexicanus*, *niger*, and *lugubris*) contain deeply divergent lineages.

In Chapter 2, I present a method for partitioning whole mitochondrial genome sequences to optimize model-fitting during phylogenetic analyses. Because standards for the rigorous and objective use of mitogenomes in phylogenetic analyses were lacking, developing such a method was a prerequisite for analyzing the mitogenomes of a clade of South American endemic genera of blackbirds that I sequenced in hopes of inferring a more robust hypothesis of the phylogenetic history of the group than had been possible using a conventional 2-kilobase sample of mitochondrial DNA. I found that the most useful categories for partitioning mitochondrial genomes into more homogenous sets of sites for phylogenetic analyses, were codon position, RNA secondary structure pairing, and the coding/noncoding distinction, and that a scheme with nine data groups outperformed all of the more complex alternatives (up to 44 data groups) that I tested. As hoped, analyses using whole mitogenomic sequences yielded better-resolved and more strongly-supported hypotheses of the phylogenetic history of that locus than did a dataset composed of the sequences of two mitchondrial genes (cytochrome *b* and ND2).

In Chapter 3, I present the first comprehensive species-level phylogeny of the Icteridae. By using mitochondrial gene sequences from all ~108 currently-recognized species, together with strategic sampling of 4 nuclear loci and whole mitochondrial genomes at the generic level, I was able to resolve most

relationships with high confidence. The best-resolved phylogeny is consistent with strongly-supported results of past studies, but it also contains many robustly-resolved inferences of relationship that eluded them. These novel hypotheses of relationship include some unexpected placements of taxa that had not been included in previous molecular phylogenies, resolution of the relationships among major subclades within Icteridae, and resolution of generic-level relationships within the largest of those subclades. I suggest taxonomic revisions based on those results.

CHAPTER 1

A complete species-level phylogeny of the grackles (*Quiscalus* spp.), including the extinct Slender-billed Grackle, inferred from mitochondrial DNA[1,2]

SUMMARY

The grackles (*Quiscalus* spp.), together with their sister genus *Euphagus*, compose a clade within the New World blackbirds (Icteridae). We used gene sequences of cytochrome *b* and NADH dehydrogenase subunit 2 (ND2) to reconstruct relationships within this group. A primary concern was determining the phylogenetic position and genetic distinctiveness of the extinct Slender-billed Grackle (*Q. palustris*)—a poorly known endemic of the Lerma Basin and the ancient lakes of the Valley of Mexico, last collected and recorded in 1910—and of the Nicaraguan Grackle (*Q. nicaraguensis*), which is likewise unusual among grackles for its restricted geographic range. Our analysis differs from previous efforts by inclusion of these taxa along with all other recognized grackle species, intraspecific sampling of Greater Antillean (*Q. niger*), Carib (*Q. lugubris*), and Great-tailed (*Q. mexicanus*) Grackles, and inclusion of additional sequence data. The recovered phylogeny reveals Slender-billed Grackle to be most closely related to one of two major haplotype clades of Great-tailed Grackle, the other being sister to Boat-tailed Grackle (*Q. major*). Nicaraguan Grackle appears sister to Carib Grackle (*Q. lugubris*). We discuss the implications of these and other relationships in the genus for species limits and biogeography.

1-1. INTRODUCTION

The grackles (*Quiscalus*) are among the most familiar of blackbirds (Icteridae). Common in anthropogenic landscapes, they are conspicuous by virtue of their gregariousness, their habit of foraging on the ground in open areas, their iridescent black or rich rusty brown (in some females) plumages, and their long wedge-shaped tails, which they flare and keel distinctively in flight and display. The seven species currently recognized (AOU 1998) are all very similar with respect to morphology, behavior, and ecology.

---

[1] A version of this paper was published:

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. Condor 110:718–728.

[2] Acknowledgements: We thank G. D. Weiblen for use of his lab during preparation of *Q. palustris* and other toe pad samples, J. M. DaCosta for sharing *Q. mexicanus* sequence data, H. Vázquez-Miranda for helpful discussion and Spanish translation of our abstract, and two anonymous reviewers for comments on the manuscript . We also thank the following institutions for maintaining and making available the specimens that made this research possible: American Museum of Natural History, James Ford Bell Museum of Natural History, Marjorie Barrick Museum of Natural History, Field Museum of Natural History, University of Kansas Natural History Museum, Natural History Museum of Los Angeles County, University of New Mexico Museum of Southwestern Biology, and the National Museum of Natural History. This study was supported in part by NSF DEB-0316092 to FKB and SML.

Nevertheless, across their collective range, which encompasses much of North America and the Caribbean and extends to the north Pacific and Caribbean coasts of South America (Fig. 1-1), a great number of forms (currently 30 ssp.; Jaramillo and Burke 1999) have been described. At least 15 of these taxa were recognized as species (Ridgway 1902), but with better understanding of the relative amounts of morphological difference among them, many species, especially island forms, were subsequently reduced to subspecific status (Peters 1921, Hellmayr 1937).

Although several studies (Avise and Zink 1988, Zink et al. 1991, Lovette et al. 1999, Kerr et al. 2007, DaCosta et al. 2008) have examined intraspecific molecular variation in grackles, only one (DaCosta et al. 2008) was a detailed phylogeographic analysis. That study, of Great-tailed (*Q. mexicanus*) and Boat-tailed (*Q. major*) Grackles, found that the former species comprises two geographically distinct clades, the eastern of which is more closely related to Boat-tailed Grackle than to the western clade, thus rendering Great-tailed Grackle paraphyletic. Similarly, Lovette et al. (1999) reported that an unpublished study of the Carib Grackle (*Q. lugubris*) found that it comprises at least two geographically distinct lineages that are >3% genetically divergent; however, the significance of this finding for grackle phylogenetics has not been investigated. All other molecular studies of the grackles (Lanyon 1994, Johnson and Lanyon 1999, Lanyon and Omland 1999, Eaton 2006) have reconstructed phylogenetic relations among recognized species using a single sample of each. Furthermore, two species, both unusual for their extremely limited distributions (Fig. 1-1), have not been included in any molecular studies—the Slender-billed (*Q. palustris*) and Nicaraguan (*Q. nicaraguensis*) grackles.

The Slender-billed Grackle (hereafter *palustris*) is the only blackbird (and one of a handful of New World nine-primaried oscines) driven to extinction in historic times. It was endemic to central Mexico, but the lake and marsh systems in which it lived have been extensively drained and diminished over the past five centuries for agricultural and urban development (Peterson and Navarro-Sigüenza 2006). The only records of *palustris* are from the Valley of Mexico (now the metropolitan area of Mexico City), where the type specimen was collected in 1827 and the species was reportedly still present in ~1890 (Peterson 1998), and from the headwaters of the Río Lerma, where specimens were collected in 1904 and 1910 (Dickerman 1965). The Nicaraguan Grackle (hereafter *nicaraguensis*) is an endemic of the marshes and lowlands around Lakes Managua and Nicaragua, where populations appear secure (IUCN 2008).

Because of their distinctive morphologies, recognition of *palustris* and *nicaraguensis* as species has rarely been questioned, but traditional hypotheses of grackle evolutionary relationships, and of these species particularly, have varied. Past taxonomies (e.g. Ridgway 1902, Hellmayr 1937) have grouped the species of grackles into three genera—*Quiscalus*, *Holoquiscalus*, and *Cassidix*—which are still sometimes recognized as subgenera. The long, straight, fine bill of *palustris* was unlike that of any other grackle (Jaramillo and Burke 1999), but since that species was otherwise similar to *mexicanus* and *major*, all authorities have grouped *palustris* with them in *Cassidix* (Table 1-1). Blake (1968) suggested *palustris* might have been a local race of *mexicanus*, whereas Selander and Giller (1961) speculated on a sister pairing with *major* and possible connection with *nicaraguensis* based on the marsh-nesting habits of these

species. Generally, *nicaraguensis* has been included in *Cassidix*, perhaps because of its long tail and marsh association. Hellmayr (1937) considered *nicaraguensis* to be "allied" with *palustris*. However, some authorities (Bond 1950, Lack 1976) thought *nicaraguensis* to be closely related to Greater Antillean (*Q. niger*) and Carib (*Q. lugubris*) grackles. The latter two species have been considered a superspecies in the subgenus *Holoquiscalus* (Jaramillo and Burke 1999). Although Ridgway (1902) placed *nicaraguensis* in *Cassidix* (using the synonym *Megaquiscalus*), he noted that it shares a feather structural character with most forms of *Holoquiscalus*. He considered these subgenera more closely related to one another than to the Common Grackle (*Q. quiscula*) on the basis of morphological similarities (regarding *nicaraguensis* and *lugubris* as intermediate forms), as did Jaramillo and Burke (1999) based on similarities in plumage and voice. Yang and Selander (1968) noted that displays and vocalizations of *nicaraguensis* are similar to *quiscula, niger* and *lugubris*. They also proposed a derivation of *quiscula* from *niger*.

Previous formal analyses of grackle phylogeny have used either morphological or molecular characters. Björklund's (1991) phylogeny, based on 23 morphological characters, placed *nicaraguensis* sister to Rusty Blackbird (*Euphagus carolinus*) and did not recover *Quiscalus*, *Euphagus*, or their union as monophyletic. However, a reanalysis of Björklund's data (by Johnson and Lanyon 1999) revealed that only one node (pairing *major* and *niger*) in his tree had ≥50% bootstrap support. Johnson and Lanyon (1999) used mtDNA to analyze relationships among the grackles and related blackbirds using parsimony; Eaton (2006) reanalyzed the same dataset using maximum likelihood and Bayesian methods. These analyses recovered *Quiscalus* as monophyletic and sister to *Euphagus*. Subgenus *Holoquiscalus* was rendered paraphyletic by the closer relationship of *niger* to the *Cassidix* group than to its previously supposed sister taxon, *lugubris.*

Our primary objectives in this study were to determine the phylogenetic positions and genetic distinctiveness of *palustris* and *nicaraguensis* in the context of a wider reevaluation of grackle relationships inferred from the sequences of two protein-coding mitochondrial genes, cytochrome *b* and NADH dehydrogenase subunit 2 (ND2). Our analysis differs from the most comprehensive previous efforts (Johnson and Lanyon 1999, Eaton 2006) insofar as we used ~20% more sequence data (2292 base pairs total), included for the first time all recognized species of grackles, and included multiple representatives of species known—as in the cases of *lugubris* and *mexicanus*—or suspected (as in *niger*) to harbor deep phylogenetic divergences among populations.

1-2. METHODS

1-2.1. Taxon sampling.—Ingroup sampling (Table 1-1) included at least one individual from each species of *Quiscalus* (Sibley and Monroe 1990, AOU 1998). Further, we included representatives (provided by J. M. DaCosta at the Marjorie Barrick Museum) of the principle haplotype clades known from *mexicanus* (DaCosta et al. 2008), and we sequenced specimens of previously unsampled subspecies (*graysoni*, *obscurus*) in the western portion of the species' range. Similarly, we included samples of the divergent lineages within *lugubris* (Lovette et al. 1999) and specimens of *niger* from four of the five islands (or island

groups) on which it occurs. Previous molecular analyses of blackbird phylogeny (Johnson and Lanyon 1999, Lanyon and Omland 1999, Eaton 2006) recovered *Quiscalus* as monophyletic and sister to *Euphagus* with unequivocal support, so outgroup sampling was limited to the two recognized species of the latter genus.

1-2.2. Laboratory procedures.—We extracted genomic DNA from tissue samples (Table 1-1) using a DNeasy Tissue Kit (Qiagen, Valencia, California) following manufacturer instructions, except that for specimens sampled from toe pads, 30 µl of 100 mg per ml dithiothreitol (Gold Biotechnology, St. Louis, Missouri) was added to the digestion, and 50 µl Buffer AE used for each DNA elution. Because the *palustris* sample was taken from a toe pad of a skin prepared in 1904, its intact DNA concentration was low, so we amplified cytochrome *b* in six fragments using primers pairs ND5emb1 and H15103 (Barker et al. 2008), L15069-Qp (5'–CTAGCCATACACTACACAGCAGAC–3') and H15305-Qp (5'–CGGTAGCGCCTCAGAATGATATTT–3'), L15259-Qp (5'–GTTGGAGTCATTCTCCTCCTAA–3') and H15460-Qp (5'–GTGAACTAGGGTAAGTCCTACGAT–3'), L15410 (Barker et al. 2008) and H15709 (Barker 2004), L15656-Qp (5'–AACCTCCTAGGAGATCCAGA–3') and H15934 (Barker et al. 2008), and L15848-Qp (5'–CAAAACTACGATCAATGACYTTCCG–3') with H16137 (Sorenson et al. 1999). The *mexicanus* toe pad samples were treated similarly but were amplified in five fragments (H15305 and L15259-Qp were not used), and L15656 (Helm-Bychowski and Cracraft 1993) was substituted for L15656-Qp. Reaction preparation and cycling parameters were as described by Barker et al. (2008). The initial products were reamplified when necessary to obtain sufficient concentrations for sequencing.

We amplified cytochrome *b* from frozen tissue specimens using ND5emb2 (5'–GGYCTAAYCAAAGCCTAYCTA–3') and H16137, reamplifying when necessary with primer pairs ND5emb1 and H15305-Qp, L15069-Qp and H15709, and L15656-Qp with H16137. To amplify ND2 from frozen tissue specimens we used primers LMET (Hackett 1996) and H1064 (Drovetski et al. 2004). We obtained the complete sequence of ND2, and ~890 bp of cytochrome *b* sequence of seven individuals in our study from GenBank; for these same individuals, in order to complete the cytochrome *b* sequences, we amplified the regions at each end of the gene using primers ND5emb1 with H15103, and L15848-Qp with H16137.

We purified PCR products through enzymatic digestion using exonuclease 1 and shrimp alkaline phosphatase (USB Corporation, Cleveland, Ohio), following Werle et al. (1994), and sequenced them following manufacturer recommendations on an ABI 3700 automated sequencer (BigDye v3.1, Applied Biosystems, Foster City, California) at the Biomedical Genomics Center of the University of Minnesota. We used the same primers as for PCR, except that for whole-gene amplifications, we used some additional, internal primers. These were L5758emb (Barker et al. 2008) and H5766emb (Barker et al. 2008) for ND2, and for cytochrome *b*, they were L15191 (Lanyon and Hall 1994), L15656, H15709, and H15298 (Helm-Bychowski and Cracraft 1993). We used Sequencher 4.7 (Genecodes, Ann Arbor, Michigan) to align and view chromatograms of complimentary reads and overlapping fragments to produce a consensus sequence.

1-2.3. Phylogenetic analyses.—Two sets of analyses were performed. The primary set utilized all data acquired, which for most individuals comprised the complete ND2 (1040 bp) and cytochrome *b* (1143 bp) gene sequences, plus a spacer region and tRNA sequence (108 bp) adjoining the latter. The second set utilized only the cytochrome *b* (with spacer and tRNA) data, and was conducted to determine whether the results of the primary set were distorted by the lack of ND2 data from individuals sequenced from toe pads (Table 1-1).

We used PAUP* 4.0b10 (Swofford 2002) to infer phylogenetic relations among taxa under maximum parsimony and maximum likelihood criteria, and used MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) to employ Bayesian methods. Parsimony analyses were conducted with branch-and-bound searches. Support for nodes was assessed, after excluding uninformative characters, with 10000 nonparametric bootstrap replicates. We used DT-ModSel (Minin et al. 2003) to select the most appropriate models for analyses of the data using maximum likelihood and Bayesian methods. Parameter values for maximum likelihood analyses were obtained through estimation on maximum parsimony trees using PAUP*. Heuristic searches for the best maximum likelihood tree for each dataset were conducted using tree bisection and reconnection branch-swapping with 1000 random addition sequences of taxa. Clock- and nonclock-like models of sequence evolution on this tree were not significantly different ($\alpha = 0.05$) for the complete dataset according to a likelihood ratio test ($\delta = 14.1$, df = 19, $P = 0.78$), but *were* different for cytochrome *b* alone ($\delta = 31.7$, df = 19, $P = 0.03$). We conducted a second set of heuristic searches for the best maximum likelihood tree for each dataset with clock-like evolution enforced. The topologies of the best clock and nonclock trees for cytochrome *b* alone were slightly different within the eastern clade of *mexicanus*, but clock and nonclock models were not significantly different when evaluated on these trees ($\delta = 14.0$, df = 19, $P = 0.79$). Support for nodes in maximum likelihood trees were assessed with 500 bootstrap replicates (10 random addition sequences each). Sequence divergences were calculated in PAUP* using nonclock maximum likelihood parameter estimates.

We used empirical Bayes factors (Kass and Raftery 1995, Nylander et al. 2004) to select partitioning schemes for Bayesian analyses of each dataset. We tested unpartitioned analyses, partitioning by gene, codon position, and codon position by gene, and the effect of assuming clock-like evolution. For both datasets, a maximally partitioned scheme (including a partition for the spacer and tRNA sequence) and enforcing a strict clock was best ($2\log_e B_{10} = 19$ for the complete dataset and $2\log_e B_{10} = 25$ for cytochrome *b* in comparison to the next best models). We conducted final analyses for each dataset with and without enforcing clock-like evolution. For each analysis, we ran four coupled chains for three million generations, sampling every 100. Samples prior to reaching stationarity were discarded as "burn-in" and the remaining subsamples used to create 50% majority-rule consensus trees.

1-3. RESULTS

Parsimony, likelihood, and Bayesian methods recovered similar patterns of relationship. Well supported nodes inferred using cytochrome *b* sequences alone (Fig. 1-2, Table 1-2) were consistent with the better supported and more finely resolved topology generated using both cytochrome *b* and ND2 (but lacking ND2 data for four specimens), so we consider the latter (Fig. 1-3, Table 1-2) our best estimate of the phylogeny of the group.

In accordance with DaCosta et al. (2008), we found two highly divergent clades of *mexicanus*. One clade is distributed mostly west of the Sierra Madre Occidental, and the other, to the east of that range. All analyses strongly supported sister relationships between *palustris* and the western clade of *mexicanus* and between *major* and the eastern clade. Amounts of cytochrome *b* sequence divergence (Table 1-3) were modest within the western and eastern clades of *mexicanus*, averaging 0.3% (range: 0.3%–0.4%) and 0.5% (range: 0.2%–0.8%) respectively, in comparison to the 3.1% divergence between those clades and to their divergences from their sister taxa. *Q. palustris* was 2.0% divergent from western *mexicanus*, whereas *major* was 1.4% divergent from the eastern clade. A sister relationship between the *palustris*-western *mexicanus*, and *major*-eastern *mexicanus* clades was weakly to modestly supported using the full dataset, but the cytochrome *b* data alone were unable to dichotomously resolve the relationships among these clades and *niger*. The *niger* samples composed a strongly supported monophyletic unit; sequence divergences among islands averaged 1.3% (range: 0.5%–1.9%).

We found *nicaraguensis* to be sister to *lugubris*, but relationships among *nicaraguensis* and the two highly divergent (3.9% in cytochrome *b*; Table 1-3) *lugubris* subspecies were not well resolved. Using the complete dataset, all analyses recovered these three lineages as a clade with modest to strong support, and Bayesian and maximum likelihood analyses recovered *lugubris* as monophyletic with weak support (Fig. 1-3). Using the cytochrome *b* data alone, maximum likelihood and clock-enforced Bayesian analyses recovered these three lineages as a clade, but support for this relationship, and for a monophyletic *lugubris* was lacking. Under parsimony, a sister relationship between *nicaraguensis* and the Lesser Antillean lineage of *lugubris* received weak bootstrap support, whereas maximum likelihood and Bayesian analyses grouped *nicaraguensis* and mainland *lugubris* with weak support (Fig. 1-2).

1-4. DISCUSSION

Our best resolved phylogeny of the grackles (*Quiscalus* spp.) is consistent with earlier molecular analyses (Johnson and Lanyon 1999, Eaton 2006) but reveals a more complex pattern of relationships than previously recognized. We discovered the phylogenetic positions of two species—*palustris* and *nicaraguensis*—not included in previous molecular studies, found these taxa genetically distinct from their closest relatives, and found that some named species comprise deeply divergent lineages. Our phylogeny conflicts with some traditional notions about the relationships among grackle species.

1-4.1. Sister lineage divergences and intraspecific variation.—Our analyses confirmed that *Q. mexicanus* comprises two deeply divergent haplotype clades (Wehtje 2004, DaCosta et al. 2008) that are

also geographically distinct, suggesting that they represent lineages that split ~2 million years ago (assuming 1.6% divergence per million years; Fleischer et al. 1998, but see Weir and Schluter 2008). The western clade corresponds to subspecies *nelsoni* and *graysoni*, notable for being the smallest *mexicanus* forms and having very pale female plumage (Jaramillo and Burke 1999). Prior to the recent expansion of *nelsoni* northward into California and the desert southwest of the United States (Phillips 1950, Wehtje 2003, 2004), this clade was restricted to Sonora and coastal Sinaloa west of the Sierra Madre Occidental (Fig. 1-4). The eastern clade, composed of the larger *mexicanus* forms with dark brown female plumages, was represented in our study by subspecies *mexicanus*, *monsoni*, and *obscurus*, but DaCosta et al. (2008) found that it also includes *prosopidicola* and *peruvianus*, and thus likely encompasses all *Q. mexicanus* outside the western clade. Prior to its recent expansion to the Gulf Coast, Great Plains, and southwest of the United States (Wehtje 2003, 2004) this clade was distributed from north-central Mexico east to the Gulf Coast and south through Central America to coastal northern South America. Despite its wide distribution, we found that divergences within this clade were shallow, poorly supported, and imperfectly congruent with named subspecies (see also DaCosta et al. 2008).

In spite of their genetic and morphological divergence, the eastern and western *mexicanus* clades interbreed freely (Johnson and Peer 2001, Wehtje 2004) in the southwestern United States, where their distributions now overlap due to expansion over the past ~60 years. Ours is the first molecular study to include *graysoni* and *obscurus* and thereby discover that these subspecies belong to the western and eastern *mexicanus* clades respectively, a biogeographically important finding because their point of contact (Fig. 1-4) along the Pacific slope of the Sierra Madre Occidental just north of its juncture with the Sierra Madre del Sur represents the only known zone of overlap between these clades prior to recent range expansions. The extent to which *graysoni* and *obscurus* interbreed is unknown, but interestingly, these forms sing "noticeably different" songs (Johnson and Peer 2001) and represent the extremes of female plumage variation within *Q. mexicanus*, with *graysoni* being pale grey-buff and many female *obscurus* being nearly black (AFLAP pers. obs., Jaramillo and Burke 1999).

The sister relationship between *palustris* and the western *mexicanus* clade that we discovered, and the sister relationship between *major* and the eastern clade (see also DaCosta et al. 2008) render *mexicanus* paraphyletic on two counts. Although this result might lead some to question the specific status of both *palustris* and *major*, both species are morphologically and ecologically distinct, and their genetic divergences from the *mexicanus* clades to which they are most closely related are large in comparison to the average amounts of divergence we found within those clades and the 0.2% maximum found within *major* (DaCosta et al. 2008). Furthermore, interbreeding is rare enough between *major* and eastern *mexicanus* where their ranges now overlap along the Gulf Coast that they are distinct under the Biological Species Concept (Selander and Giller 1961, Post et al. 1996). Thus, interbreeding of the *mexicanus* clades likely reflects retention of ancestral compatibilities as seen also in geese (Paxinos et al. 2002). It seems likely that the morphological similarities, including large size and long tail, of the *mexicanus* clades, *major*, and *palustris* are indicative of their collective monophyly, and that the small size and short tail of *niger*

represents retention of ancestral characteristics also seen in *lugubris*; however, support for this constellation of relationships was weak. It is conceivable that the *mexicanus-palustris* and *mexicanus-major* clades are not sister taxa, and that the morphological similarity of *niger* to *lugubris* is a result of convergence or reversion to ancestral characteristics brought about through adaptation to the Caribbean island context.

The 1.3% average genetic divergence among *niger* samples was much greater than that seen within the *mexicanus* clades and *major* and the 0.3% known from *quiscula* (Zink et al. 1991). The magnitudes of divergences among islands suggest relatively great evolutionary independence among populations compared to these other species, and their pattern suggests a history of island colonization from Puerto Rico westward. However, this pattern could be coincidental, and additional sampling would be necessary to establish whether our samples characterize interisland haplotype differences (i.e., whether island populations are reciprocally monophyletic).

The 3.9% difference between haplotypes of *lugubris* samples from South America and the Lesser Antilles is a result we anticipated based on comments by Lovette et al. (1999). They reported that *lugubris* haplotypes in Barbados were identical to some in Trinidad, but that samples from St. Vincent and the rest of the Lesser Antilles differed from these by >3% in mtATPase sequences. This substantial genetic divide does not correspond to known patterns of morphological similarity among subspecies (e.g., the forms on Barbados and St. Vincent, which have blackish rather than brown female plumages, were once classified together as a species separate from all other *lugubris* ssp.; Peters 1921). Somewhat surprisingly, we found that *nicaraguensis*, which is morphologically and behaviorally distinct, is at most only marginally more genetically divergent from these lineages than they are from one another. Given the lack of support for resolving relationships among these three lineages and their apparently very long histories of evolutionary independence, the *lugubris* lineages are probably best regarded as separate species. The similar morphologies of Antillean *lugubris*, mainland *lugubris*, and *niger* suggest that their appearances are conserved from the common ancestor of all *Quiscalus* apart from *quiscula*.

1-4.2. Implications for higher-level relationships, grackle evolution, and biogeographic history.— Our analyses produced a robust phylogeny for *Quiscalus* that contradicts some past suppositions about grackle relations. First, it does not support the use of the subgenera *Holoquiscalus* and *Cassidix*. The former is rendered paraphyletic, even if *nicaraguensis* were incorporated into it, by the closer relationship of *niger* to *mexicanus*, *major*, and *palustris* than to *lugubris*. The latter subgenus is polyphyletic due to the position of *nicaraguensis*; in addition, support for the monophyly of the remainder of *Cassidix* is only modest. Second, the species with the strongest ties to marsh habitat—*palustris*, *major*, *nicaraguensis*, and the *peruvianus* form of *mexicanus*—do not compose a clade, nor is it clear that this association is symplesiomorphic, as has been asserted (Selander and Giller 1961, Yang and Selander 1968). Third, the richly buff and pale plumage components of females in the aforementioned taxa (Selander and Giller 1961, Jaramillo and Burke 1999), and the strong sexual dimorphism of tail length in these species are likewise not indicative of close relationships. Finally, *quiscula* appears to be sister to all other *Quiscalus*, so the

hypothesis that it was derived from an ancestral *niger* isolated in Florida (Yang and Selander 1968) is not supported.

The basal divergences of *Euphagus* and *quiscula* suggest a continental North American origin for *Quiscalus*, but our phylogeny does not lend itself to simple inferences about the early biogeography of the genus (the sister to the *Quiscalus-Euphagus* clade, *Dives* (FKB and SML, unpubl. data), is represented by one species each in Central America, South America, and the Caribbean). Rather, it perhaps accurately reflects a history of dynamic fluctuations in distributions during the climatically dynamic late Pliocene and Pleistocene when diversification of the clade appears to have occurred. Such a history is expected given that grackles are water-associated creatures of savanna, woodland edge, and open marsh—ephemeral habitats that vary considerably in their distribution and extent over time—and, as seen from the range expansions of several species in the past century, they are capable of rapid population and distributional responses to habitat availability.

The curiously localized *palustris* appears to have diverged ~1.2 million years ago from its geographically distant sister, western *mexicanus*. *Q. mexicanus* is now common throughout the interior of Mexico, but until the last century, its distribution was more limited, and it was most common on the coastal plains (Christensen 2000). Consequently, eastern *mexicanus*, western *mexicanus*, and *palustris* had mostly allopatric distributions. It is the eastern clade, in the form of subspecies *mexicanus*, that has very recently spread into central Mexico where *palustris* once occurred. Whether these taxa ever came into contact is an interesting subject for speculation in light of questions of species limits and the possibility that ecological competition had a role in the extinction of *palustris*. The first modern occurrence of *Q. mexicanus* in the Valley of Mexico was in ~1960 at Xochimilco (Dickerman 1965), the same area to which *palustris* was confined in ~1890 when last reported in the valley (Peterson 1998). However, historical accounts indicate that, in ~1500, *mexicanus* was introduced to the valley from the Gulf Coast by the Aztecs (Haemig 1978). Common in ~1570, it perhaps declined over the next century and disappeared due to changes in land use (Christensen 2000).

The species of grackles, judging from pairwise divergences, appear to be quite recent in comparison to other congeneric North American birds (Klicka and Zink 1997), making them well suited to studies of processes related to speciation. The group allows for comparison of closely related species that have diverged substantially in size or in sexual size dimorphism, as well as those that are genetically divergent yet morphologically similar. Furthermore, range expansions in the past century have brought several of these formerly allopatric species and lineages into secondary contact or more extensive sympatry, thus allowing for study of interspecific (or interclade) interactions with respect to ecological competition, behaviors related to mate attraction and selection, and the consequences of interclade hybridization and introgression. In addition, phylogeographic and population genetic studies of *mexicanus*, *niger*, and *lugubris* are needed to better understand their present population structuring and evolutionary histories. We hope that this first complete species-level phylogeny of the grackles will stimulate further work on these and other aspects of the group's evolution, ecology, and behavior.

Table 1-1. Specimens sequenced for reconstructing phylogenetic relationships among the grackles (*Quiscalus* spp.) and *Euphagus* blackbirds using mitochondrial cytochrome *b* and ND2 gene sequences.

Genus [Subgenus], species, and subspecies	English name	Specimen [a] [Source][b]	Collecting locality	GenBank numbers (cytochrome *b*; ND2)
Euphagus carolinus	Rusty Blackbird	FMNH 333317 [G]	USA: Illinois; Chicago [c]	AF089023; AF109950
Euphagus cyanocephalus	Brewer's Blackbird	FMNH 341985 [G]	USA: California; Jolon	AF089024; AF109951
Quiscalus [*Quiscalus*] *quiscula versicolor*	Common Grackle	FMNH 341733 [G]	USA: Illinois; Chicago	AF089058; AF109956
Q. [*Holoquiscalus*] *lugubris lugubris*	Carib Grackle	FMNH 339797 [G]	Venezuela: Falcon; Tucacas	AF089054; AF109952
Q. [*Holoquiscalus*] *lugubris contrusus* [d]	Carib Grackle	USNM 612608 [F]	Saint Vincent	FJ389562; FJ389553
Q. [*Cassidix*] *nicaraguensis* [d]	Nicaraguan Grackle	MBM 4375 [F]	Nicaragua: Tipitapa	FJ389558; FJ389549
Q. [*Holoquiscalus*] *niger niger* [d]	Greater Antillean Grackle	KUMNH 6400 [e] [F]	Dominican Republic	FJ389559; FJ389550
Q. [*Holoquiscalus*] *niger brachypterus* [d]	Greater Antillean Grackle	FMNH 350794 [F]	Puerto Rico: Cabo Rojo	FJ389561; FJ389552
Q. [*Holoquiscalus*] *niger caymanensis* [d]	Greater Antillean Grackle	AMNH DOT6818 [F]	Cayman Islands: Grand Cayman	FJ389560; FJ389551
Q. [*Holoquiscalus*] *niger crassirostris*	Greater Antillean Grackle	FMNH 331153 [G]	Jamaica: Trelawny Parish	AF089057; AF109955
Q. [*Cassidix*] *major major*	Boat-tailed Grackle	FMNH 341918 [G]	USA: Louisiana; Cameron	AF089055; AF109953
Q. [*Cassidix*] *mexicanus mexicanus*	Great-tailed Grackle	MBM 6978 [D]	Honduras: Departamento Copán	FJ389563; FJ389554
Q. [*Cassidix*] *mexicanus mexicanus*	Great-tailed Grackle	MBM 838 [D]	Mexico: Distrito Federal; Mexico City	FJ389565; EU414578
Q. [*Cassidix*] *mexicanus monsoni*	Great-tailed Grackle	MBM JMD1014 [D]	USA: Texas; El Paso	FJ389564; FJ389555
Q. [*Cassidix*] *mexicanus monsoni*	Great-tailed Grackle	MSB NK 116184 [D]	USA: New Mexico; Belen	FJ389567; EU414562
Q. [*Cassidix*] *mexicanus obscurus* [d]	Great-tailed Grackle	MMNH 19363 [T]	Mexico: Guerrero; Amatlán	FJ389569; NA
Q. [*Cassidix*] *mexicanus obscurus* [d]	Great-tailed Grackle	MMNH 10972 [T]	Mexico: Jalisco; Puerto Vallarta	FJ389568; NA
Q. [*Cassidix*] *mexicanus graysoni* [d]	Great-tailed Grackle	MMNH 22343 [f] [T]	Mexico: Nayarit; Tecuala	FJ389570; NA
Q. [*Cassidix*] *mexicanus nelsoni*	Great-tailed Grackle	LACM 6173 [D]	USA: California; Fillmore	FJ389566; FJ389556
Q. [*Cassidix*] *mexicanus nelsoni*	Great-tailed Grackle	FMNH 341975 [G]	USA: California; Calipatria	AF089056; AF109954
Q. [*Cassidix*] *palustris* [d]	Slender-billed Grackle	USNM 194170 [T]	Mexico: Estado de México; Lerma	FJ389557; NA

[a] Museum catalog number. Abbreviations: AMNH = American Museum of Natural History, MMNH = James Ford Bell Museum of Natural History, FMNH = Field Museum of Natural History, KUMNH = University of Kansas Natural History Museum, LACM = Natural History Museum of Los Angeles County, MBM = Marjorie Barrick Museum of Natural History, MSB = University of New Mexico Museum of Southwestern Biology, USNM = National Museum of Natural History.

[b] Specimen type and sequence source abbreviations: D = sequenced by J. M. DaCosta from frozen tissue, F = sequenced from frozen tissue specimen, G = downloaded from GenBank and sequence added from frozen tissue of same specimen, T = sequenced from toe pad of dried study skin.

[c] Given the location and date (1 April 1986), a migrant.

[d] Taxon not included in previous phylogenies of *Quiscalus*.

[e] A second specimen from the same locality (KUMNH 6401) was of identical haplotype.

[f] A 277 base-pair fragment of cytochrome *b* from another *graysoni* (MMNH 20301; Mexico: Sinaloa; Mazatlán) matched this specimen's sequence.

Table 1-2. Dataset description, model parameter values, and analysis summaries for the two mtDNA datasets (ND2 with cytochrome *b*, and cytochrome *b* alone) used in reconstructing phylogenetic relationships among the grackles (*Quiscalus* spp.) and *Euphagus* blackbirds. For maximum likelihood (ML) and some Bayesian parameters, values are given for models with and without clock-like sequence evolution enforced. CI = ensemble consistency index, RI = ensemble retention index, π_i = base frequency, Nst = number of substitution types, TI/TV ratio = transition/transversion ratio, p_{iv} = proportion of invariant sites, -ln(ℓ) = negative natural log likelihood of best tree.

	ND2 + cytochrome *b*	cytochrome *b*
Sequence length	2292	1251
Number of variable sites	373	177
	Parsimony Analyses	
Number of informative sites	189	97
Number of trees	10	12
Tree length	535	278
CI	0.736	0.680
RI	0.729	0.718
	ML and Bayesian Analyses	
	ML (ML with clock) [Bayes; with clock]	ML (ML with clock) [Bayes; with clock]
Model of sequence evolution	HKY+I [a]	HKY+I [a]
Number of ML trees	1 (1)	2 (1)
Tree length	0.283 (0.286) [0.313; 0.285]	0.268 (0.265) [0.328; 0.272]
π_A	0.298 (0.299)	0.289 (0.292)
π_C	0.341 (0.342)	0.337 (0.339)
π_G	0.119 (0.119)	0.133 (0.128)
π_T	0.241 (0.241)	0.241 (0.241)
Nst	2	2
TI/TV ratio	8.18 (8.49)	8.94 (9.87)
p_{iv}	0.739 (0.740)	0.797 (0.804)
-ln(ℓ)	6016.1 (6023.1) [5618.5; 5616.3]	3236.0 (3242.9) [3022.8; 3021.8]

[a] Hasegawa et al. 1985.

Table 1-3. Genetic divergences (average pairwise %) among the grackles (*Quiscalus* spp.) and *Euphagus* blackbirds based upon analysis of cytochrome *b* (with adjoining spacer region and tRNA) under a maximum likelihood model of sequence evolution without enforcing molecular clock. Eastern and western clades of *Q. mexicanus* are listed separately. Average pairwise within-taxon divergences are shown on the diagonal for taxa with two or more samples.

		1	2	3	4	5	6	7	8	9
1	*E. carolinus*	–								
2	*E. cyanocephalus*	4.7	–							
3	*Q. quiscula*	6.8	7.2	–						
4	*Q. lugubris*	7.4	8.3	4.4	3.9					
5	*Q. nicaraguensis*	7.0	9.1	4.5	3.3	–				
6	*Q. niger*	8.1	8.8	4.6	4.3	4.2	1.3			
7	*Q. major*	6.8	8.3	4.7	3.9	3.8	2.9	–		
8	*Q. mexicanus E*	7.5	9.1	4.9	4.1	3.7	3.1	1.4	0.5	
9	*Q. mexicanus W*	7.2	8.6	5.7	4.6	4.5	3.6	2.9	3.1	0.3
10	*Q. palustris*	7.4	8.3	4.9	4.1	3.9	2.9	2.6	2.8	2.0

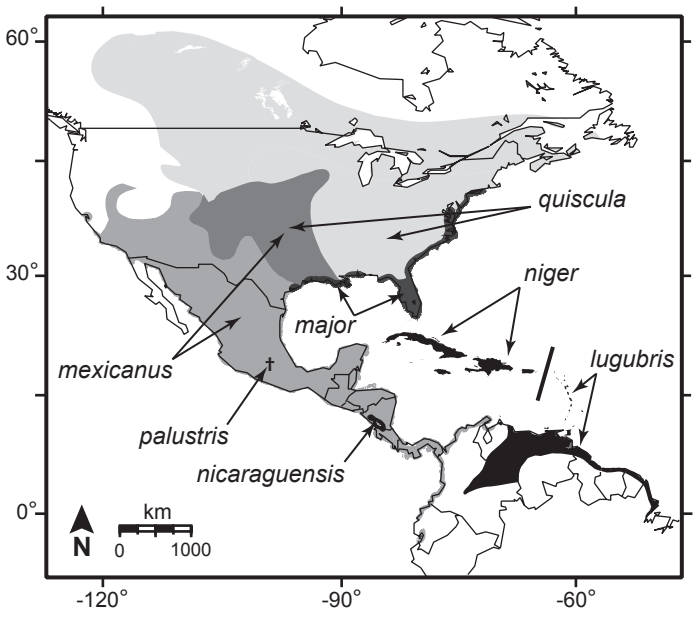

Figure 1-1. Breeding distributions (after Ridgely et al. 2007) of the currently recognized grackles (*Quiscalus* spp.). Slender-billed Grackle (*Q. palustris*) was known from two small areas, both located approximately at the center of the cross-shaped symbol. The Caribbean island distributions of the Greater Antillean (*Q. niger*) and Carib (*Q. lugubris*) grackles occur on opposite sides of the slanted black bar.

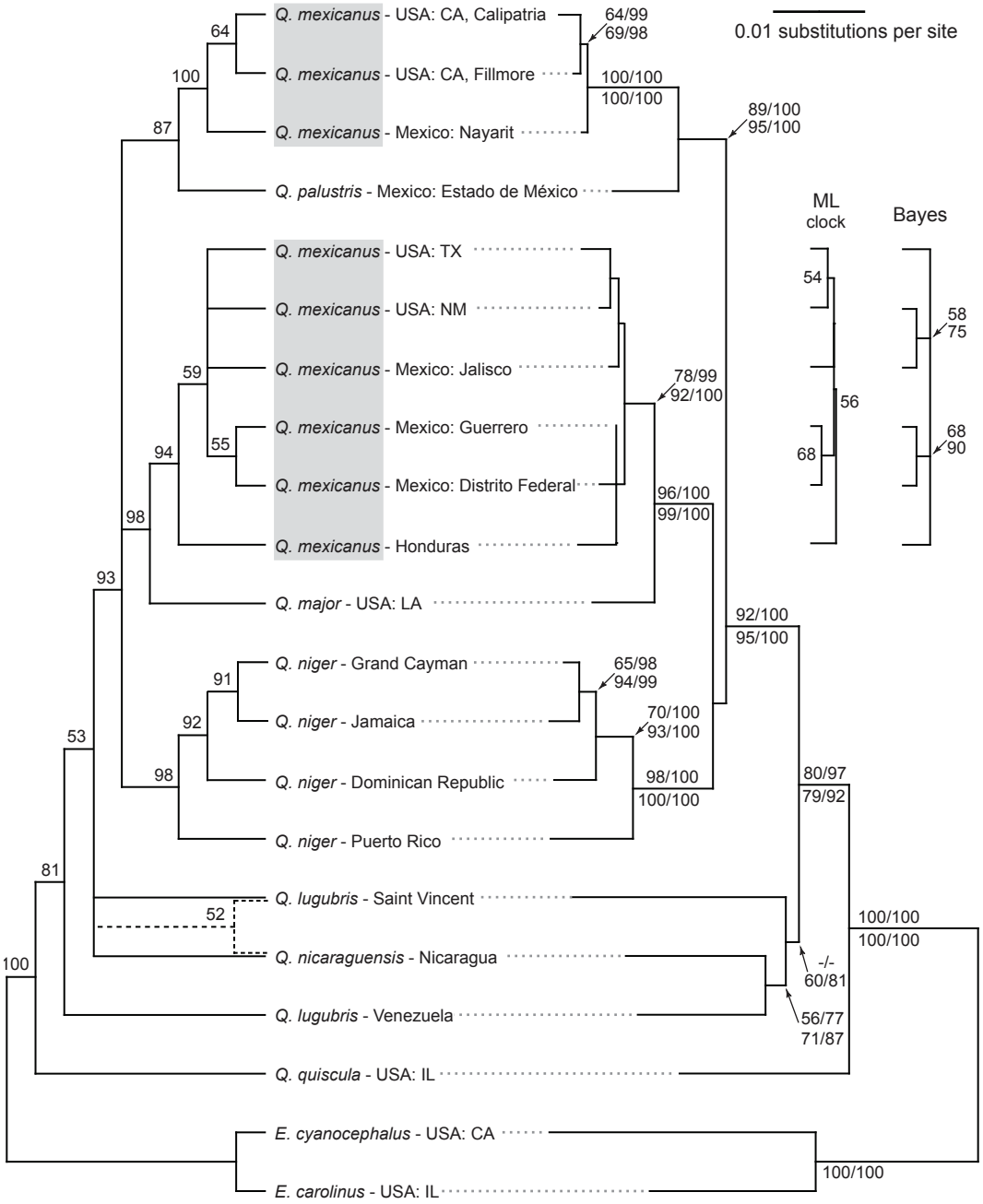

Figure 1-2. Phylogeny of the grackles (*Quiscalus* spp.; rooted using *Euphagus*), determined from analysis of mitochondrial cytochrome *b* with adjoining spacer and tRNA sequence. Left: strict consensus of 12 equally parsimonious trees with nonparametric bootstrap support values. Right: one of two best maximum likelihood trees without enforcing molecular clock. The relationships recovered within the eastern clade of *Q. mexicanus* differed according to the method of analysis; topologies found with clock-enforced maximum likelihood (ML) and with Bayesian analyses are shown at far right. Nonparametric bootstrap support values followed by estimated Bayesian posterior probabilities (×100) of nodes without molecular clock enforced are shown above values with molecular clock enforced. Arrows connect support values to nodes when they could not be fitted above and below the adjacent branches. Nodal support was <50% when indicated with a dash or not given.

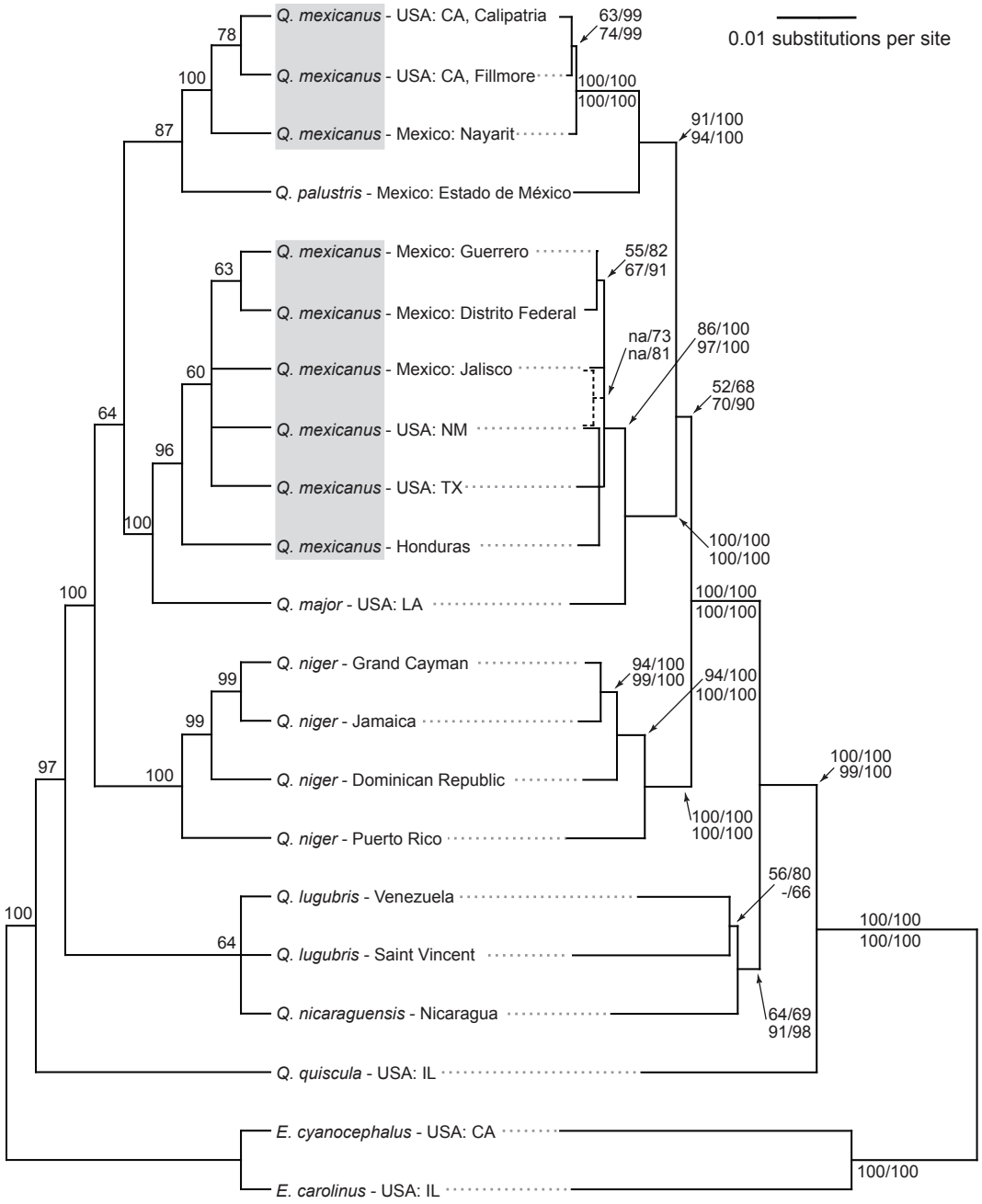

Figure 1-3. Phylogeny of the grackles (*Quiscalus* spp.; rooted using *Euphagus*) determined from analysis of mitochondrial cytochrome *b* and ND2 gene sequences. Left: strict consensus of 10 equally parsimonious trees with nonparametric bootstrap support values. Right: single best maximum likelihood tree without enforcing molecular clock (where different, topology found with Bayesian analysis shown with dashed line. Nonparametric bootstrap support followed by Bayesian posterior probabilities (×100) of nodes without molecular clock enforced are shown above values with molecular clock enforced. Arrows connect support values to nodes when they could not be fitted above and below the adjacent branches. Nodal support was <50% when indicated with a dash or not given.

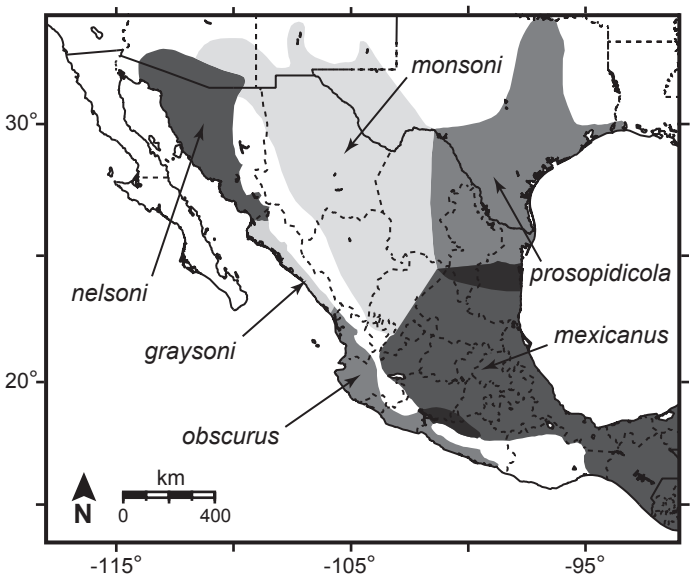

Figure 1-4. Distributions in ~1960 of the subspecies of Great-tailed Grackle (*Quiscalus mexicanus*) in Mexico and the United States (after Selander and Giller 1961). Unlabeled dark shading indicates areas of range overlap. Dashed lines indicate state boundaries.

CHAPTER 2

Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study[1,2]

SUMMARY

Whole mitochondrial genome sequences have been used in studies of animal phylogeny for two decades, and current technologies make them ever more available, but methods for their analysis are lagging and best practices have not been established. Most studies ignore variation in base composition and evolutionary rate within the mitogenome that can bias phylogenetic inference, or attempt to avoid it by excluding parts of the mitogenome from analysis. In contrast, partitioned analyses accommodate heterogeneity, without discarding data, by applying separate evolutionary models to differing portions of the mitogenome. To facilitate use of complete mitogenomic sequences in phylogenetics, we (1) suggest a set of categories for dividing mitogenomic datasets into subsets, (2) explore differences in evolutionary dynamics among those subsets, and (3) apply a method for combining data subsets with similar properties to produce effective and efficient partitioning schemes. We demonstrate these procedures with a case study, using the mitogenomes of species in the grackles and allies clade of New World blackbirds (Icteridae). We found that the most useful categories for partitioning were codon position, RNA secondary structure pairing, and the coding/noncoding distinction, and that a scheme with nine data groups outperformed all of the more complex alternatives (up to 44 data groups) that we tested. As hoped, we found that analyses using whole mitogenomic sequences yielded much better-resolved and more strongly-supported hypotheses of the phylogenetic history of that locus than did a conventional 2-kilobase sample (i.e. sequences of the cytochrome *b* and ND2 genes). Mitogenomes have much untapped potential for phylogenetics, especially of birds, a taxon for which they have been little exploited except in investigations of ordinal-level relationships.

---

[1] A version of this paper was published:

Powell, A. F. L. A., F. K. Barker, and S. M. Lanyon. 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. Molecular Phylogenetics and Evolution 66:69–79.

[2] Acknowledgements: We thank J. C. Avise and D. Walker for purified mitochondrial samples of *Agelaius* and *Molothrus*, and the following institutions for maintaining and making available the other specimens that made this study possible: Academy of Natural Sciences of Drexel University, American Museum of Natural History, Field Museum of Natural History, Louisiana State University Museum of Natural Science, and the Marjorie Barrick Museum of Natural History. We thank S. A. Jansa and F. J. Cuthbert for comments on the manuscript. This study was supported in part by NSF DEB-0316092 to FKB and SML and by the University of Minnesota.

2-1. INTRODUCTION

Mitochondrial genomes (mitogenomes) are an attractive source of data for molecular phylogenetic studies of animal taxa. Because of their rapid time to coalescence, relatively high substitution rates, and large size (~17,000 bp), mitogenomes are more likely than other loci to evolve in concert with, and harbor evidence of, the population histories of species (Moore 1995). Moreover, their high copy number, haploidy, and lack of recombination make mitogenomes especially easy to obtain, sequence, and analyze (Avise 1998, Berlin et al. 2004). Given their merits, we contend that mitochondrial DNA (mtDNA) sequences should be included as one marker among many (Fisher-Reid and Wiens 2011) in coalescent-based "species tree" and other multilocus analyses, rather than being abandoned for use in phylogeny construction, as some have advocated (e.g. Ballard and Whitlock 2004, Galtier et al. 2009; reviewed by Rubinoff and Holland 2005). Even as technological advances reduce the cost and difficulty of sequencing large numbers of nuclear loci, so should there be a concomitant increase in the use of mitogenomes, as they too are more readily acquired, whether intentionally or as by-products of genomic sequencing (e.g. Nabholz et al. 2010). Consequently, we argue that the routine practice of utilizing only 1–2 kilobases of mtDNA sequence in phylogenetic analyses should be replaced by the use of whole mitogenomes so as to take full advantage of the potential resolving power of the locus, especially with groups of closely-related organisms in which genetic distances are small. Although mitogenomic data have great potential, standards for their rigorous and objective use in phylogenetic analyses are currently lacking.

Of particular relevance to developing best methods for phylogenetic analyses of mitogenomes is that they exhibit heterogeneity in base composition and evolutionary rates at various scales across the molecule (Anderson et al. 1982, Cummings et al. 1995), which suggests that such analyses should benefit from data partitioning (Yang 1996, Nylander et al. 2004). Partitioning improves model fit by dividing alignments into relatively homogeneous sets of sites before selecting and optimizing a substitution model for each set independently. Nevertheless, data partitioning is not widely used with mitogenomes. To survey current practice, we examined 71 papers with phylogenies (Appendix 1-A), published in association with recent submissions of metazoan whole mitogenome sequences to Genbank, and found that only about a third employed a partitioning scheme. To further review practices of the researchers most likely to employ exemplary methods, we reviewed an additional 40 papers (Appendix 1-B), most published in the past 5 years, which we selected for their focus on recovering phylogeny (rather than describing novel mitogenomes). We found that while 78% used some sort of partitioning—still a remarkably low proportion in our view—there was little uniformity of approach. Generally, protein-coding sites were sorted by gene and/or by codon position, and RNA sites by type, gene, and/or transcript structure (e.g. stems vs. loops), but other criteria (e.g. template strand, evolutionary rate) were used in some cases. Of greater concern, discussion of partitioning options, consideration of their evaluation, or references to model studies were generally lacking. Consequently, we found much unexplained variation among partitioned analyses; for example, the number of data subsets utilized ranged from two to 42, with a mode of five groups.

One pervasive feature of mitogenomic studies is exclusion of data. Although some researchers have argued against that practice (Cameron et al. 2007, Kjer and Honeycutt 2007), of the 111 studies that we reviewed, under 10% made use of all alignable sequence positions. Most studies did not justify data omission, but those that did gave reasons of intragenomic heterogeneity, substitutional saturation, and unreliability of the signal from certain portions of the mitogenome—all problems that can be ameliorated with data partitioning (Cameron et al. 2007, Kjer and Honeycutt 2007). We were surprised to find no clear relationship between use of data partitioning and data exclusion: over half the studies that used full alignments used no partitioning, and many partitioned analyses discarded data, especially noncoding regions and RNA genes. In vertebrate studies, even those employing partitioning, it was common practice to exclude the ND6 gene (the only L-strand protein template) because of its markedly different base composition from other protein-coding genes.

In our view, partitioning should be favored over data exclusion as a strategy for dealing with heterogeneity within the mitogenome. However, the systematics community lacks a general strategy for selecting partitioning schemes for such datasets. Therefore, our primary goal for this study was to develop a procedure for partitioning mitogenomes that would make use of all alignable positions, accommodate among-site heterogeneity in base composition and evolutionary rates, and avoid overparameterization. In fulfilling this objective, we found inspiration in a method proposed by Li et al. (2008) for partitioning datasets composed of multiple nuclear protein-coding genes. The first steps in this approach include finely partitioning the data according to *a priori* categories, estimating evolutionary model parameter values for each subset, and grouping subsets with a clustering algorithm. The resulting clustering hierarchy is then used to define a nested set of alternative partitioning schemes, which are evaluated using tools derived from information theory. We believe that the procedure presented here, though more involved than current practices in mitogenomic phylogenetics, offers more explicit criteria for selecting a partitioning scheme and is efficient and flexible enough to serve as a model for future studies.

Another objective of this paper was to test the utility of whole mitogenome datasets for avian phylogenetics, particularly for resolving relationships among species within families that have undergone rapid diversification. Use of mitogenomes had a troubled early history in ornithology. The first studies, which examined interordinal relationships (Härlid et al. 1999, Mindell et al. 1999), generated results that were so at odds with other evidence and traditional views that they garnered considerable skepticism, both toward their findings and the general value of mitogenomic data (e.g. Johnson 2001). Whereas mitogenomes are often used within other vertebrate classes (e.g. Teleostei), they have remained unpopular in avian phylogenetics, even though the spurious results of early studies were later explained as artifacts of inadequate evolutionary models and taxon sampling (Braun and Kimball 2002, Slack et al. 2007). Moreover, avian mitogenomic studies have focused on higher-level relationships (e.g. Paton et al. 2002, Harrison et al. 2004, Gibb et al. 2007, Morgan-Richards et al. 2008, Pratt et al. 2009, Pacheco et al. 2011) even though such data likely have more promise for resolving recent divergences, which present fewer issues with signal saturation and inter-taxon base compositional heterogeneity. A new era of avian

mitogenomic phylogenetics may now be emerging as evidenced by the very recent publication of the first species-level mitogenomic phylogenies of select clades—15 crane species (Krajewski et al. 2010), 19 Hawaiian honeycreepers and 28 related species (Lerner et al. 2011), and 9 swallows, (Cerasale et al. 2012). Incredibly, despite the fact that the order Passeriformes accounts for over half of extant bird species diversity, the mitogenomes of only ten passerines had been published prior to the last two studies. Our study adds to this roster by inferring relationships within a passerine subfamily, the grackles and allies clade of New World blackbirds (Icteridae), using newly sequenced mitogenomes of 23 species.

2-2. METHODS

2-2.1. Taxon sampling.—Most ingroup sampling (Table 1) was designed to infer relationships within a clade of New World blackbirds (Icteridae) that is endemic to South America ("group 1" of Johnson & Lanyon 1999). We included one individual from 16 of 19 recognized species, thereby representing 12 of 13 genera (Gill & Donsker 2011, Remsen et al. 2010). Missing were *Curaeus forbesi*, *Macroagelaius subalaris*, and the monotypic *Hypopyrrhus pyrohypogaster*. To further examine generic relationships within the grackles and allies—the subfamily to which the South American clade belongs (Lanyon & Omland 1999)—we included one representative of each of the remaining six genera in that taxon. Recent molecular analyses of nine-primaried oscine phylogeny (Barker et al. *submitted*) found New World orioles sister to the grackles and allies, so we used *Icterus mesomelas* as an outgroup.

2-2.2. Laboratory procedures.—We obtained *Agelaius phoeniceus* and *Molothrus aeneus* specimens as purified mitochondrial DNA extracts, prepared as described by Lansman et al. (1981; samples provided by J. C. Avise & D. Walker). For all other species, we extracted genomic DNA from frozen tissue samples using a DNeasy Tissue Kit (Qiagen, Valencia, CA) following manufacturer's instructions. We aimed first to amplify the complete mitochondrial genome of *Agelaius phoeniceus* in 11 overlapping fragments, each ~2000 bp in length, and then sequence the products with 21 complementary pairs of overlapping ~1000 bp reads. Some primers were unreliable or unsuccessful, and so were modified to achieve a better match. The resulting primer set was used for amplification and sequencing of the remaining blackbird taxa, but obtaining some fragments from some species required additional primers (Table 2-S1).

All reactions were performed in 12.5-μL aqueous solution with working concentrations of reagents as follows: 0.4 μM for each primer, 0.4 μM dNTP, 3 mM $MgCl_2$, 1× Green GoTaq Flexi Buffer (Promega Corporation, Madison, WI), 0.025 units/μL GoTaq Hot Start DNA polymerase, and, when we suspected problems due to secondary structure formation in some rRNA-coding fragments, 1M glycine betaine. Cycling parameters for initial PCR were usually as follows: denaturation at 95°C for 2 min; 30 or 35 cycles of 95°C for 30 s, 55°C for 30 s, and 72°C for 2 min; final extension at 72°C for 3 min. Sometimes, we used an annealing temperature of 58 or 60°C to increase primer specificity. For other

problematic amplifications, we used a touch-down protocol, substituting 5 cycles each at annealing temperatures of 58°C, 56°C, and 54°C, followed by 30 cycles at 52°C. In a few cases of non-specific priming, we ran PCR products in an agarose gel with tris-acetate-EDTA (TAE) buffer, excised the band corresponding to the desired product, and reamplified it prior to sequencing. We purified PCR products with enzymatic digestion following Werle et al. (1994) using exonuclease 1 and shrimp alkaline phosphatase (USB Corporation, Cleveland, Ohio), and sequenced them on an ABI 3700 automated sequencer (BigDye v3.1, Applied Biosystems, Foster City, California) at the Biomedical Genomics Center of the University of Minnesota.

2-2.3. Sequence and alignment editing.—We used Sequencher 4.7 (Genecodes, Ann Arbor, Michigan) to align and edit chromatograms of complementary reads and overlapping fragments to produce consensus sequences for each species. Following import into Geneious Pro 4.8.4 (Drummond et al. 2010), the consensus sequences were aligned with one another and *Taeniopygia guttata* (Genbank accession DQ422742.1; Mossman et al. 2006) using MUSCLE (Edgar 2004) with default settings and 8 iterations. We found no differences in gene ordering, and little local length variation among species, so we made minor corrections to the alignment by hand, using stop codons as landmarks, then annotated sequences for further use and Genbank submission using *Taeniopygia* as our reference for feature identification.

To prepare the aligned sequences for phylogenetic analyses, we removed *Taeniopygia*, realigned the sequences using MUSCLE, and made minor edits by hand to produce a consensus alignment totaling 16,862 positions. We tried using Aliscore (Misof and Misof 2009) and Gblocks (Castresana 2000) to automate identification of alignment ambiguous portions, but neither program was as stringent as our own judgment, so we excluded many additional positions (but did not exclude all 1–3 bp indels, as recommended by Gblocks). Excluded sections were located in 12S and 16S rRNA, tRNA-Lys, several intergenic spacers, and the control region, and primarily consisted of areas with length variation around poly-C or highly variable stretches. In all, we excluded 185 positions, yielding a final alignment of 16,677 positions for analysis.

2-2.4. RNA structure determination.—In order to classify bases in RNA genes as corresponding to paired versus unpaired positions (e.g. in helices versus loops) in their transcripts, we inferred the secondary structures of those molecules for *Agelaius phoeniceus*, then coded each position in the multitaxon consensus alignment accordingly. We used the Quikfold application on the DINAMelt server (Markham and Zuker 2005, 2008) and hand-fit comparisons to *Gallus* (Desjardins and Morais 1990) to infer tRNA structures of all species. To determine base pairing in rRNA gene transcripts, we fit the *Agelaius* sequence to structural models of 12S for *Falco peregrinus* (Mindell et al. 1997) and *Harpactes ardens* (Cannone et al. 2002), and, for lack of an avian model (but note Raposo do Amaral et al. 2010, published subsequent to our efforts), 16S models for *Xenopus laevis* (Cannone et al. 2002) and *Bos taurus* (Burk et al. 2002, Mears et al. 2006). Some discrete sections of *Agelaius* rRNA sequence were too divergent to allow for easy match to the models, so we used Quikfold to infer their structures.

2-2.5. Exploration of data heterogeneity.—We conducted several analyses to discover overall patterns of variation within our dataset. We used χ^2 tests to check for significant among-taxon variation in base composition (Gruber et al. 2007) at variable alignment positions. We also examined whether base composition and other evolutionary model parameter estimates differed among portions of the alignment for a sliding window of 500 bp. Parameter values for the HKY+G model (Hasegawa et al. 1985) were estimated using PAUP* 4.0b10 (Swofford 2002) for each sample of alignment positions on the best maximum likelihood tree for the unpartitioned dataset. Furthermore, we examined how evolutionary model parameters differed among subsets of sites grouped according to their function, strand location, gene identity and other categories described in our methods for data partitioning.

2-2.6. Data partitioning.—To start, we divided the sequence data into 48 subsets by sorting alignment positions according to all possible combinations of the following categories: noncoding/coding, heavy/light template strand, protein/RNA-coding, gene identity (done for rRNA and protein-coding genes only), codon position, and paired/unpaired bases in RNA secondary structure. Where alignment positions were shared by overlapping genes, we gave priority to protein-coding over RNA-coding, to paired over unpaired, and to gene order for protein-coding genes; we categorized stop codons with frameshifts as noncoding. For those and subsequent analyses, all positions were described according to the bases on the L-strand.

To allow direct comparison of the evolutionary dynamics of the data subsets to one another, we estimated model parameter values for each subset under the TrN+G model (Timura and Nei 1993) in PAUP* on the best maximum likelihood tree for the unpartitioned dataset. We selected the TrN+G model after attempting to use the GTR+G model (Tavaré 1986); however, because of the close relationships between taxa in our study, some substitution classes (especially G-T and A-C transversions) were not observed within many data subsets, resulting in spurious parameter estimation under that more complex model. Using the TrN+G model yielded realistic parameter estimates for all but four subsets—three (ND6 codon 1^{st} positions, ATPase8 codon 1^{st} positions, ATPase8 codon 2^{nd} positions) that were deficient in some transitions, and one (tRNA L-strand paired positions) that had no transversions. We set those subsets aside during the next few steps of our procedure. The estimate of gamma shape for ND4 codon 3^{rd} positions was infinity, and the A-G transition rate of ND3 codon 2^{nd} positions was zero, so we substituted, respectively, the next highest and lowest estimates observed amongst the other subsets.

Next, we $\log_e$-transformed each parameter estimate, then standardized and ordinated them by principal component analysis (PCA) in R 2.10.1 (R Development Core Team 2010). We used the PCA axis scores as input for a clustering analysis of the data subsets based upon linkage by average pairwise Euclidean distance (UPGMA). Guided by the hierarchy of the resulting clustering dendrogram, we created a nested set of 44 partitioning schemes that ranged, by clustering the two most similar subsets in each step, from maximally partitioned (i.e. equivalent to the 44 data subsets used in the clustering analysis), to one subset less than maximally partitioned, and so forth through to the unpartitioned dataset. For each partitioning scheme, we estimated parameter values of the TrN+G model in PAML 4.4 (Yang 2007) on the

best maximum likelihood tree for the unpartitioned dataset, then applied the Bayesian information criterion (BIC; Schwarz 1978) to score the fits to the model and thereby decide which was best. R scripts implementing our clustering evaluation are available on request from the authors.

Having chosen an optimal partitioning scheme under a uniform model parameterization, we identified the best substitution model for each data cluster using the BIC as implemented in jModelTest 0.1.1 (Posada 2008). To re-incorporate the four data subsets that we had set aside, we took each subset in turn and calculated its likelihood in PAUP* on the best maximum likelihood tree for the unpartitioned dataset, using parameter values estimated for the best model for each data cluster, then pooled the subset with whichever cluster it best fit as judged by the BIC. Prior to phylogenetic analysis, jModelTest was again used to select the best substitution model for each of the amended clusters.

To directly compare the performance of our best scheme to some alternatives from the literature, we reanalyzed our dataset using our methods and those of other mitogenomic studies after following their practice of first excluding all noncoding sequences and the ND6 gene. These analyses were performed in PAML using the TrN+G model for all data subsets, and results were scored with the BIC.

2-2.7. Phylogeny inference.—We inferred phylogenetic relationships among the sampled taxa under maximum parsimony, maximum likelihood (ML), and Bayesian criteria, using both the unpartitioned and the partitioned datasets. Parsimony analysis was executed in PAUP* using a heuristic search with 10,000 addition sequence replicates and tree bisection and reconnection (TBR) branch swapping. Nodal support was assessed, after excluding uninformative characters, with 500 nonparametric bootstrap replicates using TBR and 10 addition-sequence replicates each. Heuristic searches for the ML tree were conducted in GARLI 1.0 (Zwickl 2006) using the unpartitioned dataset, and in GARLI-PART 0.97 (Zwickl 2006) with partitioned data, using 50 random starting points; nodal support was similarly evaluated with 500 bootstraps, each starting from a random starting point. To infer phylogenies with Bayesian methods, we used MrBayes-3.1.2 (Huelsenbeck and Ronquist 2001, Ronquist and Huelsenbeck 2003) with Metropolis coupling (four chains with default heating), sampling every 100 generations. Tracer 1.4.1 (Rambaut and Drummond 2007) and the AWTY server (Wilgenbusch et al. 2004) were used to check that effective sample sizes for parameter estimation in these analyses were adequate (i.e. >200) and that estimates of nodal posterior probability had converged. We ran the unpartitioned analysis for three million generations with a burn-in of 1000 samples, and the partitioned analysis for 12 million generations with a 10,000 sample burn-in. To allow comparison of the inferential power obtained using whole mitogenomes to that from a dataset of conventional size and composition, we analyzed the combined sequences of cytochrome *b* and NADH dehydrogenase subunit 2 (ND2) using the same methods as for the unpartitioned ML analysis of the whole mitogenomes.

2-3. RESULTS

Our results are organized sequentially around three overall topics. In section 2-3.1 and Appendix 2, we describe the mitogenomes of the grackles and allies with respect to their organization, composition,

and variation. These details, little reported for birds (especially passerines), are relevant to demonstrating the need for data partitioning, identifying and applying the categories used to subset the data in the initial steps of our partitioning procedure, and to consideration of whether transformations (e.g. RY-coding) are needed. In section 2-3.2, we report on variation among data subsets, the relationship of that variation to the clustering hierarchy used to define partitioning options, and our discoveries as to which categories were most relevant to such variability, and thus, most useful for data partitioning. In section 2-3.3, we present the phylogeny inferred from our dataset with particular attention to effects of partitioning and dataset size on topology and nodal support.

2-3.1.1 Sequencing and mitogenome organization.—Fragment amplification and sequencing was generally unproblematic. However, in six instances (particular fragments from particular taxa), we obtained likely pseudogenous sequences which did not match overlapping portions of adjoining fragments or other taxa; further effort, using different primers, yielded apparently correct products. The complete mitochondrial genomes ranged from 16,757 to 16,782 bp in length (Table 1); their consensus alignment totaled 16,862 positions. Gene order (Table 2-S2) was the same as *Gallus* (Desjardins and Morais 1990) and most other bird groups (Mindell et al. 1998). Protein-coding gene lengths, and the start and end sequences of RNA genes, were virtually identical to those of *Taeniopygia*, making identification of their boundaries unproblematic (see Appendix A2-1 for additional details).

2-3.1.2. RNA structures.—The transcribed tRNA sequences of all species folded into typical cloverleaf-shaped secondary structures with plausible acceptor stems and with anticodon loops that conform to the vertebrate mitochondrial genetic code. We found 12S rRNA in *Agelaius phoeniceus* to be structurally similar to that of *Falco* (Mindell et al. 1997). Although most base substitutions occurred in loops, helices had many as well. Even in places where the sequences were very different, they rarely differed in length, and even then, only in loops and usually only by 1–2 bp. Nearly all of the differences seen in stems involved both bases of a putative pair, or switches between C and T paired with G, reflecting a history of compensatory substitutions that conserved structural features. The *Agelaius* helices (Nos. 8, 23, 24, 28, 47; see Mindell et al. 1997) and adjacent loops that were compositionally different enough to require analysis with Quikfold were found to be qualitatively similar in structure to their *Falco* homologs. All sequence length differences among *Agelaius* and the other blackbirds in our study (indels of 1–3 bp) were located in loops. For the most part, 16S rRNA from *Agelaius* was structurally similar to the *Xenopus* and *Bos* models; however, two portions—part of domain III, and the area between helices 42 and 44—were very divergent among the three taxa. We found the fit of *Agelaius* to the models in these areas (including two different proposals for domain III in *Bos*; Mears et al. 2006) to be ambiguous or undeterminable; these alignment positions (1810–1899, 2336–2348, 2371–2381) were pooled with unpaired positions in analyses.

2-3.1.3. Heterogeneity among taxa and across genomes.—Base composition of the whole mitogenomes (averaged across taxa: 32% A, 33% C, 13% G, 22% T) matched known avian and general vertebrate patterns (Broughton et al. 2001), including GC content of ~46%, and a deficit of G and T on the L-strand. Although base composition differed among taxa, and we found *Dives* to be significantly different

from the among-taxon average (see Appendix A2-2 for further details), the magnitudes of these heterogeneities were small and appeared not to affect inferred tree topology since they were uncorrelated with the phylogenetic patterns that we found.

Our sliding window analyses of base composition and evolutionary model parameter estimates revealed much regional variation across the alignment (Fig. 2-S1). We expected to find spatial patterning in that variation resulting from strand-asymmetric replication processes, as reported for other vertebrates, especially mammals (Faith and Pollock 2003, Krishnan et al. 2004, Gibson et al. 2005, Raina et al. 2005, Broughton and Reneau 2006), but we found no linear trends, such as compositional gradients, across comparable sites (e.g. codon 3rd positions). Such trends may not exist in these taxa; no origin of L-strand replication has been identified in birds (Desjardins and Morais 1990), and replication may initiate at many sites across the avian mitochondrial genome (Reyes et al. 2005; for further discussion, see Appendix A2-2)

We found substantial differences among sites grouped by functional type, which, to the extent that such types are not randomly distributed at a fine scale across the genome, might explain some patterns of regional variation. For example, protein genes generally have higher C and lower G content on the L-strand, higher substitution rates, and moderate among-site rate heterogeneity (higher α) in comparison to RNA genes (Fig. 2-S1). The only L-strand template protein gene, ND6, has high A and C, and low G and T content (41% A, 39% C, 8% G, 11% T) in comparison to other parts of the genome. The non-coding control region exhibited very low transition bias. However, many of the biggest differences exist between sets of sites that are spatially intermixed (Fig. 1). In comparison to the overall genome, third positions of H-strand template proteins have very high A, high C, low G, and low T content, and a high substitution rate (44% A, 39% C, 6% G, 11% T, tree length 1.89), whereas second positions are notable for their low A content, very high T content, and low substitution rate (19% A, 29% C, 13% G, 39% T, tree length 0.14). Equally notable are the differences between paired and unpaired RNA sites; paired sites have low A and high GC content and a low substitution rate (24% A, 26% C, 26% G, 24% T, tree length 0.19), whereas unpaired sites have high A and low GC content and evolve more quickly (44% A, 23% C, 10% G, 23% T, tree length 0.52).

2-3.2. Data partitioning and model selection.—Principal component analysis (PCA) of parameters from 44 data subsets revealed that the first three axes accounted for 96% of variation in model parameters (Table 2). The first axis, which was strongly positively correlated with T frequency and strongly negatively correlated with all other parameters, accounted for 83% of total variation. The second axis, primarily related to CT transition rate versus AG transition rate and C frequency, accounted for 8% of the variance. The third axis accounted for 5% of total variance and contrasts CT transition rate with A frequency.

All data subsets composed of codon 3rd positions grouped together separately from other subsets along the first PCA axis (Fig. 1), and they were collectively distinguished from other subsets at the highest level of the clustering hierarchy (Fig. 2). Correspondingly, partitioning the data into two clusters—one composed of codon 3rd positions, and the other of everything else—yielded a far greater improvement in likelihood and BIC score than did any additional partitioning; nevertheless, the optimal scheme along the

hierarchy, under BIC, was to partition into nine clusters (Fig. 3). The second partitioning step separated most codon 2nd positions from other sites and yielded the second largest improvement in model score. The order of remaining steps to the nine-cluster scheme, as dictated by the clustering hierarchy, was unrelated to their value for score improvement. All gains were quite small, but the largest subsequent improvements came from separating paired and unpaired RNA sites, codon 1st positions, and noncoding sequences from one another. The nine clusters of the optimal scheme were quite homogeneous with respect to codon position and RNA base-pairing (Table 3, Fig. 2). We also found that the data subsets that we set aside prior to the clustering analysis fit best into clusters that they matched with respect to those characteristics (Table 3).

Reanalysis of our data, after first excluding the ND6 gene and noncoding positions, returned a nine-cluster partitioning scheme, similar to that for the complete dataset. The BIC score of that nine-cluster scheme was superior to all but one of the alternatives from other mitogenomic studies that we tested on the trimmed version of our dataset (Table 4). Partitioning by only codon position and RNA pairing, for a total of five data groups, allowed for a better fit than our scheme (ΔBIC = 17). Its performance further improved (ΔBIC = 104) after we applied partition-specific models, rather than the TrN+G model, within each scheme. In comparison, approaches that partitioned by protein gene, or by gene × codon, rather than by codon alone, were much inferior (ΔBIC ≈ 8495 and ≈ 1421 respectively), as were those that partitioned by rRNA gene and tRNA rather than by secondary structure pairing (ΔBIC ≈ 520).

2-3.3. Phylogeny inference.—Using the full dataset, tree topology (Fig. 4) differed very little according to analytical approach. Parsimony and likelihood analyses yielded single best trees, with all nodes dichotomously resolved. No conflicts were found among the strongly supported nodes of different analyses. Support was generally lacking for nodes at the base of the tree, and consequently, for the pattern of relationships among the genera *Agelaius*, *Molothrus*, and *Nesopsar*. On the other hand, all analyses found strong support for a clade comprising the South American endemic genera, and for a sister relationship between that clade and one composed of *Dives*, *Euphagus*, and *Quiscalus*. In general, nodal support values were lower in analyses of the partitioned dataset, though the greatest differences were at poorly supported nodes (Fig. 4).

Whole mitogenomes yielded a better resolved and more strongly supported tree in our comparison of unpartitioned ML analyses than that generated using a dataset of conventional size, consisting of cytochrome *b* and ND2 gene sequences (Fig. 5). In particular, with one exception, all nodes receiving ≥ 50% bootstrap support in analyses of cytochrome *b* and ND2 attained bootstrap values of >95% in analyses of the full data set, whereas nodes with <50% support variously increased or decreased in support. With the mitogenomic dataset, we resolved all 20 nodes of a majority-rule consensus tree with bootstrap support >50%, and 17 of those with ≥70% support, whereas the conventional dataset resolved only 13 and 11 nodes at those thresholds. Of the 17 well-supported nodes in the mitogenomic analysis, 11 were equivalent to the well-supported nodes found with the conventional dataset, three were recovered in the latter with 35–54% support, and three were not found at all in that analysis (in single best trees) due to differences in tree

topology. Of the three nodes with <70% support using the mitogenomic analysis, two were not found in the conventional analysis due to differences in topology. Of the nine nodes with <70% support using the conventional dataset, five were not found in the mitogenomic analysis due to differences in topology.

2-4. DISCUSSION

In this study, we present a method for partitioning mitogenomic datasets to optimize model-fitting during phylogenetic analyses, and we demonstrate the utility of whole mitogenome sequences for inferring better-resolved and more strongly-supported hypotheses of the phylogenetic history of that locus than possible using an exemplary 2-kilobase sample. In examining the mitogenomes of a clade of New World blackbirds (Icteridae), we found a great deal of variation in base composition and substitution rates among functional categories of sites. Such heterogeneity is typical of vertebrates and other animals, and thus calls for routinely partitioning mitogenomic datasets in phylogenetic analyses, yet only a minority of recent studies do so, and they utilize a wide array of (largely unjustified) approaches. We suggest several standards for best practice.

2-4.1. Mitogenomic dataset partitioning.—Our procedure for partitioning mitogenomes derives from a method first proposed by Li et al. (2008) for multiple protein-coding nuclear genes. The overall approach involves dividing the data into subsets based on *a priori* categories, then generating a tractable set of alternatives for combining those subsets into homogenous clusters, and finally, selecting the best from among that set of schemes. Though designed to yield an efficient and effective solution, like any heuristic method, this procedure does not promise discovery of *the* optimal partitioning scheme for a given dataset, even from amongst the universe of possible combinations of the predefined data subsets; in fact, for a modified version of our dataset that lacked ND6 and noncoding positions, our procedure failed to match or better a previously-published (though not widely utilized) scheme with five data groups (Table 4; Harrison et al. 2004, Phillips et al. 2010).

The component methods of our procedure likely impose a number of limitations on its ability to accomplish globally optimal solutions. First, as with all partitioning methods, the effectiveness of the final scheme is limited by how well the categories used to define data subsets capture the variation in evolutionary tendencies that exists among individual sites. Second, the parameter values used to describe data subsets are point estimates that may be imprecise, especially for subsets with few variable sites or for which the model applied (in our study, TrN+G) is a nonoptimal parameterization. Third, clustering is critical to making our procedure practical by reducing the dimensions of subset variation to a single set of summary distance measures, but doing so is a drastic simplification and some particulars of the outcome may be sensitive to the algorithm employed. Also, strict adherence to the structure of the clustering hierarchy entailed designating several small groupings of data subsets as clusters in our final partitioning scheme. These clusters had few variable sites (clusters 7–9; Table 3), so were nearly devoid of phylogenetic signal. Consequently, they had negligible effect on likelihood but added many parameters to the final scheme and thus compromised BIC score performance. We did not take the trouble to do so

(because it would not have improved phylogeny inference), but a step could be added to our procedure to combining each such "empty" cluster with whichever larger cluster it had the best likelihood fit.

Despite its limitations, the approach advocated here should reliably yield near-optimal partitioning schemes—as it did for our dataset—because it is designed to maximize within-cluster data homogeneity, which we expect is the strongest determinant of performance. Although apportioning data into homogeneous groups is ostensibly the goal all partitioning efforts, most studies that we reviewed either employed a single approach without explaining its merits, or they tested a very limited set of alternatives, often erring on the side of the most complex scheme with its better likelihood score. Some strategies were apparently products of faulty reasoning, for example those that grouped genes by their name (i.e. all ND versus COX versus ATPases; e.g. Okajima and Kumazawa 2009, Podnar et al. 2009), or data subsets according to their best-fit model parameterization (i.e. HKY versus GTR etc; e.g. Mulcahy and Macey 2009) rather than similarity of parameter values.

Perhaps the best feature of the partitioning method presented here is that it does not presuppose which of the categories used to initially define data subsets are most strongly correlated with variation in evolutionary processes. Rather, it makes those determinations through explicit testing. For example, a notable feature of the partitioning scheme for our dataset is that the data clusters are quite homogeneous with respect to codon position and RNA secondary-structure pairing. In effect, we created a close approximation of the five data group scheme that bettered our own (Table 4), so our results provide explicit empirical justification (heretofore lacking) for that simple scheme and suggest that it deserves wider use (It was employed in only four of the 111 studies that we reviewed—Harrison et al. 2004; Gibb et al. 2007; Phillips et al. 2006, 2010), perhaps after adding a sixth category for noncoding sequences. Moreover, our analysis allows us to reject, at least for our dataset, a number of popular partitioning categories. Many studies partition proteins by gene, but we found that data subsets did not cluster by gene (Fig. 2), that partitioning proteins by gene performed very poorly (Table 4), and that partitioning proteins by gene in addition to codon position added tremendous complexity with comparatively little improvement in likelihood. Likewise, partitioning rRNA by gene, or RNA into rRNA and tRNA, brought little benefit. We also found that template strand was a fairly unimportant dimension for both RNA and protein-coding genes, though many studies presume its relevance when choosing to exclude the ND6 gene. Finally, our analysis indicates that evolutionary rate by itself is not a good basis for clustering; for example, noncoding sequences, codon 3rd positions, and unpaired RNA positions, though all rapidly evolving, did not group together in our analysis of overall similarity.

We do not yet know the extent to which the results of our procedure as applied to our dataset may be similar to those for datasets from other animal taxa or with greater taxonomic scope. We suspect, given that the final partitions correspond to conserved organizational attributes of vertebrate mitogenomes, that broad commonalities will be found. Some features may be universal. Our findings that codon positions clustered together, and that partitioning 3rd positions from all other sites provided the largest improvement in model fit, were identical to the results of Li et al. (2008), even though they used a very different

dataset—ten nuclear genes from a broad taxonomic sample of ray-finned fishes (Actinopterygii). Nevertheless, because our method is flexible and adaptable, it does not require other datasets to behave like ours. Furthermore, should other datasets require additional processing prior to phylogenetic analysis, those treatments might be incorporated into our procedure. For example, studies with broader taxonomic scope sometimes confront effects of saturation and significant differences in base composition among taxa, either of which can undermine phylogeny estimation if not addressed. We note that remedial strategies such as RY-coding (Phillips and Penny 2003, Gibson et al. 2005) could be applied to partitions after using our methods to define them.

2-4.2. Best practices in phylogenetic analyses of mitogenomic data.—Partitioning benefits phylogenetic analyses of mitogenomes in two interrelated ways. First, by improving model-fit for such demonstrably heterogeneous data sets, evolutionary dynamics that might bias unpartitioned analyses are accommodated. Second, partitioning may obviate the unfortunate practice of selectively purging or transforming data to reduce heterogeneity, and thereby maximize dataset size and utilization of phylogenetic signal. Our analysis of blackbird mitogenomic data indicates that, at a minimum, the best partitioning schemes recognize codon position, base-pairing in RNA, and non-coding regions. Although each of these categories has been employed in previous studies, to our knowledge our analysis provides the most explicit justification for their importance relative to other alternatives (e.g. coding strand, gene identity, evolutionary rate, location along molecule, or complex idiosyncratic combinations thereof) and demonstration of their combined superiority over other schemes. We note that coding positions within RNA genes according to base-pairing in their transcripts is a tedious process when done by carefully hand-fitting them to structural models. An efficient alternative is to make protein and RNA structural annotations to an exemplar genome, align multiple genomes, and then map the coding and structural assignments of the exemplar to the others. That procedure works quite well for closely related species, such as the set of blackbirds sampled here, but for more distantly-related taxa, substantial editing of the alignment may be necessary (Kjer 1995). To streamline this process for analyses of birds (especially closely-related passerines), we provide (available from authors upon request) a gene and structural annotation of the *Agelaius phoeniceus* mitogenome.

We found it almost universal practice to exclude non-coding regions and the ND6 gene. Most studies also leave out all RNA sequences to avoid their differing base compositions and evolutionary rates in comparison to the H-strand proteins. Likewise, it is common practice to translate protein-coding genes into amino acids, sometimes simply to avoid differences in evolutionary characteristics among codon positions. In taxa where significant rearrangements and duplications around the control region raise questions of orthology (e.g. Abbott et al. 2005), or when significant base compositional heterogeneity exists among taxa (e.g. Gibson et al. 2005), data exclusion or transformation may be justified; otherwise, we suggest using all alignable positions with appropriate partitioning to account for their distinctiveness.

Generally, inclusion of more data—provided their evolution is adequately modeled—should foster less biased and more accurate inferences of phylogeny. That was certainly the case for our study when the

results of analyses using our exhaustive whole mitogenome (>16 kb) dataset are compared to those based on our cytochrome *b* with ND2 (2-kb) dataset, which has often been considered an exemplary sample of the mitochondrial locus. Six of 17 nodes that were resolved with strong support using the full dataset did not have strong support using the smaller one, and three were not recovered by its single best topology or bootstrap consensus (with even plurality support). On the other hand, although we expect their inferences of phylogeny to be more reliable, there is no reason to expect that better-fitting models will yield higher nodal support, which is sometimes erroneously used as an optimality criterion for model selection or misinterpreted as a measure of inference accuracy. In fact, in our partitioned analyses, support values were generally slightly lower than in unpartitioned analyses, though strongly-supported nodes were little affected. This result suggests that model underparameterization led to overconfidence in nodal support values.

2-4.3. Conclusions.—Our review of current practices for phylogenetic analysis of the mitochondrial genome revealed that most studies employ unsophisticated approaches that do not adequately account for its internal heterogeneities or fully exploit the locus. No explicitly recognized or *de facto* standards for dataset partitioning exist among otherwise exemplary studies. We present a method (adapted from Li et al. 2008) for developing efficient, effective, and empirically justified partitioning schemes for such datasets, and we suggest that certain categories, such as codon position and RNA secondary-structure base-pairing, may be more salient than others for dataset partitioning. Greater use of species-tree methods in phylogenetics will reduce interest in partitioning methods that achieve efficiency by grouping sites from different genetic loci (as in Li et al. 2008), but the utility of partitioning within the mitochondrial locus will remain. As mitogenomic datasets become commonplace, we encourage use of these methods to take full advantage of the historical signal that they contain.

Table 2-1. Specimens sequenced for inferring phylogenetic relationships within the grackles and allies subfamily of New World blackbirds (Icteridae).

Taxon	English name[a]	Specimen[b]	Collecting locality	Mitogenome length (bp)	GenBank number
Icterus mesomelas	Yellow-tailed Oriole	LSUMZ 109279	Panama: Darién	—[c]	JX516068
Agelaius phoeniceus	Red-winged Blackbird	BB-96 Tordoff	USA: Minnesota	16775	JX516062
Nesopsar nigerrimus	Jamaican Blackbird	FMNH 331150	Jamaica: Portland	16780	JX516054
Molothrus aeneus	Bronzed Cowbird	BB-73 James	Mexico: Puebla	16757	JX516067
Dives dives	Melodious Blackbird	MBM 7100	Honduras: Copán	16766	JX516061
Euphagus cyanocephalus	Brewer's Blackbird	FMNH 342000	USA: California	16765	JX516072
Quiscalus quiscula	Common Grackle	FMNH 341733	USA: Illinois	16768	JX516064
Lampropsar tanagrinus	Velvet-fronted Grackle	ANSP 177921 LSUMZ B-103505	Peru: Loreto	—[d]	JX516057
Macroagelaius imthurni	Golden-tufted Mountain Grackle	FMNH 339783	Venezuela: Bolívar	16768	JX516073
Gymnomystax mexicanus	Oriole Blackbird	FMNH 339743	Venezuela: Falcón	16781	JX516075
Amblyramphus holosericeus	Scarlet-headed Blackbird	FMNH 334662	Bolivia: El Beni	16768	JX516063
Curaeus curaeus	Austral Blackbird	AMNH 826156	Chile: Magallanes	16771	JX516070
Gnorimopsar chopi	Chopi Blackbird	FMNH 334679	Bolivia: Santa Cruz	16775	JX516055
Agelasticus thilius	Yellow-winged Blackbird	FMNH 334615	Bolivia: Oruro	16771	JX516069
Agelasticus xanthophthalmus	Pale-eyed Blackbird	FMNH 324094	Peru: Madre de Dios	16769	JX516059
Agelasticus cyanopus	Unicolored Blackbird	FMNH 334636	Bolivia: El Beni	16771	JX516076
Agelaioides badius	Baywing	FMNH 330801	Brazil: Rio Grande do Sul	16774	JX516074
Oreopsar bolivianus	Bolivian Blackbird	FMNH 334687	Bolivia: El Beni	16777	JX516058
Chrysomus ruficapillus	Chestnut-capped Blackbird	FMNH 330775	Brazil: Rio Grande do Sul	16773	JX516056
Chrysomus icterocephalus	Yellow-hooded Blackbird	FMNH 339772	Venezuela: Sucre	16773	JX516060
Xanthopsar flavus	Saffron-cowled Blackbird	FMNH 330747	Brazil: Rio Grande do Sul	16773	JX516065
Pseudoleistes guirahuro	Yellow-rumped Marshbird	FMNH 330795	Brazil: Rio Grande do Sul	16770	JX516071
Pseudoleistes virescens	Brown-and-yellow Marshbird	FMNH 330796	Brazil: Rio Grande do Sul	16769	JX516066

[a] Gill and Donsker 2011

[b] Specimens with "BB" codes are unvouchered and were obtained from J. C. Avise and D. Walker; all other codes are museum catalog numbers. When tissue and skin specimens are housed at different institutions, both are listed. Abbreviations: ANSP = Academy of Natural Sciences of Drexel University; AMNH = American Museum of Natural History; FMNH = Field Museum of Natural History; LSUMZ = Louisiana State University Museum of Natural Science; MBM = Marjorie Barrick Museum of Natural History.

[c] Incompletely sequenced; missing two portions totaling ~460 bp, one from the spacer following $tRNA^{thr}$ through the first part of ND6 (consensus alignment positions 14973–15273), and the other in the control region (positions 16311–16470).

[d] Incompletely sequenced; missing a ~515 bp section from the last part of ND4 through most of $tRNA^{ser}$ (consensus alignment positions 11299–11813).

Table 2-2. Principle components and their correlations with the model parameter estimates (with TrN+G parameterization) from which they were derived in analyses of the compositional and evolutionary properties of 44 portions of the mitochondrial genomes of species in the grackles and allies subfamily of New World blackbirds (Icteridae). Abbreviations: π_i = frequency of base *i*, r_{ij} = substitution rate between bases *i* and *j*.

	Principal component axes						
	1	2	3	4	5	6	7
Eigenvalue square roots	2.4870	0.7760	0.6080	0.3621	0.3013	0.1821	0.0784
Proportion of variance	0.8340	0.0812	0.0498	0.0177	0.0122	0.0045	0.0008
Cumulative proportion of variance	0.8340	0.9150	0.9648	0.9825	0.9947	0.9992	1.0000
Component coefficients:							
π_A	−0.813770	0.299650	−0.450160	0.142352	−0.028560	0.121057	0.097980
π_C	−0.764200	−0.445010	0.251137	−0.347060	−0.157160	0.074407	0.064825
π_T	0.910879	−0.170940	0.149272	−0.015220	0.317220	0.106373	0.081380
r_{AG}	−0.661620	−0.570380	0.123597	0.470139	0.023347	−0.009050	−0.000480
r_{CT}	−0.537920	0.551383	0.623377	0.129778	0.001530	0.029464	0.017451
Gamma distribution shape (α)	−0.944340	0.031770	−0.050950	−0.133650	0.228047	−0.172480	0.070767
Tree length	−0.952540	−0.009140	−0.061530	−0.125060	0.210668	0.099245	−0.137580

Table 2-3. Characteristics of data sets and subsets used in phylogenetic analyses of mitogenomic data from the grackles and allies subfamily of New World Blackbirds (Icteridae).

Dataset	Number of positions			Model[a]	Parameter values for ML analysis (r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, r_{GT}), (π_A, π_C, π_G, π_T), p_{iv}, α	Data subset membership
	Total	Variable	Parsimony informative			
Cluster 1	3794	2950	2057	TIM3+I+G[b]	(0.43, 28.98, 1, 0.43, 17.93, 1), (0.426, 0.423, 0.042, 0.109), 0.029, 2.105	Codon 3rd positions (all)
Cluster 2	2773	503	320	GTR+I+G	(0.54, 8.98, 0.67, 0.00, 15.41, 1), (0.300, 0.308, 0.213, 0.180), 0.697, 0.741	Codon 1st positions of COX1–2 and ND1–6
Cluster 3	2451	138	80	TIM3+I+G[b]	(4.55, 67.93, 1, 4.55, 15.65, 1), (0.187, 0.299, 0.114, 0.400), 0.865, 0.431	Codon 2nd positions of ATP6, COX2–3, ND1–2, and ND4–5
Cluster 4	2362	166	90	K80+I+G[c]	(1, 35.25, 1, 1, 35.25, 1), (0.25, 0.25, 0.25, 0.25), 0.873, 0.715	RNA paired positions; codon 1st positions of COX3
Cluster 5	2144	340	211	TrN+I+G[b]	(1, 10.00, 1, 1, 14.71, 1), (0.440, 0.232, 0.119, 0.209), 0.689, 0.539	RNA unpaired positions; codon 2nd positions of ND6
Cluster 6	1983	467	282	HKY+I+G	(1, 6.69, 1, 1, 6.69, 1), (0.292, 0.310, 0.141, 0.257), 0.606, 0.572	Noncoding positions; codon 1st positions of ATP6, ATP8, ND4L, and Cyt *b*
Cluster 7	517	10	3	HKY+I	(1, 6.42, 1, 1, 6.42, 1), (0.180, 0.274, 0.149, 0.397), 0.961, NA	Codon 2nd positions of COX1
Cluster 8	381	9	3	HKY	(1, 22.04, 1, 1, 22.04, 1), (0.206, 0.256, 0.125, 0.413), NA, NA	Codon 2nd positions of Cyt *b*
Cluster 9	272	25	13	HKY+I[d]	(1, 16.31, 1, 1, 16.31, 1), (0.154, 0.331, 0.134, 0.381), 0.890, NA	Codon 2nd positions of ATP8, ND3, and ND4L
Unpartitioned mitogenome	16677	4608	3059	TVM+I+G[b]	(1.42, 16.32, 0.98, 0.26, 16.32, 1), (0.316, 0.329, 0.135, 0.220), 0.612, 0.900	
Cytochrome *b* and ND2	2184	754	509	TPM2uf+I+G[b]	(2.62, 33.58, 2.62, 1, 33.58, 1), (0.307, 0.367, 0.114, 0.212), 0.573, 1.416	

[a] Abbreviations as used in jModelTest 0.1.1 (Posada 2008).

[b] GTR+I+G model implemented in MrBayes.
[c] HKY+I+G model implemented in MrBayes.
[d] This was the 2nd-best model under the BIC (ΔBIC = 1.4057, −lnL of 577.3 versus 580.8) in jModelTest; the "best" model, TPM3uf+I, was inappropriately complicated and led to spurious parameter estimates.

Table 2-4. Comparison of model fit for alternative partitioning schemes used in phylogenetic analyses of mitochondrial genomes. Scores are based on analysis of mitogenomic sequences (excluding noncoding positions and the ND6 gene) from the grackles and allies subfamily of New World blackbirds (Icteridae), using the TrN+G substitution model. Bolded values in columns indicate the preferred partitioning schemes under the specified optimality criteria.

Partitioning scheme	No. Data groups	No. Parameters	−lnL	BIC	ΔBIC	AIC	ΔAIC
Codon position, RNA	4	70	65100	130873	437	130340	828
Codon position, rRNA, tRNA	5	77	65077	130894	458	130308	796
Codon position, RNA paired, RNA unpaired	5	77	64848	**130436**	0	129850	338
Result from this study	9	105	64722	130453	17	129654	142
Protein gene, RNA paired, RNA unpaired	14	140	68793	138931	8495	137865	8353
Protein gene, 12S, 16S, tRNA	15	147	69012	139437	9001	138317	8805
Codon position by protein gene, RNA paired, RNA unpaired	38	308	**64448**	131857	1421	**129512**	0
Codon position by protein gene, 12S, 16S, tRNA	39	315	64682	132393	1957	129994	482

Table 2-S1. Primers used for amplifying and sequencing mitochondrial genomes of species in the grackles and allies subfamily of New World blackbirds (Icteridae). Position numbers refer to the L-strand sequence of *Agelaius phoeniceus*, starting with the first position of $tRNA^{phe}$. References in parentheses indicate primer sequences that were altered from the original published versions.

	L-strand primers				
Mitogenome fragment	Lab label	Primer sequence	Positions 3'–5'	Gene at 3' end	Reference[a]
Primer pairs used for most PCR:					
1	ND5emb2	GGYCTAAYCAAAGCCTAYCTA	13620–13600	ND5	Barker 2004
2	L15848	CCAAACTACGATCAATRACYTTCCG	14632–14608	Cyt *b*	(Groth 2000)
3	LCR3e	TCCAACAGCCTTCAAGAACA	15988–15969	CR	(Tarr 1995)
4	L1753	AAACTGGGATTAGATACCCCACTAT	527–503	12S	Sorenson et al. 1999
5	L3450_Agpho	GAAGACCCTGTGGAACTTTAA	2200–2180	16S	Lee et al. 1997
6	L5793	ATCCTAGCCTTCTCCTCCATCTC	4554–4532	ND2	
7	L7612_Agpho	CCAACTGGAATCAAAGTATTCAGCTG	6370–6345	COX1	
8	L8740	GGCCATTTCCGCCTACTAGAAGT	7506–7484	COX2	Lee et al. 1997
9	L10167_Agpho	GTACAAAAAGGATTACGATACGG	8945–8923	COX3	
10	L12369-mod	GCTTACTCATCCGTMAGCCAYATAGG	11150–11125	ND4	(Sorenson et al. 1999)
11	L12976	CAAGAACTGCTAATTCTCGCATCTG	11757–11733	tRNA-Ser	Sorenson et al. 1999
Primers used for most sequencing (used in conjunction with above PCR primers):					
1b	L15410	TGAGGCGGATTCTCYGTMGACAA	14194–14172	Cyt *b*	(Groth 2000)
2b	polyC_for	CCCCCCCCAGTACATTT	15619–15603	CR	Barker 2004
3b	F304_emb	CTTGGCACTGATGCACTTTG	16406–16387	CR	(Baker & Marshall 1997)
3bb	polyT_for	TTTTTATTTTTTTTTATCAAACAATAAAACC	16577–16547	CR	Barker 2004
4b	L2724_Agpho	ATCGAGCTGGGTGATAGCTG	1479–1460	16S	(Sorenson et al. 1999)
5b	L4500_alt	GTAGCMCAAACAATCTCMTATGAAG	3250–3226	ND1	(Sorenson et al. 1999)
5bb	L3827_alt	GCAATCCAGGTCGGTTTCYATC	2576–2555	16S	(Sorenson et al. 1999)
6b	L6958_Agpho	AACAACATAAGCTTCTGACT	5713–5694	COX1	(Sorenson et al. 1999)
7b	L8232_Agpho	YTGGTTTCAAGCCAACCGC	6990–6972	tRNA-Ser	(Sorenson et al. 1999)
8b	L9233_Agpho	CCTGACCATGAACCTAAGCTTCTT	8027–8004	ATP6	(Sorenson et al. 1999)
9b	L11122_Agpho	CAAGGCGGACTAGAATGAGCAGA	9900–9878	ND3	(Sorenson et al. 1999)

11b	L14080_Agpho	TCAACCCAYGCMTTCTTTAAAGC	12863–12841	ND5	(Sorenson et al. 1999)
Other PCR primers:					
5	F5_gap_for	GGGTCACTATGATAAAGTGAACATGG	3831–3806	tRNA-Ile	
6	L5793_alt	ATCYTAGCYTTTTCCTCCATTTC	4554–4532	ND2	
Other sequencing primers:					
2a	polyC_for2	CCCCCCCCAGTGCGTTT	15619–15603	CR	(Barker 2004)
2 before polyC	ND6_for	TGCAACCGCCGCYACAAC	15400–15383	ND6	
3 polyT	F304_emb (see above)				
4	F4_polyC_for	CCCCCCCCCATAAACTAATA	975–956	12S	
5b_end	L5216_Agpho	GCCCATACCCCGAAAATG	3976–3959	tRNA-Met	(Sorenson et al. 1999)
11a	L12976_Ictmes	CGAGAACTGCTAACTCTTGTATCTG	11757–11733	tRNA-Ser	(Sorenson et al. 1999)

Mitogenome fragment	**H-strand primers** Lab label	Primer sequence (actual, 5'–3')	Positions 3'–5'	Gene at 3' end	Reference[a]
Primer pairs used for most PCR:					
1	H16065	TTCATCTCCGGTTTACAAGAC	14848–14869	tRNA-Thr	Helm-Bychowski & Cracraft 1993
2	BBCR_Rev1	CCAAGTGTAGGAGGTCGGTAT	16223–16243	CR	
3	H1861-12sd	ATCGATTACAGAACAGGCTCCTC	635–657	12S	Raxworthy et al. 2008
4	H3754 (trunc)	GTTAATCGTTGAACAAACGAACC	2503–2525	16S	
5	H6030_Agpho	TTAGTTAGTTCTTGGATGATGAGTCA	4790–4815	ND2	
6	H7662	AGGAAGATGAATCCTAGGGCTCA	6420–6442	COX1	Lee et al. 1997
7	H9036	TYTAGCTTAAAAGGCTARCGC	7802–7822	tRNA-Lys	(Sorenson et al. 1999)
8	H10441	TATTCTATGGCTTGGAGGGCAGT	9220–9242	COX3	
9	H12488_Agpho	ATTCGACTGTGGGTTCGTTC	11269–11288	ND4	(Sorenson et al. 1999)
10	L13525-revcomp	GATATGATTCCTACYCCTTCTCAGC	12284–12308	ND5	(Sorenson et al. 1999)
11	H15103	TCAGCCGAATTGKACGTCTCGGCA	13887–13910	CytB	(Groth 2000)
Primers used for most sequencing (used in conjunction with above PCR primers):					
1a	H15709	GCGTAGGCGAATAGGAAGTATCA	14493–14515	CytB	Barker 2004
2a	polyC_rev	GGGGGGGGTGGAGTGA	15583–15598	CR	Barker 2004
3a	Emb1 (= H886)	AATATGTCCGGCAACCATTACA	16453–16474	CR	Barker 2004
3aa	polyT_rev	AAAAAAAAAAAATGATGCGTAAAA	16516–16539	CR	Barker 2004

4a	H2826_Agpho	TTCTTTTTTAAAGGAGCTGTACC	1573–1595	16S	(Sorenson et al. 1999)
5a	H4644_alt	TCGAATGGGGCRCGGTTTGTYTC	3394–3416	ND1	(Sorenson et al. 1999)
5aa	H5191	GGGGTATGGGCCCGATAGC	3951–3969	tRNA-Met	Sorenson et al. 1999
6a	H7032_Agpho	TTGCCTGCTAGTGGAGGGTA	5787–5806	COX1	(Sorenson et al. 1999)
7a	H8400_Agpho	ATTATTAGGGRGTGGTCGTG	7166–7185	COX2	(Sorenson et al. 1999)
8a	H9855_Agpho	ACGTAGGCYTGGATTATTGCTACTGC	8626–8651	ATP6	(Sorenson et al. 1999)
9a	H11660_Agpho	AGGGGAGAGGAGATTTGGTC	10441–10460	ND4	(Sorenson et al. 1999)
10a_rev	H13238_Agpho	GGTGTTTTTGAGGTTGTTGGA	11932–11952	ND5	
11a	H14127_alt	GTGTTCCTATTAGGGCTAGG	12984–13003	ND5	(Sorenson et al. 1999)
Other PCR primers:					
1	H16191-Pass	TCTCGWGGGGCGATTCGGGC	14985–15004	ND6	(Sorenson et al. 1999)
5	F5_gap_rev	GAGATTAATGGGAGGATGGC	4115–4134	ND2	
Other sequencing primers:					
2a	polyC_rev2	GGRGGATGATCTAGGCGTTC	15565–15584	tRNA-Glu	
2 before polyC	H417	AGTAGCTCGGTTCTCGTGAG	16027–16046	CR	Tarr 1995
3 polyT	H1248	CATCTTCAGTGTCATGCT	20–37	tRNA–Phe	Tarr 1995
4	F4_polyC_rev	GGGGGGKCGCCTGCGA	944–959	12S	
5a_endrev	H4169_Agpho	CTACTAATGTGAGGAAGGC	2872–2890	ND1	
5b_endrev	H5902_Agpho	GGCGTATAGGTAGAAGTTGAG	4607–4627	ND2	
11a	H14127_lctmes	GTGTTCCTATTAGGGCGAGA	12984–13003	ND5	(Sorenson et al. 1999)

[a] References:

Baker, A. J., H. D. Marshall. 1997. Mitochondrial control region sequences as tools for understanding evolution, pp. 51–82. *In* D. P. Mindell [ed.], Avian Molecular Evolution and Systematics. Academic Press, San Diego, CA.

Barker, F. K. 2004. Monophyly and relationships of wrens (Aves: Troglodytidae): a congruence analysis of heterogeneous mitochondrial and nuclear DNA sequence data. Mol. Phylogenet. Evol. 31:486–504.

Groth, J. G. 2000. Molecular evidence for the systematic position of Urocynchramus pylzowi. Auk 117:787–791.

Helm-Bychowski, K., and J. Cracraft. 1993. Recovering phylogenetic signal from DNA sequences: relationships within the corvine assemblage (class Aves) as inferred from complete sequences of the mitochondrial DNA cytochrome-b gene. Mol. Biol. Evol. 10:1196–1214.

Lee, K., J. Feinstein, and J. Cracraft. 1997. The phylogeny of ratite birds: resolving conflicts between molecular and morphological data sets, pp. 173–211. *In* D. P. Mindell [ed.], Avian Molecular Evolution and Systematics. Academic Press, San Diego, CA..

Raxworthy, C. J., R. G. Pearson, B. M. Zimkus, S. Reddy, A. J. Deo, R. A. Nussbaum, and C. M. Ingram. 2008. Continental speciation in the tropics: contrasting biogeographic patterns of divergence in the *Uroplatus* leaf-tailed gecko radiation of Madagascar. Journal of Zoology 275:423–440.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Mol. Phylogenet. Evol. 12:105–114.

Tarr, C. L. 1995. Primers for amplification and determination of mitochondrial control-region sequences in oscine passerines. Mol. Ecol. 4:527–529.

Table 2-S2. Mitochondrial genome organization in the grackles and allies subfamily of New World blackbirds (Icteridae). Underlined text identifies variants found in *Agelaius phoeniceus*.

Gene	Template strand	Consensus alignment positions	Start codon	Stop codon	*Agelaius phoeniceus* positions
$tRNA^{phe}$	H	1–69			1–68
12S rRNA	H	70–1051			69–1042
$tRNA^{val}$	H	1052–1121			1043–1112
16S rRNA	H	1122–2728			1113–2710
$tRNA^{leu}$	H	2729–2803			2711–2785
ND1	H	2823–3801[a]	ATG	AGA, AGG,TAG	2800–3777
$tRNA^{ile}$	H	3809–3880			3785–3856
$tRNA^{gln}$	L	3886–3955+3966[b]			3862–3932
$tRNA^{met}$	H	3966–4034			3932–4000
ND2	H	4035–5075	ATG	TAA	4001–5041
$tRNA^{trp}$	H	5075–5144			5041–5110
$tRNA^{ala}$	L	5146–5214			5112–5180
$tRNA^{asn}$	L	5225–5295			5190–5260
$tRNA^{cys}$	L	5297–5363			5262–5328
$tRNA^{tyr}$	L	5363–5433			5328–5398
COX1	H	5435–6985	ATG, GTG	AGG	5400–6950
$tRNA^{ser}$	L	6977–7049			6942–7014
$tRNA^{asp}$	H	7056–7124			7020–7088
COX2	H	7133–7819[c]	ATG	TAA, T—	7097–7780
$tRNA^{lys}$	H	7821–7891			7782–7851
ATP8	H	7893–8060	ATG	TAA, TAG	7853–8020
ATP6	H	8051–8734	ATG	TAA	8011–8694
COX3	H	8741–9524	ATG	T—	8701–9484
$tRNA^{gly}$	H	9525–9594			9485–9553
ND3	H	9595–9945	ATA, ATG	TAA, TAG	9554–9904
$tRNA^{arg}$	H	9947–10016			9906–9975
ND4L	H	10018–10314	ATG	TAA	9977–10273
ND4	H	10308–11685	ATG	T—	10267–11644
$tRNA^{his}$	H	11686–11755			11645–11714
$tRNA^{ser}$	H	11756–11822			11715–11780
$tRNA^{leu}$	H	11822–11892			11780–11850
ND5	H	11893–13710	ATG	AGA	11851–13668
Cyt *b*	H	13722–14864	ATG	TAA	13677–14819
$tRNA^{thr}$	H	14868–14937			14823–14891
$tRNA^{pro}$	L	14973–15043			14899–14968
ND6	L	15053–15571	ATG	TAA, TAG	14978–15496
$tRNA^{glu}$	L	15573–15643			15498–15568
Control region		15644–16862			15569–16775

[a] Some species have a 1 bp insertion that makes position 3801 part of the subsequent spacer.
[b] *Quiscalus quiscula* has a 10 bp insertion that completes the tRNA at position 3956 and creates a unique spacer in positions 3957–3965.
[c] *Pseudoleistes guirahuro* has an insertion that makes position 7820 the final base of COX2.

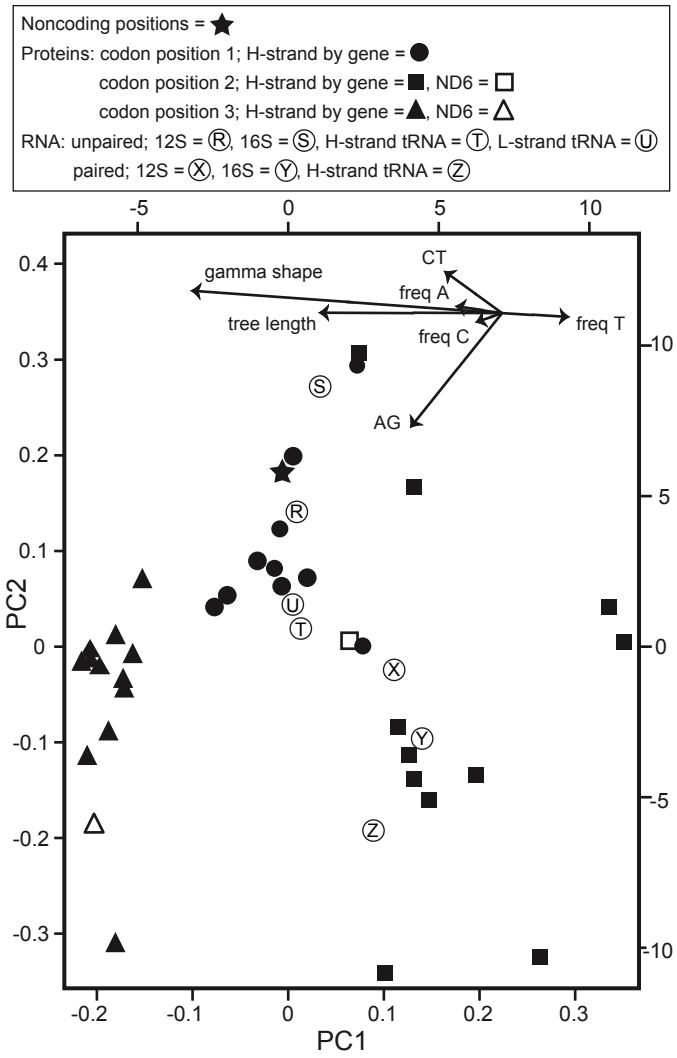

Figure 2-1. Principal component analysis (PCA) of data subsets from mitochondrial genome sequences of the grackles and allies clade of New World blackbirds (Icteridae). Shown are the first two axes of variation. The inset vector graph shows the correlation of partition-specific model parameter variation with these axes, using the TrN+G model with the unpartitioned phylogeny.

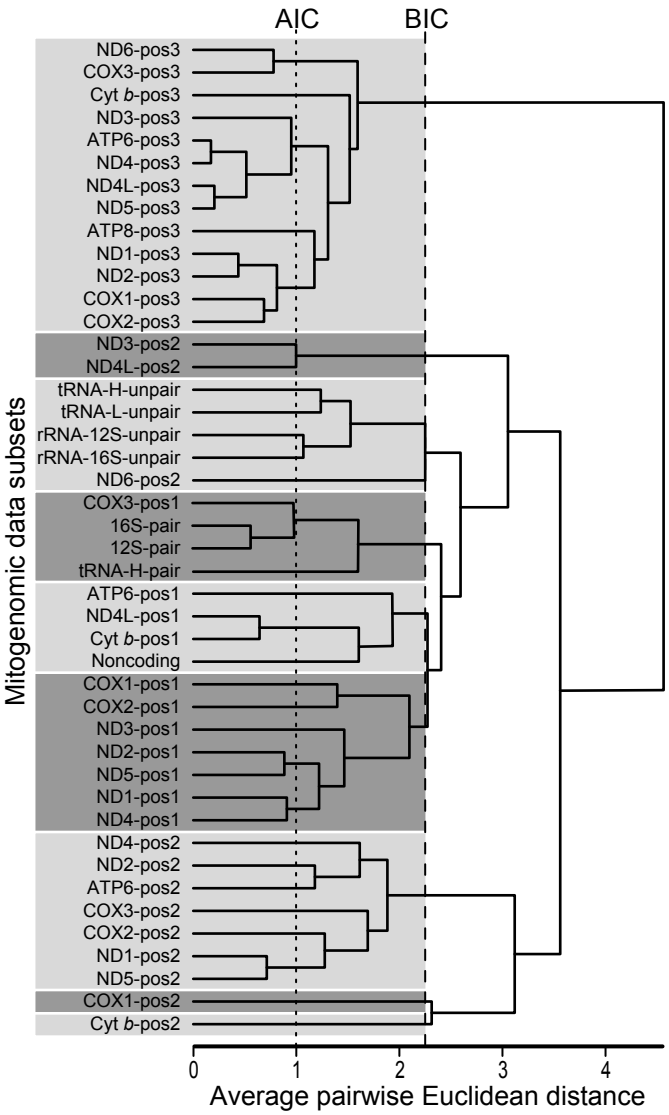

Figure 2-2. Clustering of mitogenomic data subsets from the grackles and allies subfamily of New World blackbirds (Icteridae), based on Euclidean distances in principal component space (Fig. 2-1). Shaded blocks denote the nine groups at the level of partitioning with the best BIC score.

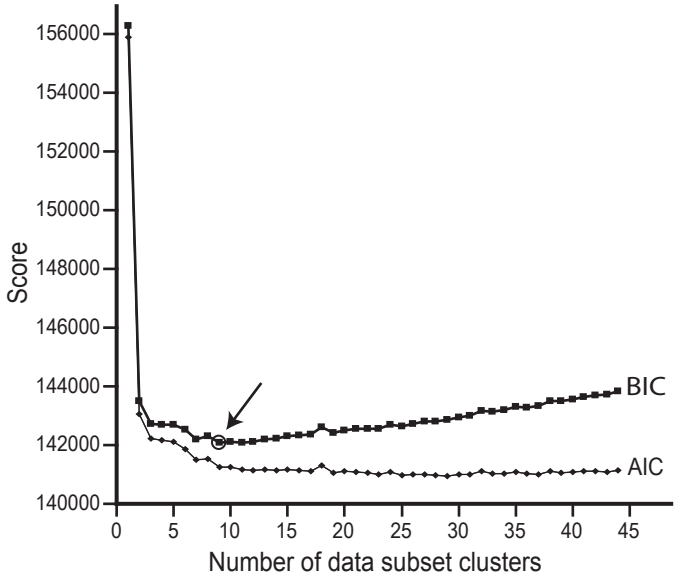

Figure 2-3. BIC and AIC scoring of model fit and efficiency of a series of nested data partitioning schemes for phylogenetic analysis of mitogenome sequences of the grackles and allies subfamily of New World blackbirds (Icteridae). The scheme with the best BIC score is circled and indicated with an arrow.

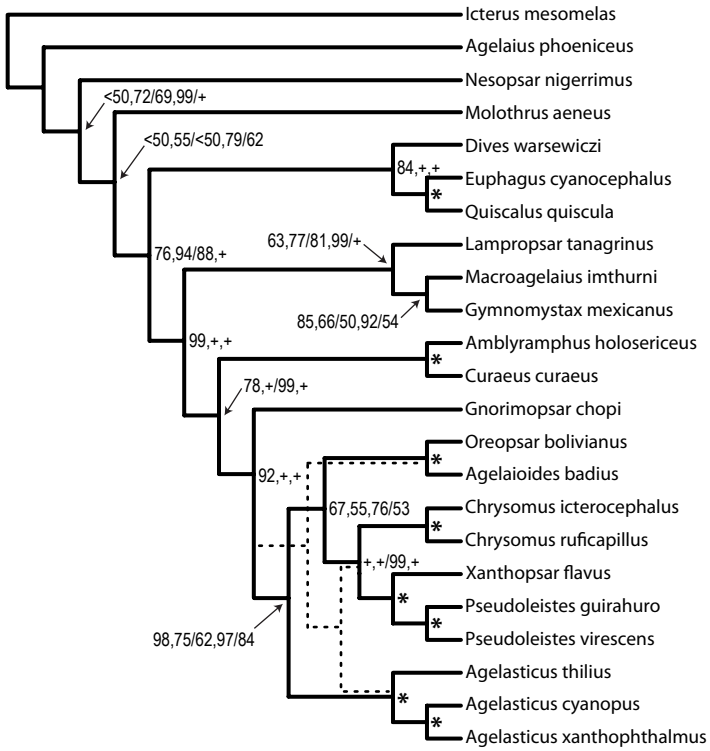

Figure 2-4. Phylogeny of the grackles and allies subfamily of New World blackbirds (Icteridae) inferred from whole mitochondrial genome sequences. Support values at nodes are nonparametric bootstrap percentages from parsimony and maximum likelihood (ML) analyses (unpartitioned/partitioned) followed by Bayesian posterior probabilities (×100; unpartitioned/partitioned). When identical, values from unpartitioned and partioned analyses are not reported separately. Support of 100% is indicated with a plus symbol; nodes that received 100% support in all analyses are marked with a single large asterisk. Dashed lines show the topology of the single best ML tree from analysis of the partitioned dataset where it differed from other ML and Bayesian analyses.

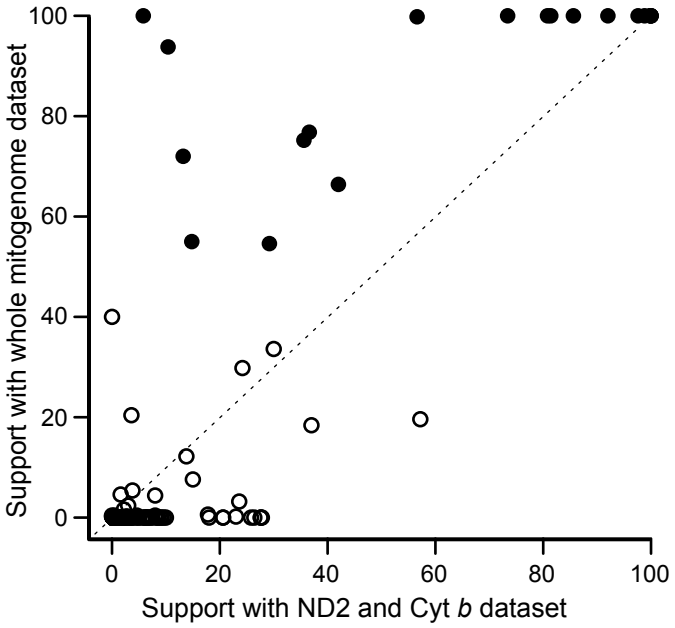

Figure 2-5. Comparison of support for bipartitions found in bootstrapped maximum-likelihood analyses of phylogeny within the grackles and allies clade of New World blackbirds (Icteridae) using two different datasets, (1) whole mitochondrial genome sequences, and (2) combined sequences of ND2 and cytochrome *b*. Bipartitions that occur in the single best tree using the whole mitogenome dataset are indicated with filled circles (and occur only in the upper half of the graph).

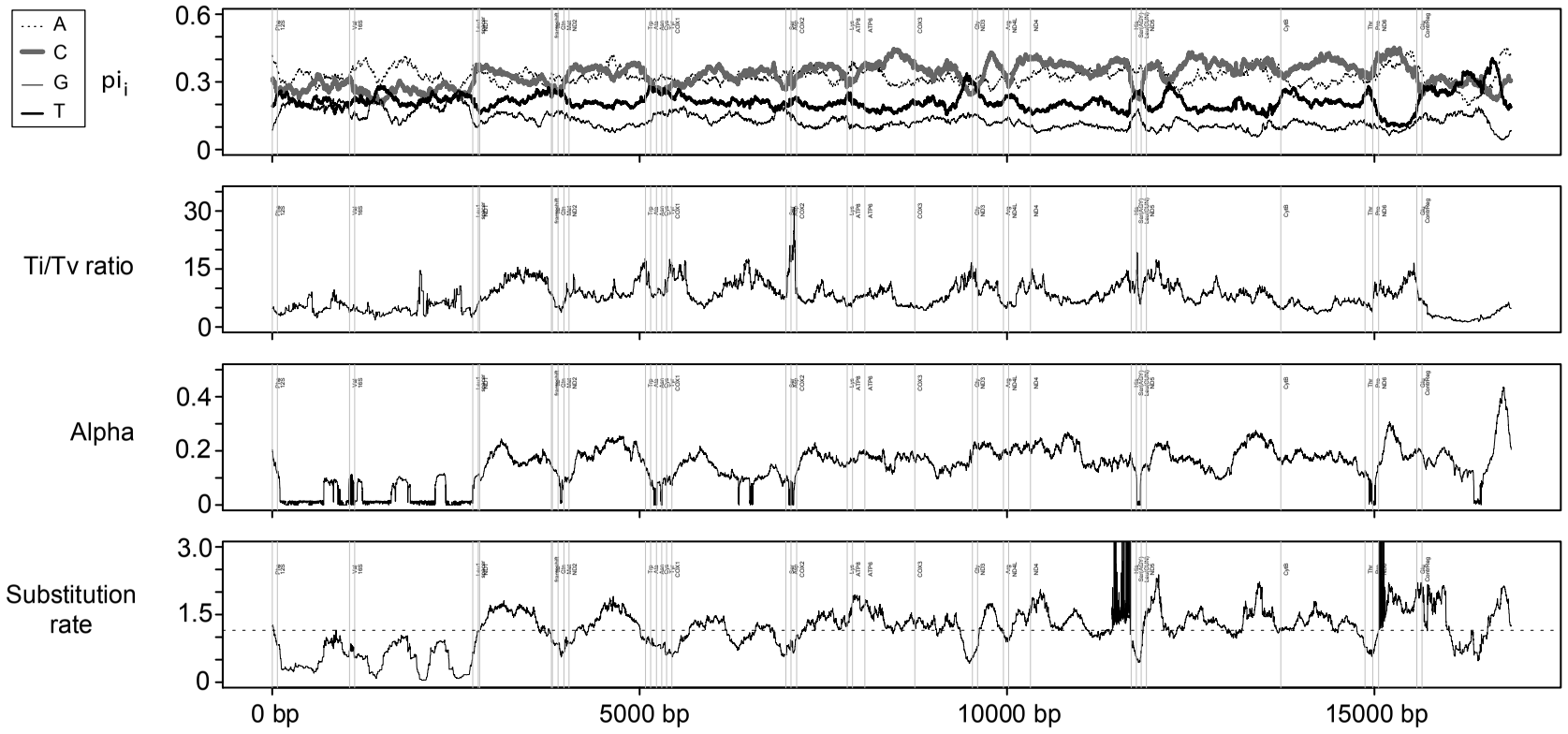

Figure 2-S1. Sliding window (500 bp) analysis of aligned whole mitochondrial genome sequences of species in the grackles and allies subfamily of New World blackbirds (Icteridae) using the TrN+G substitution model.

CHAPTER 3

A comprehensive species-level molecular phylogeny of the New World blackbirds (Icteridae)[1,2]

SUMMARY

The New World blackbirds (Icteridae) are among the best known songbirds, both through exemplar species, such as the Red-winged Blackbird (*Agelaius phoeniceus*), and collectively, through service as a model clade in numerous studies of morphological, ecological, and behavioral trait evolution. Knowledge of phylogeny is a prerequisite for reconstructing evolutionary patterns, and it is the basis for systematic classification, but as of yet there has been no analysis of blackbird phylogeny with comprehensive species-level sampling, or that offers robust support for most intergeneric relationships. Using mitochondrial gene sequences from all ~108 currently-recognized species and 7 additional distinct lineages, together with strategic sampling of 4 nuclear loci and whole mitochondrial genomes at the generic level, we were able to resolve most relationships with high confidence. Our best-resolved phylogeny is consistent with the strongly-supported results of past studies, but also contains many novel inferences of relationship, including unexpected placements of some newly-sequenced taxa, resolution of the relationships among the major subclades within Icteridae, and resolution of generic-level relationships within the largest of those subclades, the grackles and allies. We suggest taxonomic revisions based on our results.

3-1. INTRODUCTION

The New World blackbirds (Icteridae) are among the best known and studied songbirds, both through exemplar species, such as the Red-winged Blackbird (*Agelaius phoeniceus*), and collectively, through service as a model clade in numerous studies of morphological, ecological, and behavioral trait

---

[1] A version of this paper is being prepared for publication:

Powell, A. F. L. A., F. K. Barker, S. M. Lanyon, K. J. Burns, J. Klicka, and I. J. Lovette. A comprehensive species-level molecular phylogeny of the New World blackbirds (Icteridae), including multilocus and mitogenomic evaluation of generic relationships, with comments on taxonomy.

[2] Acknowledgements: We thank G. D. Weiblen for use of his lab during preparation of toe pad samples. H. Vázquez-Miranda helpfully performed some extractions. We thank the museums that loaned tissues for sequencing that was new to this project—the Academy of Natural Sciences of Drexel University; American Museum of Natural History; Field Museum of Natural History; Natural History Museum of Los Angeles County; Louisiana State University Museum of Natural Science; Museo Argentino de Ciencias Naturales, Bernardino Rivadavia; Marjorie Barrick Museum of Natural History; Museu Paraense Emílio Goeldi; University of Washington, and the Burke Museum of Natural History and Culture—as well as the many others (listed in Table 3-1) for maintaining the specimens and vouchers associated with the GenBank sequences that we used. We also thank J. C. Avise, D. Walker, J. M. DaCosta, K. E. Omland, and F. Jacobsen for providing extracts or sequences. S.A. Jansa provided helpful comments on the manuscript. This study was supported in part by NSF DEB-0316092 to FKB and SML and by the University of Minnesota.

evolution. The size of the group (~108 species) and its variability along several dimensions of general theoretical interest—both within and among the major subgroups it comprises—make the Icteridae especially attractive for comparative studies. Some topics that have been investigated in the family or its major subclades, using phylogenetically-informed methods, are as follows: mating systems (Searcy et al. 1999), brood parasitism (in cowbirds: Lanyon 1992), sexual size dimorphism (Webster 1992), sexual dichromatism (Irwin 1994; in orioles: Hofmann et al. 2008, Friedman et al. 2009), plumage pattern divergence (in orioles: Omland and Lanyon 2000; caciques and oropendolas: Price and Whalen 2009), chemical bases of plumage color (Friedman et al. 2011; in orioles: Hofmann et al. 2006, 2007, 2008), ultraviolet and structural color (in grackles and allies: Eaton 2006, Shawkey et al. 2006), ecological correlates of plumage color (in grackles and allies: Johnson and Lanyon 2000), ecological correlates of female song (Price 2009, Price et al. 2009), song divergence (in caciques and oropendolas: Price and Lanyon 2002b, 2004a; orioles: Price et al. 2007), migration (in orioles: Kondo and Omland 2007), biogeographic history (in orioles: Sturge et al. 2009), and ecological niche divergence (Eaton et al. 2008).

Knowledge of phylogeny is a prerequisite for reconstructing evolutionary patterns, and it is the basis for systematic classification, but as of yet there has been no formal analysis of blackbird phylogeny with comprehensive species-level sampling, nor does a tree exist that offers robust support for most intergeneric relationships. The first molecular phylogenies of Icteridae with broad taxonomic sampling provided revolutionary insights into relationships within the family (Lanyon 1994, Freeman and Zink 1995, Lanyon and Omland 1999) and within its subgroups (Johnson and Lanyon 1999; Omland et al. 1999; Price and Lanyon 2002a, 2004a; Barker et al. 2008), and were a huge advance over the diffuse hypotheses of relationship presented in taxonomic reviews, which were based on either informal evaluation of the external anatomy of museum skins (Ridgway 1902, Hellmayr 1937, Blake 1968) or very limited molecular sampling (Sibley and Monroe 1990). Sequence-based molecular studies, for the first time, brought together large numbers of informative characters with objective analytical methods to comprehensively resolve relationships among species and clades, tasks for which morphological characteristics had proven to be weakly informative (e.g. Björklund 1991) and, in combination with informal and speculative methods of inference, often (in hindsight) misleading (Beecher 1950, 1951).

In contrast to the other diverse families within the New World nine-primaried oscine clade—namely, the tanagers (Thraupidae), cardinal-grosbeaks (Cardinalidae), New World sparrows (Emberizidae), and New World warblers (Parulidae)—among which many species and genera have recently been shuffled, the constitution of the Icteridae has been unaffected by results from molecular phylogenetic studies of that radiation (e.g. Burns 1997, Klicka et al. 2000, Burns et al. 2002, Lovette and Birmingham 2002, Yuri and Mindell 2002, Burns et al. 2003, Klein et al. 2004, Klicka et al. 2007, Alström et al. 2008, Lovette et al. 2010, Barker et al. *in press*). Apparently, the features that have traditionally been used to recognize blackbirds, such as bill shape (casque of maxilla, see Webster 2003), morphology related to gape-feeding by many species (Beecher 1951, Orians 1985), and general similarities in shape, plumage, voice, display, and ecologies have led to their accurate diagnosis. The only contrary assertions have been placement of

Spiza in Icteridae (Beecher 1951, Raikow 1978), which was immediately disputed (see e.g. Tordoff 1954) and is not currently supported (molecular evidence places *Spiza* deep within the Cardinalidae; Klicka et al. 2007), and unsubstantiated yet prominently-presented suggestions of the affinity of *Compsothraupis loricata* to Icteridae (Jaramillo and Burke 1999) and lack thereof for *Amblycercus* (Fraga 2011). Comprehensive generic-level multi-locus molecular sampling of the nine-primaried oscines strongly supports the monophyly of Icteridae (Barker et al. *in press*) as traditionally defined.

Although molecular phylogenetic studies of Icteridae (e.g. Lanyon 1992, 1994; Freeman and Zink 1995; Lanyon and Omland 1999) did not lead to its redefinition, they shed considerable light on relationships within the family, including recognition of constituent clades and discovery that several genera—*Molothrus*, *Agelaius*, *Cacicus*, and *Psarocolius*—as then defined, were not monophyletic. Lanyon and Omland (1999) found that the Icteridae comprises 5 deeply-divergent lineages—the meadowlarks and allies (*Sturnella*, *Dolichonyx*, *Xanthocephalus*), cup-nesting caciques (*Amblycercus*), caciques and oropendolas (*Cacicus*, *Psarocolius*, *Ocyalus*), orioles (*Icterus*), and a large set of genera collectively referred to as the grackles and allies (e.g. *Agelaius*, *Quiscalus*, *Molothrus*)—but they were unable to resolve the basal divergences among those lineages. Similarly, Johnson and Lanyon (1999) examined the grackles and allies clade and found strong support for several groups, including the cowbirds (*Molothrus*), marsh blackbirds (*Agelaius*), and grackles (*Quiscalus*), among others, but poor support for the relationships among those lineages. Among the more surprising findings of these studies was existence of a clade of South American endemics ("group 1" of Johnson and Lanyon 1999) within the grackles and allies, composed largely of morphologically and ecologically enigmatic genera (many of them monotypic) together with species that had been thought to be members of genera (*Molothrus*, *Agelaius*) outside that clade. Subsequent studies have explored relationships *within* the basal icterid clades, especially the orioles (e.g. Omland et al. 1999, Jacobsen et al. 2010) and the caciques and oropendolas (Price and Lanyon 2002a, 2004a), but until now (this study, and Barker et al. *in press*) no additional effort has been made to resolve the relationships *among* the basal icterid clades, or among major groups within the grackles and allies (but see Powell et al. 2013), with additional sequence or taxon sampling.

By providing a set of highly-resolved hypotheses of the relationships among a large sample of species, molecular phylogenetic studies enabled substantial taxonomic revision of the Icteridae, and, for the first time, comparative investigations of the patterns generated by and processes involved in their diversification. The blackbirds hold much promise for additional taxonomic and comparative work, but all such work is ultimately limited by the resolution and accuracy of available phylogenetic hypotheses. Though many findings from the initial round of molecular investigations of the Icteridae had robust support, many did not. Inference of phylogeny is itself limited by taxon sampling, the sequences used, and analytical methods. Also, all past phylogenies of the Icteridae, except within the orioles (Allen and Omland 2003, Jacobsen et al. 2010, Jacobsen and Omland 2011) and some meadowlarks (Barker et al. 2008), have relied solely upon mitochondrial DNA. Given the passage of time, and improved capacity in all these areas, a revision of the phylogeny of the Icteridae as a whole, using new methods and data, is in order.

The overall goal of the present study was to infer, for the first time, a hypothesis of the phylogenetic relationships among all ~108 species of New World blackbirds (Icteridae), using both mitochondrial and nuclear DNA sequences. Key objectives were to (a) sample all currently-recognized species not included in previous studies, (b) robustly resolve relationships among major clades within Icteridae, (c) robustly resolve relationships among the grackles and allies, especially within a phenotypically and ecologically diverse clade of South American endemics, which previous studies have failed to resolve with confidence, (d) compare patterns of relationship found in previous studies, which used only mitochondrial markers, to results from nuclear loci, and (e) suggest taxonomic revisions based on our results. Preliminary results from this project (i.e. phylogenies inferred from less comprehensive versions of our dataset) have already informed studies of female song (Price 2009, Price et al. 2009) and plumage color evolution (Friedman et al. 2011), so we hope that the phylogeny presented here proves a useful reference and impetus for further work with the Icteridae.

3-2. METHODS

3-2.1 Taxon and character sampling.—Our analyses encompassed 115 ingroup and 4 outgroup taxa (Table 3-1). Sampling within Icteridae included all species currently recognized by taxonomic authorities (Dickinson 2003, Remsen et al. 2012, Gill and Donsker 2012) or in prominent references (Jaramillo and Burke 1999, Fraga 2011), with the following exceptions: we did not obtain samples of *Agelaioides badius fringillarius* or *Molothrus aeneus armenti* (Dugand and Eisenmann 1983), and we chose not to include (see Discussion) samples of *Psarocolius angustifrons alfredi* and *Agelaius phoeniceus gubernator*. Approximately 10% of the sampled taxa had not been included in previous molecular phylogenies of Icteridae, including 3 meadowlarks (*Sturnella militaris*, *S. loyca*, *S. defilippii*), 4 caciques and oropendolas (*Cacicus koepckeae*, *Psarocolius b. bifasciatus*, *P. cassini*, *P. guatimozinus*), an oriole (*Icterus jamacaii*), and 3 members of the grackles and allies subfamily (*Dives atroviolaceus*, *Curaeus forbesi*, *Macroagelaius subalaris*). We included more than one sample of a species if particular of its subspecies appeared to represent deeply divergent and geographically distinct lineages that likely merit specific status. Outgroups were selected based on results of recent molecular analyses of family and generic-level relationships within the New World nine-primaried oscines (Barker et al. *in press*), and consisted of *Teretistris fernandinae*, *Seiurus aurocapillus*, *Oreothlypis gutturalis* and *Icteria virens*.

We sequenced 4 nuclear loci (5266 bp total) from a set of 46 taxa (Table 3-2) that included at least one representative from 26 out of 28 ingroup genera (lacking only *Hypopyrrhus* and *Clypicterus*) and all 4 outgroups. From each of those taxa, we sequenced one protein-coding autosomal gene, two autosomal introns, and one sex-linked (Z chromosome) intron, those loci being, respectively, recombination activating gene 1 (RAG1), myoglobin intron 2 (MB-I2), β-fibrinogen intron 5 (FGB-I5), and aconitase 1 intron 9 (ACO1-I9). We also sequenced MB-I2 and ACO1-I9 from 4 additional taxa (including *Clypicterus*) and added ACO1-I9 or FGB-I5 sequences of another 4 taxa that were available on GenBank from past studies by ourselves and others (Table 3-2).

How comprehensively the mitochondrial locus was represented in our dataset varied considerably among taxa. We obtained sequences of the cytochrome *b* gene (1143 bp) for all 119 taxa in our study except *Icterus jamacaii*, which we included only as a COX1 sequence from GenBank. For 9 rarely collected or extinct species that were sampled using DNA extractions from toe-pads of museum skins, cytochrome *b* was the only gene that we sequenced, but for all other taxa, we also obtained ND2 gene sequences (1041 bp). Preliminary phylogenetic analyses, based on the concatenated nuclear, ND2, and cytochrome *b* sequences, yielded well-resolved trees with strong nodal support, with the notable exception of some generic-level relationships within the grackles-and-allies clade (Lanyon and Omland 1999) and among the 19 species that compose a clade of South American endemics ("group 1" of Johnson and Lanyon 1999) therein. The nuclear loci in our dataset were unable to resolve any but a few trivial relationships within the South American clade, so we turned to more extensive sampling of the mitochondrial genome as a source of phylogenetic signal. We sequenced whole mitochondrial genomes (~16775 bp) of 23 species (20 of them within our 46-taxon set; Table 3-2). For 5 other ingroup taxa in the 46-taxon set, and for the 4 outgroups, we obtained the sequence of a ~5000 bp fragment encompassing ND2, COX1, COX2, ATP8, ATP6, and several tRNA genes. Further, we filled in remaining missing sequence for each taxon to the extent possible using GenBank records, provided we could be confident (e.g. from locality information) of their taxonomic identities. Most additional mitochondrial sequences were from the COX1, ATP6, and 12S rRNA genes. Whenever possible, all nuclear and mitochondrial gene sequences were obtained from the same specimen, but for 45 taxa, we assembled chimaeric sequences from two or more specimens.

3-2.2 Laboratory procedures and sequence preparation.—Genomic DNA was extracted from frozen tissue and toe-pad samples as described in Powell et al. (2008) or with conventional phenol/chloroform methods (e.g. as in Lanyon 1994). To avoid contamination, we processed toe-pad specimens in a lab not otherwise used for extraction or amplification of avian DNA. Target DNA fragments were amplified via polymerase chain reaction (PCR). See listed references (given in parentheses) for details of primers and cycling parameters used to amplify the following sequences: cytochrome *b* and ND2 (Barker et al. 2008, Powell et al. 2008), whole mitogenomes and large fragments (Powell et al. 2013), RAG1 (Barker et al. 2002), MB-I2, FGB-I5, and ACO1-I9 (Barker et al. 2008, *in press*). Purification of PCR products, sequencing, sequence editing, and alignment were as described in Powell et al. (2013) except that some products were sent to Beckman Coulter Genomics (Danvers, MA) for sequencing.

3-2.3 Data partitioning and phylogeny inference.—To probe for potentially spurious effects of character and taxon sampling on phylogeny inference, we assembled the following datasets to analyze and compare: (1) concatenated (i.e. to analyze with standard phylogenetic inference) and (2) unconcatenated (i.e. to analyze with species tree methods) nuclear sequences of the 46 taxa for which all 4 such loci were sampled; (3) concatenated nuclear sequences of the 54 taxa with nuclear data; (4) cytochrome *b*, (5) combined ND2 and cytochrome *b*, and (6) full mitochondrial alignments of the 46-taxon and (7–9) 118 or 119-taxon sets. Based on results from those datasets, we assembled the following datasets for our final

analyses: (10) concatenated and (11) unconcatenated combined nuclear and full mitochondrial alignments of the 46-taxon sample, and (12) concatenated combined nuclear and full mitochondrial alignments of all 119 taxa. To maximize sequence coverage for *Molothrus* in the 46-taxon analyses of combined nuclear and mitochondrial loci, we utilized a chimaeric sequence, composed of the mitogenome of *aeneus* together with nuclear sequences from *ater*, since we lacked full sampling for either species.

All datasets were partitioned for analysis. Partitioning was accomplished by finely dividing each dataset according to *a priori* categories (such as gene and codon position), then using PartitionFinder 1.0.1 (Lanfear et al. 2012)—set to assess all models, using the greedy algorithm, under the Bayesian information criterion (BIC, Schwarz 1978)—to find an optimal scheme for grouping subsets according to similarities in evolutionary tendencies. The most complicated datasets were the full-length mitochondrial alignments. As described in Powell et al. (2013), alignment positions of those datasets were sorted into 48 initial subsets according to all possible combinations of the following categories: noncoding/coding, heavy/light template strand, protein/RNA-coding, gene identity (done for rRNA and protein-coding genes only), codon position, and paired/unpaired bases in RNA secondary structure. Initial subsetting of nuclear markers was limited to separation according to locus and, for RAG1, codon position. On occasion, PartitionFinder returned an inappropriately complicated model for a data subset, which led to spurious parameter estimates; to reassess those cases, and when we needed to identify best models for individual data blocks (as with the introns, for unconcatenated analyses), we used the BIC in jModelTest 2 (Darriba et al. 2012).

For single-locus and concatenated-loci datasets, we inferred phylogenetic relationships under maximum likelihood (ML) using GARLI 2.0 (Zwickl 2006) and with Bayesian methods using MrBayes-3.2.1 (Ronquist et al. 2012). We also used Bayesian methods as implemented in *BEAST 1.7.4 (Drummond et. al. 2012) to infer species trees from our unconcatenated multilocus 46-taxon datasets. Most GARLI analyses were run on the CIPRES Science Gateway (Miller et al. 2010), where we conducted heuristic searches for ML trees using 50 random starting points (i.e. searchreps), and evaluated nodal support with 500 bootstraps, each with a random starting point. Analyses with MrBayes used Metropolis coupling (four chains with default heating), and generally ran for 6–12 million generations, sampling every 100 generations, and with a burn-in of 10–25%. We found that default settings in MrBayes yielded unrealistically long tree-length estimates in partitioned analyses, so following Marshall (2010), we set a shorter prior on mean branch length (brlenspr = unconstrained:exp(100.0)). Analyses using *BEAST ran for 200 million generations, sampling every 10,000 generations, with a burn-in of 10%. For all partitions or loci in those analyses, we used a lognormal relaxed clock model of evolutionary rate, with an exponential prior (mean = 0.1). All mitochondrial partitions in *BEAST analyses were linked under the same tree model. We used Tracer 1.5 (Rambaut and Drummond 2009) and the AWTY server (Wilgenbusch et al. 2004) to check that effective sample sizes for parameter estimation in Bayesian analyses were adequate (i.e. >200) and that estimates of nodal posterior probability had converged.

Because the completeness of sequence and locus sampling varied substantially among taxa, we examined the results of the various datasets that we assembled (as previously described) to assess their

sensitivity to completeness of genetic marker and taxon sampling, and to check for congruence between inferences from nuclear loci and the mitochondrial genome. We looked for significant differences between analyses in their support for hypotheses of relationship, especially instances of strong conflict in pairwise comparisons (i.e. cases in which each of two incongruent hypotheses of relationship were supported by ≥ 70% of bootstrap replicates or ≥ 95% of posterior samples at incompatible nodes).

3-3. RESULTS

3-3.1 Partitions and substitution models.—Optimal partitioning and model parameterizations were achieved using relatively few data groups. The best partitioning scheme for cytochrome *b*, for both 46 and 118-taxon analyses, was into 3 groups by codon position. The 46 and 118-taxon ND2 plus cytochrome *b* datasets were partitioned into 4 data groups based on codon position and by gene for 3rd positions. As in Powell et al. (2013), we found that the most salient categories for mitogenomic partitioning were codon position, RNA secondary structure pairing, and the coding/noncoding distinction (Table 3-3). The best schemes for the 46 and 119-taxon analyses divided the mitochondrial data into 9 and 7 groups, respectively (Those schemes were quite similar, except that the 9 group scheme subdivided codon 1st positions and 3rd positions a bit more). Nuclear markers sorted separately from the mitochondrial data groups. Codon position was a significant variable within RAG1, but since some nuclear loci or subsets were similar enough that they grouped together (i.e. ACO1-I9 with FGB-I5, MB-I2 with 3rd positions of RAG1), the best schemes for the concatenated datasets utilized only 4 groups for nuclear data (Table 3-3). Parameter values estimated in the ML analysis of the 119-taxon combined mitochondrial and nuclear dataset are given in Table 3-3; they were very similar to the estimates obtained from our other analyses and datasets.

3-3.2 Phylogenies.—Analyses of the datasets that we assembled, with their various combinations of taxon and DNA sequence sampling (described in Methods), and using GARLI, MrBayes, and *BEAST, yielded a set of > 20 summary phylogenetic trees. The primary purpose of most trees was to allow for comparisons to investigate the sensitivity of results to sampling and inference methods. Most of those trees are not shown, but comparisons among them are described in Sections 3-3.3 to 3-3.7. A representative set of trees, including those we consider to be our best estimates of phylogeny, are presented as follows: 119-taxon analyses of the full mitochondrial dataset (Fig. 3-1), 46-taxon analyses of the nuclear dataset (Fig. 3-2), 46-taxon analyses of the combined mitochondrial and nuclear datasets (Fig. 3-3), and the 119-taxon analyses of the combined mitochondrial and nuclear datasets (Fig. 3-4)

3-3.3 Effects of mitochondrial locus sampling on phylogeny inference.— We found that adding sequence, even when sampled unevenly across taxa, led to the addition of strongly-supported nodes (and not to switches in patterns of strongly supported relationships) in phylogeny reconstructions, as compared to results from smaller datasets. For the 46-taxon analyses, we obtained cytochrome *b* and ND2 sequences of all species. Using cytochrome *b* alone, we recovered a ML tree in which 17 of 43 nodes, most of them

uniting taxa at the tips of branches (often congeners), were resolved with strong (≥70% nonparametric bootstrap) support. Adding ND2 sequences yielded a tree with 26 strongly-supported nodes, including all those found with cytochrome *b* alone. Adding all remaining sequence from the full mitochondrial alignment had a similar effect—35 nodes were recovered with strong support, including all but one node found with the ND2 plus cytochrome *b* dataset. The latter dataset recovered *Nesopsar* as sister to the rest of the grackles and allies clade (86% bootstrap support for the monophyly of all grackles and allies exclusive of *Nesopsar*), whereas with the full dataset, the pattern of relationships among *Nesopsar*, *Agelaius*, and *Molothrus* (and thus which of those taxa—individually or in combination—is sister to all other grackles and allies) was not resolved with confidence.

Results of the 118 or 119-taxon analyses were qualitatively nearly identical to those of the 46-taxon datasets, even though sequence sampling across taxa was quite heterogeneous (i.e. many taxa had a large percentage of missing data). The full mitochondrial dataset yielded a ML tree (Fig. 3-1) with strong support for 94 of 116 nodes, including all 54 and 79 strongly-supported nodes recovered, respectively, with the cytochrome *b* and the ND2 plus cytochrome *b* datasets. Likewise, support for the position of *Nesopsar* differed, in the same manner as described for the 46-taxon datasets, between results of the ND2 plus cytochrome *b* and the full mitochondrial datasets.

3-3.4 Effects of taxon sampling on phylogeny inference.—Taxon addition can sometimes bolster phylogeny inference, but in this study, after pruning trees to include only the taxa in the less comprehensive analyses, we found that taxon sampling had almost no effect on the pattern or number of strongly-supported nodes. By those measures, ML reconstructions from the 46 (Fig. 3-2) and 54-taxon concatenated nuclear-only datasets were identical, as were results from the 46 and 119-taxon datasets of combined nuclear and mitochondrial sequences (Figs. 3-3 and 3-4). Only the trees generated exclusively from mitochondrial data exhibited any significant differences in support at equivalent nodes. The larger ND2 plus cytochrome *b* datasets and full mitochondrial alignments robustly (71 and 95% bootstrap support, respectively) placed *Icterus mesomelas* closer to *cucullatus* than *parisorum*, whereas the 46-taxon analyses failed to resolve those relationships with confidence or to recover that same topology with even plurality support. Furthermore, the tree from the 119-taxon full mitochondrial dataset (Fig. 3-1) placed *Agelaius phoeniceus* closer to *tricolor* (72% bootstrap) than to *xanthomus*, whereas support for that relationship was weak (62%) using the equivalent 46-taxon dataset. On the other hand, all of the 46-taxon mitochondrial datasets strongly (85–93% bootstrap) supported a closer relationship of *Ocyalus* to *Cacicus sclateri* than to *Psarocolius*, whereas support for that pattern was weak (49–64%) using the mitochondrial datasets with 118 or 119 taxa. Also, the 46-taxon ND2 plus cytochrome *b* analysis strongly (89% bootstrap) supported a clade composed of *Quiscalus* and *Euphagus*, but support for that relationship was lacking (29%) using the equivalent 118-taxon dataset due to instability in the placement of those taxa relative to *Dives atroviolaceus*.

3-3.5 Concordance of inferences from nuclear and mtDNA.—Phylogenies generated from separate nuclear and mitochondrial datasets showed strong support for a majority of relationships in the 46-taxon

analyses, but yielded somewhat different topologies; however, with only one exception, topological differences occurred at nodes that were poorly supported by at least one of the two datasets. The single instance of strong conflict concerned basal relationships within the meadowlarks and allies. According to mitochondrial data, *Xanthocephalus* and *Dolichonyx* are sister taxa (97% bootstrap support, 99% posterior probability) that compose a clade sister to *Sturnella*. By contrast, phylogenies inferred from nuclear data place *Xanthocephalus* sister to a strongly (92% bootstrap, 100% posterior) supported *Dolichonyx-Sturnella* clade.

We did not produce a 54-taxon mitochondrial phylogeny to allow for direct comparisons with our 54-taxon nuclear phylogenies, but placements of the 8 taxa with partial data in those nuclear trees (not shown, but match following descriptions to Fig. 3-2) were congruent with the tree from the 119-taxon full mitochondrial alignment (Fig. 3-1). The nuclear data put (a) *Sturnella bellicosa* and *loyca* with *superciliaris*, thus supporting monophyly of the red-breasted meadowlarks (83% bootstrap, 100% posterior probability), (b) *Sturnella lilianae* and *magna* together (73, 98), and that pair sister to *neglecta* (99, 100), thus supporting monophyly of the yellow-breasted meadowlarks, (c) *Cacicus solitarius* and *Clypicterus* into a poorly-resolved grouping with *Ocyalus* and *Cacicus sclateri* (90, 100), (d) *Cacicus melanicterus* outside a well-supported (76, 99) clade containing *Psarocolius, Ocyalus, Clypicterus,* and all other *Cacicus,* and (e) *Agelasticus xanthopthalmus* with *cyanopus* (99, 100).

3-3.6 Inference power of separate and combined mtDNA and nuclear loci.—Direct comparison of the 46-taxon ML tree built using all 4 nuclear loci (Fig. 3-2) to the 46-taxon tree inferred from the full mitochondrial alignment illustrates the relative strengths and weaknesses of those two classes of markers in our study, and provides insight into their separate contributions to our reconstructions of phylogeny using the combined dataset. A striking feature of the ML analysis of nuclear loci was its nearly complete failure to robustly resolve relationships within the South American endemic clade—only 1 node out of 13, uniting *Xanthopsar* with *Pseudoleistes*, received strong (85% bootstrap) support, whereas the mitochondrial analysis recovered 11 well-supported nodes within that clade. By contrast, with respect to resolving the rest of the tree, the mitochondrial dataset performed no better, overall, than the nuclear dataset—both datasets resolved 24 of 30 nodes with confidence. The superior performance of mitochondrial sequences for resolving short internodes was obvious within the South American clade, which was inferred using full mitogenomic sequences of most species, but was less evident in the rest of the tree, built from less comprehensive sampling of the locus. Of the 6 nodes outside the South American clade that were not resolved in ML trees using the nuclear dataset, 4 were failures to resolve relationships among congeners (within *Psarocolius*, *Molothrus*, *Agelaius*, and *Euphagus*), and two involved uncertainty in the resolution of basal relationships among major clades within the grackles and allies—specifically, the relationships among *Molothrus*, *Nesopsar-Agelaius*, *Dives-Euphagus-Quiscalus*, and the South American endemic clade. To compare, two of the nodes that the mitochondrial dataset failed to resolve were relationships among congeners (within *Icterus* and *Agelaius*), two involved basal relationships within the grackles and allies—specifically, the relationships among *Molothrus*, *Nesopsar*, *Agelaius*, and a clade composed of *Dives-*

Euphagus-Quiscalus plus the South American endemics—and two involved resolution of relationships among 4 basal icterid clades, namely *Amblycercus*, all other caciques and oropendolas, *Icterus*, and the grackles and allies.

To the extent that their strengths were complementary, the nuclear and mitochondrial datasets had potential for additive or synergetic fusion of their best qualities in the 46-taxon ML analysis of the combined data (Fig. 3-3). However, some of their limitations were congruent, and (as described previously) the datasets did exhibit some conflict with one another, so although confidence values of many nodes increased, the ML phylogeny built from the combined data had a net gain of only two strongly-supported nodes more than the tree from the mitochondrial dataset, for a total of 37 strongly-supported nodes out of 43. The ML tree from the combined data was very similar to the mitochondrial tree except that it exhibited (a) strong support at 4 nodes robustly recovered by nuclear loci, including resolution of relationships within *Icterus* and *Agelaius*, the sister relationship of *Amblycercus* to all other caciques and oropendolas, and the sister relationship of *Icterus* to the grackles and allies, (b) lower confidence for some nodes within the South American blackbirds, including one that dropped from 85% to 69% bootstrap support, and (c) placement of *Xanthocephalus* and *Nesopsar* in the positions robustly recovered with nuclear loci (i.e. sister to the other meadowlarks, and sister to *Agelaius*, respectively), but with poor support due to conflict with the mitochondrial signal. Each of those 3 features were also found in the ML analysis of the combined datasets for all 119 taxa (Fig. 3-4), and together with robust placement of *Cacicus solitarius* as sister to the other caciques (excepting *C. melanicterus* and *Amblycercus*) due to the contribution of nuclear loci, constituted its main differences from the mitochondrial tree (Fig. 3-1). The ML tree from the combined dataset contained 97 well-supported nodes—a total of 3 more than the mitochondrial tree.

Because a species-tree analysis could not be conducted using the mitochondrial dataset alone, we compared the 46-taxon tree using the nuclear loci to the tree from the combined dataset and found no significant incongruence. Placement of *Nesopsar* sister to *Agelaius* got strong (98% posterior) support using nuclear sequences, but not with the combined dataset (87% posterior); the other 14 strongly-supported nodes in the former analyses were recovered in the latter, which found robust support for 21 of 43 nodes.

3-3.7 Concordance of results from different methods of analysis.—We found no strongly-supported topological differences between trees inferred from a given dataset using different optimality criteria. Furthermore, analyses using GARLI and MrBayes almost always agreed with one another in assigning strong support, or not, to nodes (according to the thresholds that we selected for comparing bootstrap support to posterior probability, i.e. ≥ 70% and ≥ 95%, respectively), though in a few cases, assessments of support were sharply discordant.

Trees generated by GARLI and MrBayes using the full mitochondrial dataset for all 119 taxa (Fig. 3-1) had equivalent topologies and differed in assignments of strong support at only 4 nodes: (1) placement of *Sturnella bellicosa* with *loyca* and *defilippii* (72% bootstrap, 78% posterior), (2) placement of *Psarocolius bifasciatus yuracares* (78, 80) sister to the *montezuma-cassini-guatimozinus* clade, (3) pairing

Icterus icterus with *croconotus* exclusive of *jamacaii* (86, 80), and (4) monophyly of *Quiscalus-Euphagus* exclusive of *Dives atroviolaceus* (73, 73).

For the 46-taxon nuclear locus dataset, the single best tree from GARLI and the consensus tree from MrBayes had identical topologies; the same was true for the 54-taxon trees from nuclear data. Apart from uniting *Xanthopsar* with *Pseudoleistes*, the pattern of relationships within the South American endemic clade received poor support under ML, so it would not warrant a mention except that the same pattern was recovered by MrBayes, and with strong support at 4 additional nodes (Fig. 3-2). Of those 4 nodes, one was not found in the ML bootstrap consensus (it had 8% support) yet had a posterior probability of 96%, another had respective bootstrap and posterior supports of 24% and 97%, a third node correspondingly received 22% and 95% support, and the remaining node, which united the two *Pseudoleistes* species, got 63% and 100% support. Except for monophyly of *Pseudoleistes*, these nodes were not recovered in the species-tree inferred from nuclear loci in *BEAST. That analysis yielded its own (unique) poorly-supported hypothesis of relationships within the South American clade. Otherwise, the species-tree was topologically like the concatenated nuclear-only analyses in most respects, though basal divergences within the grackles and allies more closely matched trees inferred from the combined nuclear and mitochondrial datasets.

Analyses of the 46-taxon combined nuclear and mitochondrial dataset with GARLI and MrBayes recovered identical tree topologies (Fig. 3-3), and assignments of strong support differed at only 3 nodes: (1) the sister relationship of *Agelaius phoeniceus* to *tricolor* (70% bootstrap support, 92% posterior probability), (2) the sister relationship of *Nesopsar* to *Agelaius* (60, 100), and (3) monophyly of all other grackles and allies exclusive of *Nesopsar-Agelaius* (62, 100). The *BEAST species-tree of that dataset differed topologically from the trees from concatenated analyses at a few nodes, but all except that pairing *Xanthocephalus* and *Dolichonyx* (as found with mitochondrial sequence) were within the South American clade, and none was strongly-supported.

Using the 119-taxon combined dataset, we found that the topologies of the single best tree from GARLI and the consensus tree from MrBayes were identical, even at poorly supported nodes, with one exception—the Bayesian tree found *Curaeus forbesi* sister to *Gnorimopsar* with poor (61% posterior) support, whereas in ML, those lineages were sequentially nested branches in relationship to other taxa (Fig. 3-4). Assessments of whether support was robust, given our chosen thresholds, agreed at all but 8 of 116 nodes. Five inferences received strong support only under ML: (1) exclusion of *Cacicus solitarius* (with 73% bootstrap, 78% posterior support) from a clade containing all other *Cacicus* except *melanicterus*, (2) placement of *Psarocolius bifasciatus yuracares* (77, 87) sister to the *montezuma-cassini-guatimozinus* clade, (3) pairing *Icterus icterus* with *croconotus* exclusive of *jamacaii* (88, 81), (4) support for a clade composed of *Icterus galbula*, *abeillei*, *bullockii*, and *pustulatus* (71, 87), and (5) monophyly of *Quiscalus-Euphagus* exclusive of *Dives atroviolaceus* (73, 78). Three nodes were strongly supported only with Bayesian methods: (1) monophyly of *Quiscalus lugubris* (64, 98), (2) placement of *Nesopsar* sister to *Agelaius* (49, 99), and (3) monophyly of all other grackles and allies exclusive of *Nesopsar-Agelaius* (50,

99). To test whether the finding of strong support for those 3 nodes was peculiar to MrBayes, we used BEAST to analyze the same concatenated dataset and got the same results (posterior probabilities of 97, 100, and 100%, respectively).

3-4. DISCUSSION

With this study, we present the first comprehensive species-level phylogeny of the Icteridae. By using mitochondrial gene sequences from all currently-recognized taxa, together with strategic sampling of 4 nuclear loci and whole mitochondrial genomes at the generic level, we were able to resolve most relationships with high confidence. Our best-resolved phylogeny (Fig. 3-4) exhibits a topology that is consistent with the strongly-supported results of past studies, but it also contains many robustly-resolved inferences of relationship that eluded them, and which they did not recover with even plurality support. These novel hypotheses of relationship include some unexpected placements of taxa that had not been included in previous molecular phylogenies, resolution of the relationships among major subclades within Icteridae, and resolution of generic-level relationships within the largest of those subclades, the grackles and allies.

3-4.1. Congruence of results from different analyses.—Although we have no way of determining whether or how the inferences made with our most complete dataset were biased by the uneven coverage of mitochondrial and nuclear sampling across taxa, we can state with confidence that the heterogeneous addition of sequence data did not undermine recovery of relationships that received robust support with smaller datasets with more uniform coverage. The congruent results of the many different analyses that we performed, demonstrate that our findings were robust to variation in mitochondrial sampling, taxon sampling, and use of signal derived from either the mitochondrial or nuclear genomes. In general, nuclear loci were less successful than mitochondrial sequence for informing inferences of relationship at the tips of the tree, but they provided much stability to the resolution of basal relationships, and just as importantly, corroborated many of the surprising results of previous studies, which were based on mitochondrial data alone.

3-4.2. Icteridae and its major subclades.—Although the composition of Icteridae (as traditionally recognized) has rarely been questioned, until now (Barker et al. *in press*, and this study), robust support for its monophyly, and for resolving basal relationships within it, has been lacking. Lanyon and Omland (1999), using mitochondrial cytochrome *b* sequences, found strong (though, for the meadowlarks and allies, marginal) support for 5 major clades within the Icteridae, but not for their interrelationships, or for icterid monophyly (despite limited outgroup sampling). Klicka et al. (2007), using ND2 plus cytochrome *b*, found strong support for the Icteridae excluding meadowlarks and allies, but not for the family as a whole, or for monophyly of the meadowlarks and allies, or for basal icterid relationships. Our results using only ND2 plus cytochrome *b* were much the same, but using the full mitochondrial dataset, we recovered all Icteridae

(against limited outgroup sampling), the meadowlarks and allies, and the rest of Icteridae, as clades with strong support. However, the full mitochondrial alignment was no better than smaller datasets for robustly resolving the topology of the set of short internodes connecting the long basal branches of the other 4 major clades—*Amblycercus*, the caciques and oropendolas, the orioles, and the grackles and allies—within Icteridae to one another.

Nuclear loci allowed us to reconstruct basal relationships within Icteridae with high confidence, and they resolved homoplasy in the mitochondrial signal such that support values were even higher using the combined dataset. We found a graded pattern of relationship among major clades, with the meadowlarks and allies sister to the rest of Icteridae (as in previous studies), and within that, the caciques and oropendolas, *including Amblycercus*, sister to a pairing of the orioles with the grackles and allies. Note that this pattern does not match mitochondrial topologies, which grouped (with poor support) the orioles with the caciques and oropendolas, a suggestion that concorded with traditional suppositions (e.g. American Ornithologists' Union 1983).

3-4.3. Meadowlarks and allies.—The meadowlarks (*Sturnella*) generally inhabit open grasslands and are notable for their stocky build, long bill, relatively short tail, and contrastingly red or yellow breast versus cryptically-streaked dorsal color patterning. Prior to the present study, which included all 10 species of meadowlarks and allies, a thorough molecular treatment was lacking. Lanyon and Omland (1999), in their study of Icteridae, included 6 of the 10 species. Barker et al. (2008) also included 6 species in their treatment of the yellow-breasted meadowlarks, but the 3 red-breasted species served only as outgroups. We found that relationships of the 3 red-breasted species not included in previous studies fit traditional expectations: each is sister to the species with which it has sometimes been considered conspecific—*S. militaris* with *superciliaris*, *loyca* with *defilippii*. Both our mitochondrial and nuclear datasets supported monophyly of the red and yellow-breasted groups, which we note are genetically more divergent (~15%) than any other congeners within Icteridae. The meadowlarks were once divided between the genera *Sturnella*, *Leistes*, and *Pezites*, which this study demonstrates were all monophyletic as originally defined (though support for placement of *S. bellicosa* with *loyca* and *defilippii*, to constitute *Pezites*, was weak in most analyses), but Short (1968) argued for their merger for lack of substantial morphological and ecological divergence. Sibley and Monroe (1990) gave new life to *Leistes*, citing Parker and Remsen (1987), who argued for its continued recognition based on behaviors shared with *Agelaius phoeniceus* and not with *Sturnella*. When molecular studies later found *S. bellicosa* more closely related to *L. superciliaris* than to the yellow-breasted *Sturnella spp.*, the genus *Leistes* was abandoned since it made *Sturnella*, as then defined (i.e. inclusive of *Pezites*), paraphyletic. Note that based on our phylogeny, the behaviors and ecological attributes of the former *Leistes spp*. that have generally been interpreted as primitive traits (shared with *Agelaius*), are instead, derived.

One of the most surprising findings of the first molecular phylogenies of Icteridae (Lanyon 1994, Lanyon and Omland 1999), was that *Xanthocephalus* is not allied with *Agelaius*, as had always been supposed from behavioral and ecological similarities, but rather, that it is most closely related to

Dolichonyx and *Sturnella*. Our nuclear and mitochondrial datasets both supported that unexpected grouping, but it was in the precise pattern of divergences among those three genera that we encountered the only instance (in this study) of conflict between strongly-supported nodes inferred from nuclear versus mitochondrial sequences. Nuclear data placed *Xanthocephalus* sister to a *Dolichonyx*-*Sturnella* clade, whereas mitochondrial data supported a sister relationship between *Xanthocephalus* and *Dolichonyx*. We obtained all 4 nuclear loci used in this study, and a substantial amount of mitochondrial sequence, from each of these taxa, so it seems that many additional loci will be necessary to resolve these relationships with confidence. It is perhaps worth noting that although *Xanthocephalus* and *Dolichonyx* are more closely related to *Sturnella* than to other icterids, that their relationship is still a very distant one with respect to genetic divergence. These taxa are peculiar (e.g. *Dolichonyx* is unique among blackbirds, and unusual among passerines, for undergoing two complete molts per year and for being an interhemispheric migrant), and so phenotypically divergent from one another, that their morphologies and behaviors are not particularly suggestive of one resolution of their relationships over another.

3-4.4. Caciques and oropendolas, including Amblycercus—The caciques and oropendolas (~23 sp.) are inhabitants of tropical forests, where their pendant nests and displays can make them quite conspicuous, especially in the case of polygynous colonial species. They span an amazing range of sizes, from small species like *Cacicus sclateri* (23 cm, 57 g), to enormous beasts like *Psarocolius montezuma* (up to 53 cm, 560 g), which dwarf all other blackbirds (Fraga 2011). For practitioners of molecular phylogenetics, the most notable quality of the group is that it has evolved DNA sequences that when analyzed, yield well-resolved and strongly-supported hypotheses of relationship (e.g. this study, and see Price and Lanyon 2002), even when internodes are short (One explanation may be small effective population sizes in polygynous species). Our main concerns were to achieve complete taxon sampling, and to use nuclear loci to test some of the surprising findings of previous studies, which used only mitochondrial DNA. We also propose a number of taxonomic revisions, many of them already long overdue given results of previous studies (i.e. Price and Lanyon 2004a, 2004b).

Mitochondrial DNA, even with increased sample size, was not able to recover the cup-nesting cacique, *Amblycercus*, as sister to the typical caciques and oropendolas, but nuclear loci did so with very strong support (in both concatenated and species-tree analyses), as did the combined dataset. Like mitochondrial data, the nuclear loci indicate that the genetic divergence of *Amblycercus* from the other caciques and oropendolas is substantial. Nuclear markers also supported the position of *Cacicus melanicterus* outside the rest of the typical caciques and oropendolas (see Results, section 3-3.5), and the combined dataset placed it sister to them with strong support; consequently, that taxon should be restored to *Cassiculus*. The remaining caciques and oropendolas sort into two clades, one containing all species currently placed in *Psarocolius*, and the other comprising mostly *Cacicus* spp.

Mitochondrial data placed *Cacicus solitarius* sister to *Psarocolius*, but with only weak support. By contrast, nuclear loci strongly supported a sister relationship of *Cacicus solitarius* to the other *Cacicus* spp., as did analysis of the combined dataset under ML (Bayesian analysis recovers the same topology with

weak support). Consequently, our study found that *Cacicus solitarius* need not be renamed to *Procacicus*, as has been proposed (Fraga 2005), though it could be done without rendering a revised *Cacicus* (i.e. exclusive of *Cassiculus*) paraphyletic. A very surprising finding of previous studies was the sister relationship between *Ocyalus* and *Clypicterus* (Freeman and Zink 1995, Price and Lanyon 2002), and the position of those taxa well outside of *Psarocolius*. Those results led to restoring *Clypicterus* to its monotypic genus (from *Psarocolius*); however, subsequent work (Price and Lanyon 2004a), which even more surprisingly, demonstrated that those taxa are imbedded well within *Cacicus*, has thus far been ignored in taxonomic revisions. We found strong nuclear (and combined) support for placement of *Ocyalus* and *Clypicterus* in the *Cacicus* clade; consequently, those species should be renamed to that genus. We were not able to resolve the relationships among *Ocyalus*, *Clypicterus,* and *Cacicus haemorrhous* with confidence. Elsewhere within *Cacicus*, we also recovered the same pattern of relationships found by Price and Lanyon (2004a), except that we included, for the first time, *C. koepckeae*, which grouped sister to *sclateri*, as anticipated (Cardiff and Remsen 1994), with strong support. Following Price and Lanyon (2004a), we included samples of deeply divergent subspecies of *Amblycercus* and several *Cacicus* taxa, which should probably be recognized as species. Note that some authorities (e.g. Jaramillo and Burke 1999, Fraga 2011, Gill and Donsker 2012) recognize *Cacicus* (*uropygialis*) *microrhynchus* and then treat *C. u. pacificus* as a subspecies of *microrhynchus*, when in fact, *pacificus* is more closely related to *C. u. uropygialis.*

Within *Psarocolius*, we recovered the same pattern of relationships found by Price and Lanyon (2002, 2004a), but we added 3 species missing from previous studies. We found that two of the newly sampled taxa, *P. cassini* and *guatimozinus*, are sister to one another, and that *montezuma* is sister to that pair. We expected the other newly-sampled taxon, *P. b. bifasciatus*, to group with *P. b. yuracares* because those taxa are usually treated as conspecific, but instead, it grouped with *viridis*. It is perhaps worth remembering that, because of its quite different appearance, the close relationship of *viridis* to those other 5 taxa (which formerly constituted the genus *Gymnostinops*), was itself a surprise when first discovered (Price and Lanyon 2004b), although song characteristics supported the alliance; indeed, that result has yet to be embraced by taxonomic authorities (e.g. Gill and Donsker 2012, Remsen et al. 2012), who still list *viridis* between the much more distantly-related *atrovirens* and *decumanus*. Still, given the very similar appearance of *bifasciatus* to *montezuma, cassini,* and *guatimozinus*, and its very shallow genetic divergence with its putative sister taxon, *viridis* (thus limiting opportunity for phenotypic divergence of *viridis*), this hypothesis of their relationship strains credibility and will require further investigation. Note that divergences among all these taxa are very shallow, but that all nodes were strongly-supported under ML. Finally, note that we chose, for the sake of clarity, not to include multiple representatives of *P. decumanus* and *angustifrons*, even though both species contain divergent lineages; we instead refer the reader to Price and Lanyon (2002, 2004a). Both complexes require further investigation and taxonomic revision. Some authorities have suggested recognizing *P. angustifrons alfredi* as a species, but the divergence of *P. a.*

atrocastaneus from nominate *angustifrons* is much deeper and needs to be considered in conjunction with that proposal.

3-4.5. Orioles.—The orioles, a group of small arboreal or shrub-dwelling icterids, many of them with distinctive orange and black plumage patterns, represent the second-largest of the major clades within Icteridae, yet all ~33 species are classified into the same genus, *Icterus*. The orioles have been the subject of intensive systematic study (Omland et al. 1999, Lovette et al. 2001, Allen and Omland 2003, Sturge et al. 2009, Jacobsen et al. 2010, Jacobsen and Omland 2011), including very thorough sampling at the subspecies level, and use of both mitochondrial DNA and multiple autosomal and Z-linked nuclear introns. These studies have generally found high concordance between signals, and been successful at resolving speciation events separated by very short internodes (Jacobsen et al. 2010), but they have also uncovered instances of conflict between nuclear and mitochondrial markers, conflicts that are unlikely to be results of incomplete lineage sorting (Jacobsen and Omland 2011). It seems that introgressive exchanges among the ancestors of a few species, some of which are involved in different hybridization interactions at present, has complicated the histories of their genomes. Efforts by other researchers to probe these complex cases, and to reconstruct the phylogeny of *Icterus* with even more loci, are well in hand; consequently, we chose, in our study, to direct our resources elsewhere within Icteridae, rather than to match or duplicate those efforts.

Nuclear marker sampling in our study was limited to one representative from each of the three major clades within *Icterus*. Consequently, we did not have much opportunity to, nor did we, encounter any cases of conflict between nuclear and mitochondrial signals. The pattern of relationships that we recovered was equivalent to that described in previous studies that utilized essentially the same mitochondrial dataset (e.g. Omland et al. 1999). Likewise, we were no more successful than previous studies (e.g. Sturge et al. 2009) at resolving relationships within the Caribbean radiation that includes the members of the former *I. dominicensis* complex, a task for which whole mitochondrial genomes might prove useful. One outstanding problem that we did address, was the relationships among the troupials, a group of orioles that are unusual for their large size, white irides, and blue-colored bare skin around their eyes. We included *I. jamacaii*, which is the only species of oriole missing from previous studies, as a COX1 sequence obtained from GenBank (Table 3-2), and hoped to resolve its position relative to the other two troupials using that lone mitochondrial fragment. A closer relationship of *I. icterus* to *croconotus* than to *jamacaii* received strong support under ML, but not using Bayesian methods. If accurate, this relationship is contrary to treatment of *croconotus* as a subspecies of *jamacaii*, as done in classifications that recognize only two species of troupial (e.g. Sibley and Monroe 1990). Finally, we note that inclusion of samples of *I. cayanensis*, *chrysocephalus*, and *pyrrhopterus* in this study should not be interpreted as an endorsement of resolving species limits within that complex (D'Horta et al. 2008) to those taxa, but rather, was done to illustrate representative divergences within that complex using names that appear in current taxonomic lists and other references.

3-4.6. Grackles and Allies.—The grackles and allies are the most taxonomically diverse major clade of Icteridae, yet most species exemplify the attributes commonly associated with the family (at least

among English speakers), as it is to them, among New World birds, that the "blackbird" label was bestowed. Many species in the clade are entirely black, and like their Old World namesake, spend much time foraging terrestrially within the natural or anthropogenic savannah and marshy grassland habitats that they frequent.

One of the main goals of this study, was to resolve generic-level relationships within the grackles and allies. Although we succeeded for the most part, or at least better than any previous study, a definitive resolution of basal relationships within the clade remains elusive, as neither whole mitochondrial genomes nor the nuclear loci we sampled, provided consistently-robust support for all relationships. Studies based on ND2 plus cytochrome *b* (e.g. Johnson and Lanyon 1999, Eaton 2006) found *Nesopsar* and *Dives* to be sequentially sister to all other grackles and allies, whereas in this study, whole mitochondrial genomes placed *Nesopsar* in a poorly-resolved basal position relative to *Agelaius*, *Molothrus*, and a strongly-supported *Dives-Euphagus-Quiscalus* plus South American endemics clade. Note that although the finding of a sister relationship between the South American blackbirds and *Dives-Euphagus-Quiscalus* also received strong support in combination with nuclear loci, it was dependent on the signal from whole mitogenomes, and thus was novel to this study (e.g. it was not recovered by Barker et al. *in press*). Nuclear loci supported the *Dives-Euphagus-Quiscalus* clade, and further, a sister relationship between *Nesopsar* and *Agelaius*. Though the latter relationship is seemingly compatible with the poorly-supported result from the full mitochondrial dataset, the nuclear and mitogenomic signals were apparently antagonistic, such that in the combined dataset analysis, the *Nesopsar-Agelaius* pairing received inconsistent (49% ML bootstrap, 99% Bayesian posterior probability) support. When coupled with inconsistent (50% ML bootstrap, 99% Bayesian posterior) support for the monophyly of the remaining grackles and allies, these conflicts yielded a less than fully-robust final result for the resolution of basal relationships in the grackles and allies, the topology of which (Fig. 3-4) is altogether unique to this study. We are presently unable to explain the differing signals contained in the nuclear and mitochondrial genomes of *Nesopsar*, but the issue is not a result of differences in overall base composition (Powell et al. 2013).

Another goal of our study was to robustly-resolve relationships among the motley assemblage of enigmatic species that compose the South American endemic clade, which is exceptional for the diversity of plumage patterning, morphology, and habitat preferences of its members, and for reports of cooperative breeding in many species (Fraga 2008). The diversity of the group is reflected in its taxonomy—with 13 genera, 8 of them monotypic, its 19 species account for nearly half of all genera in Icteridae. We sequenced nuclear loci from most species, but analyses of those sequences resolved only 4 nodes with strong support, and two of those were almost certainly erroneous since they (a) received strong support only with Bayesian analysis of the concatenated dataset, (b) were not found, even with weak support, in trees recovered with mitochondrial data, combined nuclear and mitochondrial data, or in any of the species-tree analyses using *BEAST, and moreover, (c) they strongly contradicted several strongly-supported relationships found in other analyses, including (1) a *Macroagelaius-Gymnomystax-Lampropsar-Hypopyrrhus* clade, (2) a *Chrysomus-Xanthopsar-Pseudoleistes* clade, (3) monophyly of *Agelasticus*, (4) a sister relationship of

Curaeus curaeus with *Amblyramphus*, and (5) a sister relationship of *Oreopsar* with *Agelaioides*. In combination with ND2 plus cytochrome *b* sequences, the contribution of noise and misleading signal from the nuclear loci was substantial enough to nullify the mitochondrial signal, yielding a tree with a unique topology and only two strongly-supported nodes within the clade (Barker et al. *in press*). By contrast, trees inferred from ND2 plus cytochrome *b* alone (Johnson and Lanyon 1999, Cadena et al. 2004, Eaton 2006), though lacking support at many nodes, were topologically similar to our best inferences. We found that whole mitochondrial genomes were able to robustly resolve most nodes in the group, even in combination with the nuclear dataset (though its inclusion weakened support values). The only case of strongly-supported agreement between those markers was the sister relationship between *Xanthopsar* and *Pseudoleistes*.

Apart from stronger support for many nodes, the novel findings of this study, with respect to the South American clade, include recovery of a *Macroagelaius-Gymnomystax-Lampropsar-Hypopyrrhus* clade, and inclusion of two species that were absent from previous molecular studies. As expected, *Macroagelaius subalaris* was recovered as sister to *M. imthurni*. By contrast, *Curaeus forbesi* did not group with *C. curaeus* in any analyses, but rather, defined its own long branch in a grade between a strongly supported *C. curaeus-Amblyramphus* clade and *Gnorimopsar*. Morphologically, *C. curaeus*, *C. forbesi*, and *Gnorimopsar* are all quite similar—for example, they all have distinctively lanceolate feathers, with flattened rachides, around the head (but note that this trait is also found in distantly-related *Hypopyrrhus*)—and specimens of *forbesi* have been mistaken for *Gnorimopsar* (Short and Parkes 1979). In a few analyses, we recovered *forbesi* as sister to *Gnorimopsar* with weak support; regardless, the taxon clearly does not belong in *Curaeus*, and unless a wholesale taxonomic revision of the South American clade is undertaken to lump most of the group into a single genus, it seems that naming *forbesi* to a new monotypic genus is in order. Elsewhere within the clade, our results concord with past studies, and so taxonomic revisions of the past decade remain appropriate, including naming several former *Agelaius* spp. to *Agelasticus* and *Chrysomus* (Lowther et al. 2004). Note that taxonomies currently differ in the naming of *Oreopsar*, either as *Oreopsar badius*, or following Lowther (2001), as *Agelaioides oreopsar*, a usage that recognizes the sister relationship between that taxon and *A. badius*. Given that avian taxonomists have generally been loath to placing species with distinctly different (judged according to a threshold that is locally contextual) characteristics within the same genus, that case and the South American clade in general, presents a challenge to those who seek also to avoid monotypic genera. Certainly the strongest case for grouping genera (based on genetic divergence and the strength of inference of their relationship) could be made for renaming *Xanthopsar* to *Pseudoleistes*, but taxonomists have thus far not chosen to do so because those species are perceived as being very different in kind. The results of molecular phylogenetic studies have not led to any reappraisals of the similarities of the South American endemic species, so unless a different standard, such as closeness of relationship, is adopted to measure the utility of generic naming, the taxonomy of that clade does seem an appropriate reflection of its diversity.

Another case in which taxonomic revision is in order is that of *Dives atroviolaceus*, which we found is sister to *Quiscalus-Euphagus*, not other *Dives*. Although exact placement of *atroviolaceus* relative to *Euphagus* and *Quiscalus* was somewhat unstable (Additional sequencing, now underway using a fresh tissue specimen, may yield a definitive result.), its closer relationship to one or both of those taxa than to *Dives* received strong support. Consequently, *atroviolaceus* should be restored to its former monotypic genus, *Ptiloxena*, as is already done in one prominent current reference (Fraga 2011), based on behavioral characteristics and following the suggestion of Webster (2003), who measured divergences among species according to skeletal measurements. Ironically, because it is precisely opposite to our finding despite leading to the same nomenclatural solution, Webster (2003) argued for the distinctiveness of *atroviolaceus*, and thus its renaming, based on its morphological divergence from other *Dives* and *Quiscalus*, and he suggested the revised *Dives* and *Quiscalus* were morphologically similar enough that they might be merged. Finally, the genus *Quiscalus* warrants additional phylogeographic study and revision of species limits because several species (some of them not monophyletic) contain deeply divergent lineages (see Powell et al. 2008).

3-4.7. Concluding thoughts.—We hope that this first complete species-level phylogeny of the New World blackbirds proves a useful resource for additional comparative studies of morphological, behavioral, and ecological trait evolution in the Icteridae. For studies that require an explicit hypothesis of relationships and divergences, we recommend using the topology from our ML analysis of the complete dataset (Fig. 3-4). We also hope that this paper serves as a reference for identifying opportunities for more detailed molecular studies of phylogeography and species limits in certain taxa, and that it inspires a new generation of multi-locus and mitogenomic studies of the phylogeny of Icteridae.

Table 3-1. Taxa and specimens used in phylogenetic analyses of New World blackbirds (Icteridae).

Taxon	English name[a]	Voucher specimen or tissue[b]	Collecting locality
Outgroups:			
Teretistris fernandinae	Yellow-headed Warbler	ANSP-B5548, STRI CUTFE 5548	Cuba: Guantánamo
Icteria virens	Yellow-breasted Chat	UWBM CDS4131, STRI USIVI 4131	USA: Washington
		LSUMZ B3892	USA: Louisiana
Seiurus aurocapilla	Ovenbird	STRI-PRSAU1	Puerto Rico: Patillas
		UMMZ 224992	USA: Michigan
Oreothlypis gutturalis	Flame-throated Warbler	LSUMZ B26458	Panama: Chiriquí
Ingroup taxa used in 46-taxon analyses:			
Psarocolius wagleri	Chestnut-headed Oropendola	LSUMZ B27280	Costa Rica: Cartago
Psarocolius decumanus	Crested Oropendola	FMNH 324065	Peru: Madre de Dios
		CUMV-Bird 52534, MACN-Or-ct 1130	Argentina: Jujuy
		–	Panama
Psarocolius montezuma	Montezuma Oropendola	LSUMZ 164424	Panama: Colón
Ocyalus latirostris	Band-tailed Oropendola	ANSP 177928, LSUMZ B3625	Peru: Loreto
Cacicus sclateri	Ecuadorian Cacique	ANSP 177931, LSUMZ B103568	Peru: Loreto
Amblycercus holosericeus australis	Yellow-billed Cacique	LSUMZ 98900	Peru: Puno
Icterus mesomelas	Yellow-tailed Oriole	LSUMZ 109279	Panama: Darién
Icterus cucullatus	Hooded Oriole	BB-BEHB25	unknown
		FMNH 341931	USA: California
		UWBM 48323	USA: Arizona
Icterus parisorum	Scott's Oriole	FMNH 341943	USA: California
		FMNH 334367	USA: Arizona
Nesopsar nigerrimus	Jamaican Blackbird	FMNH 331150	Jamaica: Portland
Gymnomystax mexicanus	Oriole Blackbird	FMNH 339743	Venezuela: Falcón
Macroagelaius imthurni	Golden-tufted Mountain Grackle	FMNH 339783	Venezuela: Bolívar
Lampropsar tanagrinus	Velvet-fronted Grackle	ANSP 177921, LSUMZ B103505	Peru: Loreto
Gnorimopsar chopi	Chopi Blackbird	FMNH 334679	Bolivia: Santa Cruz
Curaeus curaeus	Austral Blackbird	AMNH 826156	Chile: Magallanes
Curaeus forbesi	Forbes's Blackbird	MPEG 72143 CPE-II 040	Brazil: Pernambuco

Amblyramphus holosericeus	Scarlet-headed Blackbird	FMNH 334662	Bolivia: El Beni
Agelasticus cyanopus	Unicolored Blackbird	FMNH 334636	Bolivia: El Beni
Agelasticus thilius	Yellow-winged Blackbird	FMNH 334615	Bolivia: Oruro
Chrysomus icterocephalus	Yellow-hooded Blackbird	FMNH 339772	Venezuela: Sucre
Xanthopsar flavus	Saffron-cowled Blackbird	FMNH 330747	Brazil: Rio Grande do Sul
Pseudoleistes guirahuro	Yellow-rumped Marshbird	FMNH 330795	Brazil: Rio Grande do Sul
Pseudoleistes virescens	Brown-and-yellow Marshbird	FMNH 330796	Brazil: Rio Grande do Sul
Oreopsar bolivianus	Bolivian Blackbird	FMNH 334687	Bolivia: El Beni
Agelaioides badius	Baywing	FMNH 330801	Brazil: Rio Grande do Sul
Molothrus rufoaxillaris	Screaming Cowbird	FMNH 330805	Brazil: Rio Grande do Sul
		–	Argentina: Formosa
Molothrus oryzivorus	Giant Cowbird	FMNH 324097	Peru: Madre de Dios
		LSUMZ 134021	Bolivia: Pando
		USNM 587829	Guyana
Molothrus ater	Brown-headed Cowbird	FMNH 350707	USA: Chicago
		UMMZ –	USA: Michigan
		BIOUG:SPP1681-70648	Canada: Ontario
		MBMC jk 96-016	USA: Minnesota
Dives dives	Melodious Blackbird	MBMC 7100	Honduras: Copán
Dives warczewiczi	Scrub Blackbird	LSUMZ 113959	Peru: Lambayeque
Agelaius phoeniceus	Red-winged Blackbird	BB-96 Tordoff	USA: Minnesota
		FMNH 341893	USA:Louisiana
Agelaius tricolor	Tricolored Blackbird	LSUMZ 130833	USA: California
		USNM 632199	USA: California
Agelaius xanthomus	Yellow-shouldered Blackbird	BB-SML 86-1	Puerto Rico
Euphagus carolinus	Rusty Blackbird	FMNH 333317	USA: Illinois
		ROM 1B-3617	Canada: Ontario
Euphagus cyanocephalus	Brewer's Blackbird	FMNH 342000	USA: California
		FMNH 341985	USA: California
Quiscalus quiscula	Common Grackle	FMNH 341733	USA: Illinois
Quiscalus mexicanus W^{c}	Great-tailed Grackle	FMNH 341975	USA: California
Quiscalus nicaraguensis	Nicaraguan Grackle	MBMC 4375	Nicaragua: Tipitapa

Sturnella superciliaris	White-browed Blackbird	FMNH 334657	Bolivia: Santa Cruz
		LSUMZ B9630	Bolivia: Pando
		USNM 635873	Uruguay: Atigas
		NRM 947221	Paraguay
Sturnella neglecta	Western Meadowlark	FMNH 341967	USA: California
		ROM 1B-1038	Canada: Ontario
		FMNH 330040	USA: California
Xanthocephalus xanthocephalus	Yellow-headed Blackbird	LSUMZ 126564	USA: California
Dolichonyx oryzivorus	Bobolink	FMNH 334721	Bolivia: Santa Cruz
		UMMZ 234583	USA: Michigan
		BIOUG: LMA 8101-91770	Canada: Ontario
Additional ingroup taxa included in 119-taxon analyses:			
Clypicterus oseryi	Casqued Oropendola	LSUMZ 120394	Peru: Loreto
Psarocolius viridis	Green Oropendola	USNM 609202	Guyana
		USNM 639199	Guyana: Cuyuni-Mazaruni
Psarocolius atrovirens	Dusky-green Oropendola	FMNH 324106	Peru: Cuzco
Psarocolius a. angustifrons	Russet-backed Oropendola	LSUMZ 120397	Peru: Loreto
Psarocolius cassini	Baudo Oropendola	ANSP 147013	Colombia: Choco
Psarocolius bifasciatus bifasciatus	Para Oropendola	MPEG 61970, FMNH 457579	Brazil: Pará
Psarocolius bifasciatus yuracares	Olive Oropendola	FMNH 324076	Peru: Madre de Dios
Psarocolius guatimozinus	Black Oropendola	LSUMZ B48620	Panama: Darién
Cacicus cela cela	Yellow-rumped Cacique	KUMNH 88289, USNM B04259	Guyana: Berbice
Cacicus cela vitellinus	Yellow-rumped Cacique	LSUMZ 163850	Panama: Colón
Cacicus haemorrhous	Red-rumped Cacique	USNM 621068	Guyana
		USNM 586489	Guyana: Barima-Waini
Cacicus uropygialis uropygialis	Subtropical Cacique	LSUMZ B6093	Ecuador: Morona-Santiago
Cacicus uropygialis microrhynchus	Scarlet-rumped Cacique	STRI PACUR-PC99	Panama
		USNM 608010	Panama: Bocas del Toro
Cacicus uropygialis pacificus	Pacific Cacique	ANSP 182884	Ecuador: Esmeraldas
Cacicus chrysopterus	Golden-winged Cacique	USNM 620761	Argentina
		MACN-Or-ct 987	Argentina: Jujuy
Cacicus chrysonotus chrysonotus	Southern Mountain Cacique	LSUMZ 103278	Bolivia: La Paz

Cacicus chrysonotus leucoramphus	Northern Mountain Cacique	ANSP 182883	Ecuador: Imbabura
Cacicus koepckeae	Selva Cacique	LSUMZ B48621	Peru: Loreto
Cacicus solitarius	Solitary Cacique	FMNH 324089	Peru: Cuzco
		MACN-Or-ct 1403	Argentina: Corrientes
Cacicus melanicterus	Mexican Cacique	UWBM 52185	Mexico: Oaxaca
Amblycercus h. holosericeus	Yellow-billed Cacique	KUMNH 1928	Mexico: Yucatán
		USNM 608009	Panama: Bocas del Toro
Icterus icterus	Venezuelan Troupial	LSUMZ B11328	Puerto Rico
		LSUMZ B48559	Guyana
Icterus croconotus	Orange-backed Troupial	FMNH 324092	Peru: Madre de Dios
		USNM 632494	Guyana: Upper Takutu-Upper Essequibo
Icterus jamacaii	Campo Troupial	LGEMA 2742	Brazil: Piaui
Icterus pectoralis	Spot-breasted Oriole	MMNH 42544	USA: Florida
		KUMNH 109733	El Salvador: La Paz
Icterus graceannae	White-edged Oriole	ANSP 181810	Ecuador: Loja
Icterus cayanensis	Epaulet Oriole	MPEG 40.357	Brazil: Rondônia
		USNM 625332	Guyana
Icterus chrysocephalus	Moriche Oriole	FMNH 339734	Venezuela: Sucre
		USNM 625748	Guyana
Icterus pyrrhopterus	Variable Oriole	FMNH 334608	Bolivia: Santa Cruz
		USNM 614726	Argentina: Entre Ríos
Icterus bonana	Martinique Oriole	STRI MA-IBO2	Martinique
Icterus laudabilis	St. Lucia Oriole	STRI SL-ILA4	St. Lucia
Icterus oberi	Montserrat Oriole	STRI MO-IOB4	Montserrat
Icterus dominicensis	Hispaniolan Oriole	AMNH NKK1112	Dominican Republic
Icterus portoricensis	Puerto Rican Oriole	STRI PR-IDO1	Puerto Rico
Icterus melanopsis	Cuban Oriole	MNHNCu 4/8/92	Cuba
Icterus northropi	Bahama Oriole	BNT REF024	Bahamas: Andros
Icterus prosthemelas	Black-cowled Oriole	KUMNH MBR 4510	–
		MMNH 42542	Mexico: Yucatán

Icterus spurius	Orchard Oriole	NCSM	USA: Colorado
		USNM 626504	USA: Florida
		FMNH 381975	USA: Illinois
Icterus fuertesi	Ochre Oriole	MMNH 42538	Mexico: Veracruz
Icterus wagleri	Black-vented Oriole	MZFC QRO-216	Mexico: Querétaro
Icterus maculialatus	Bar-winged Oriole	INIREB SRF-387	Mexico: Chiapas
Icterus auricapillus	Orange-crowned Oriole	FMNH 261843	Colombia: Boyacá
Icterus chrysater	Yellow-backed Oriole	UWBM 69019	Nicaragua: Chinandega
Icterus graduacauda	Audubon's Oriole	LSUMZ B-4023	USA: Texas
Icterus galbula	Baltimore Oriole	UMMZ 226382	USA: Michigan
		FMNH 350604	USA: Illinois
		ROM 1B-131	Canada: Ontario
Icterus abeillei	Black-backed Oriole	MZFC 9657	Mexico: Querétaro
		MZFC keo-48	Mexico: Michoacán
Icterus bullockii	Bullock's Oriole	MBMC jk95-095	USA: Oregon
		UWBM 59056	USA: Washington
		UWBM 55975	USA: Washington
Icterus pustulatus	Streak-backed Oriole	UWBM 52129	Mexico: Chiapas
		MZFC keo38	Mexico: Jalisco
Icterus leucopteryx	Jamaican Oriole	FMNH 331145	Jamaica: Trelawny
		FMNH 33144	Jamaica: Trelawny
Icterus auratus	Orange Oriole	UAM 7222	Mexico: Yucatán
Icterus nigrogularis	Yellow Oriole	STRI TR-INI1	Trinidad
		USNM 627066	Guyana
Icterus gularis	Altamira Oriole	FMNH	Mexico: Oaxaca
		UWBM 52191	Mexico: Oaxaca
Macroagelaius subalaris	Colombian Mountain Grackle	LACM 40973	Colombia: Santander
Hypopyrrhus pyrohypogaster	Red-bellied Grackle	ICN 33977, IAvH 2078	Colombia: Antioquia
Agelasticus xanthophthalmus	Pale-eyed Blackbird	FMNH 324094	Peru: Madre de Dios
Chrysomus ruficapillus	Chestnut-capped Blackbird	FMNH 330775	Brazil: Rio Grande do Sul
Molothrus aeneus	Bronzed Cowbird	BB-73 James	Mexico: Puebla

Molothrus bonariensis	Shiny Cowbird	LSUMZ 113963	Peru: Lambayeque
		MACN-Or-ct 3062	Argentina: Buenos Aires
		FMNH 334768	Puerto Rico
		–	Argentina: Buenos Aires
Dives atroviolaceus	Cuban Blackbird	FMNH 375251	Cuba: Pinar del Río
Agelaius assimilis	Red-shouldered Blackbird	MNHNCu	Cuba
Agelaius humeralis	Tawny-shouldered Blackbird	–	Cuba
Quiscalus lugubris lugubris	Carib Grackle	FMNH 339797	Venezuela: Falcón
		USNM 627469	Guyana: Mahaica-Berbice
Quiscalus lugubris contrusus	Carib Grackle	USNM 612608	St. Vincent
		STRI SV-QLU2125	St. Vincent
Quiscalus palustris	Slender-billed Grackle	USNM 194170	Mexico: Estado de México
Quiscalus mexicanus E[c]	Great-tailed Grackle	MBMC JMD1014	USA: Texas
		UWBM 52154	Mexico: Chiapas
Quiscalus major	Boat-tailed Grackle	FMNH 341918	USA: Louisiana
		USNM 626311	USA: Florida
Quiscalus niger	Greater Antillean Grackle	FMNH 331153	Jamaica: Trelawny
Sturnella militaris	Red-breasted Blackbird	FMNH 339777	Venezuela: Falcón
		USNM 625917	Guyana
Sturnella bellicosa	Peruvian Meadowlark	ANSP 178118	Ecuador: Bolívar
Sturnella defilippii	Pampas Meadowlark	AMNH 816591	Argentina: Buenos Aires
Sturnella loyca	Long-tailed Meadowlark	MACN-Or 68357	Argentina: Buenos Aires
		AMNH DOT-13514	Argentina: Río Negro
Sturnella magna	Eastern Meadowlark	FMNH 339780	Venezuela: Falcón
		UMMZ 227823	USA: Louisiana
		ROM SMM 88-1	Canada: Ontario
Sturnella lilianae	Lilian's Meadowlark	FMNH 393903	Mexico: Sonora
		ROM JCB5473	USA: Texas

[a] Following Gill and Donsker (2012) or Jaramillo and Burke (1999).

[b] Specimens with "BB" codes are unvouchered; all other codes are museum catalog numbers. When tissue and skin specimens are housed at different institutions, both are listed. Abbreviations: ANSP = Academy of Natural Sciences of Drexel University; AMNH = American Museum of Natural History; BIOUG = Biodiversity Institute of Ontario; BNT = Bahamas National Trust; CUMV = Cornell University Museum of Vertebrates; FMNH = Field Museum of Natural History; IAvH = Instituto Alexander von Humbolt; ICN = Instituto de Ciencias Naturales, Universidad Nacional de Colombia; INIREB = Instituto de Historia

Natural, San Cristóbal de las Casas, Chiapas, Mexico; KUMNH = University of Kansas Natural History Museum; LACM = Natural History Museum of Los Angeles County; LGEMA = Universidade de Sao Paulo, Departamento de Botanica; LSUMZ = Louisiana State University Museum of Natural Science; MACN = Museo Argentino de Ciencias Naturales, Bernardino Rivadavia; MBMC = Marjorie Barrick Museum of Natural History; MNHNCu = Museo Nacional de Historia Natural Cuba; MMNH = James Ford Bell Museum of Natural History; MPEG = Museu Paraense Emílio Goeldi; MZFC = Museo de Zoología de la Facultad de Ciencias, Universidad Nacional Autónoma de México; NCSM = North Carolina State Museum of Natural Sciences; NRM = Swedish Museum of Natural History; ROM = Royal Ontario Museum; STRI = Smithsonian Tropical Research Institute; UAM = University of Alaska Museum; UMMZ = University of Michigan Museum of Zoology; USNM = Smithsonian Institution, National Museum of Natural History; UWBM = University of Washington, Burke Museum of Natural History and Culture.

[c] Eastern and western lineages of *Quiscalus mexicanus* are treated as separate taxa.

Table 3-2. GenBank accession numbers for DNA sequences used in phylogenetic analyses of New World blackbirds (Icteridae). Loci not collected for a given specimen are marked with a dash. Newly-obtained sequences that have not yet been deposited in GenBank are marked with a double X.

Taxon	Voucher specimen or tissue[b]	mtDNA sample description	GenBank numbers of mtDNA samples	GenBank numbers of nucDNA loci (ACO1-I9, FGB-I5, MB-I2, RAG1)
Outgroups for all analyses:				
Teretistris fernandinae	ANSP-B5548	ND2–ATP6, Cyt b	GU932143, AF382999	GU931929, GU932036, GU932357, KC007804
Icteria virens	UWBM CDS4131	ND2–ATP6, Cyt b	GU932138, AF383028	GU931924, GU932031, GU932352, KC007812
	LSUMZ B3892	12S	AF447236	–
Seiurus aurocapilla	STRI-PRSAU1	ND2–ATP6, Cyt b	GU932043, GU932365	GU931829, GU931936, GU932257, KC007800
	UMMZ 224992	12S	AF447254	–
Oreothlypis gutturalis	LSUMZ B26458	ND2–ATP6, Cyt b	GU932041, GU932363	GU931827, GU931934, GU932255, KC007792
Ingroup taxa used in 46-taxon analyses:				
Psarocolius wagleri	LSUMZ B27280	ND2, Cyt b	AF472394, AF472369	KC007919, KC007643, KC007727, KC007834
Psarocolius decumanus	FMNH 324065	ND2, Cyt b	AF472400, AF472375	XX, XX, XX, XX
	CUMV-Bird 52534	COX1	FJ028159	–
	none	16S	AY283889	–
Psarocolius montezuma	LSUMZ 164424	ND2, Cyt b	AF472403, AF472378	KC007921, KC007645, KC007729, KC007836
Ocyalus latirostris	ANSP 177928	ND2, Cyt b	AF472407, AF472382	KC007920, KC007644, KC007728, KC007835
Cacicus sclateri	ANSP 177931	ND2, ND2–ATP6, Cyt b	AY117746, XX, AY117718	KC007922, KC007646, KC007730, KC007837
Amblycercus holosericeus australis	LSUMZ 98900	ND2, ND2–ATP6, Cyt b	AF472411, XX, AF472386	KC007923, KC007647, KC007731, KC007838
Icterus mesomelas	LSUMZ 109279	mitogenome	JX516068	XX, XX, XX, XX
Icterus cucullatus	BB-BEHB25	–	–	XX, XX, XX, XX
	FMNH 341931	ND2, Cyt b	AF099323, AF099284	–
	UWBM 48323	COX1	DQ433692	–
Icterus parisorum	FMNH 341943	ND2, Cyt b	AF099347, AF089035	KC007924, KC007648, KC007732, KC007839
	FMNH 334367	COX1	DQ432953	–

Nesopsar nigerrimus	FMNH 331150	mitogenome	JX516054	KC007925, KC007649, KC007733, KC007840
Gymnomystax mexicanus	FMNH 339743	mitogenome	JX516075	KC007926, KC007650, KC007734, KC007841
Macroagelaius imthurni	FMNH 339783	mitogenome	JX516073	KC007938, KC007663, KC007747, KC007854
Lampropsar tanagrinus	ANSP 177921	mitogenome	JX516057	KC007937, KC007662, KC007746, KC007853
Gnorimopsar chopi	FMNH 334679	mitogenome	JX516055	KC007935, KC007660, KC007744, KC007851
Curaeus curaeus	AMNH 826156	mitogenome	JX516070	KC007934, KC007659, KC007743, KC007850
Curaeus forbesi	MPEG 72143	ND2–ATP6, Cyt b	XX, XX	XX, XX, XX, XX
Amblyramphus holosericeus	FMNH 334662	mitogenome	JX516063	KC007933, KC007658, KC007742, KC007849
Agelasticus cyanopus	FMNH 334636	mitogenome	JX516076	KC007929, KC007653, KC007737, KC007844
Agelasticus thilius	FMNH 334615	mitogenome	JX516069	XX, XX, XX, XX
Chrysomus icterocephalus	FMNH 339772	mitogenome	JX516060	XX, XX, XX, XX
Xanthopsar flavus	FMNH 330747	mitogenome	JX516065	KC007928, KC007652, KC007736, KC007843
Pseudoleistes guirahuro	FMNH 330795	mitogenome	JX516071	XX, XX, XX, XX
Pseudoleistes virescens	FMNH 330796	mitogenome	JX516066	KC007932, KC007657, KC007741, KC007848
Oreopsar bolivianus	FMNH 334687	mitogenome	JX516058	KC007936, KC007661, KC007745, KC007852
Agelaioides badius	FMNH 330801	mitogenome	JX516074	KC007942, KC007667, KC007751, KC007858
Molothrus rufoaxillaris	FMNH 330805	ND2, ND2–ATP6, Cyt b	AF109961, XX, AF089044	XX, XX, XX, XX
	none	CR	EU199785	–
Molothrus oryzivorus	FMNH 324097	Cyt b	AF089060	XX, XX, XX, XX
	LSUMZ 134021	12S, ND2, ND6–CR	AF407089, AF407046, AF407132	–
	USNM 587829	COX1	JQ175403	–

Molothrus ater	FMNH 350707	–	–	KC007943, KC007668, KC007752, KC007859
	UMMZ none	12S, ND2–COX1, COX2–ATP6	AF447241, AF447291, AF447341	–
	BIOUG:SPP1681-70648	COX1	DQ434680	–
	MBMC jk 96-016	Cyt b	EF529951	–
Dives dives	MBMC 7100	mitogenome	JX516061	KC007939, KC007664, KC007748, KC007855
Dives warczewiczi	LSUMZ 113959	ND2, Cyt b	AF109962, AF089021	XX, XX, XX, XX
Agelaius phoeniceus	BB-96 Tordoff	mitogenome	JX516062	–
	FMNH 341893	–	–	KC007930, KC007654, KC007738, KC007845
Agelaius tricolor	LSUMZ 130833	ND2, Cyt b	AF109949, AF08911	XX, XX, XX, XX
	USNM 632199	COX1	JQ173923	–
Agelaius xanthomus	BB-SML 86-1	ND2, Cyt b	AF109948, AF089012	XX, XX, XX, XX
Euphagus carolinus	FMNH 333317	ND2, Cyt b	AF109950, AF089023	XX, XX, XX, XX
	ROM 1B-3617	COX1	AY666525	–
Euphagus cyanocephalus	FMNH 342000	mitogenome	JX516072	–
	FMNH 341985	–	–	KC007941, KC007666, KC007750, KC007857
Quiscalus quiscula	FMNH 341733	mitogenome	JX516064	KC007940, KC007665, KC007749, KC007856
Quiscalus mexicanus W[c]	FMNH 341975	ND2, Cyt b	AF109954, AF089056	XX, XX, XX, XX
Quiscalus nicaraguensis	MBMC 4375	ND2, Cyt b	FJ389549, FJ389558	XX, XX, XX, XX
Sturnella superciliaris	FMNH 334657	Cyt b	AF089038	KC007846, FJ154707, KC007655, KC007739
	LSUMZ B9630	12S, ND2–COX1, COX2–ATP6	AF447239, AF447289, AF447339	–
	USNM 635873	COX1	JQ176299	–
	NRM 947221	ND3	JN715497	–
Sturnella neglecta	FMNH 341967	ND2, Cyt b	FJ154698, FJ154651	–
	ROM 1B-1038	COX1	EU525529	–
	FMNH 330040	–	–	KC007931, KC007656, KC007740, KC007847
Xanthocephalus xanthocephalus	LSUMZ 126564	ND2, COX1, COX1–ATP6, Cyt b	XX, XX, XX, AF089067	KC007927, KC007651, KC007735, KC007842

Dolichonyx oryzivorus	FMNH 334721	–	–	FJ154706, KC007669, KC007753, KC007860
	UMMZ 234583	12S, ND2–COX1, COX2–ATP6, Cyt b	AF447226, AF447276, AF447326, AF447367	–
	BIOUG: LMA 8101-91770	COX1	DQ434587	_
Additional ingroup taxa included in 119-taxon analyses:				
Clypicterus oseryi	LSUMZ 120394	ND2, Cyt b	AF472408, AF472383	XX, –, XX, –
Psarocolius viridis	USNM 609202	ND2, Cyt b	AY117726, AY117698	_
	USNM 639199	COX1	JQ175997	_
Psarocolius atrovirens	FMNH 324106	ND2, Cyt b	AF472391, AF472366	_
Psarocolius a. angustifrons	LSUMZ 120397	ND2, Cyt b	AF472389, AF472364	_
Psarocolius cassini	ANSP 147013	Cyt b	XX	_
Psarocolius bifasciatus bifasciatus	FMNH 457579	Cyt b, ND2	XX, XX	_
Psarocolius bifasciatus yuracares	FMNH 324076	ND2, Cyt b	AF472404, AF472379	_
Psarocolius guatimozinus	LSUMZ B48620	Cyt b	XX	_
Cacicus cela cela	KUMNH 88289	ND2, COX1, Cyt b	AY117731, JQ174227, AY117703	–
Cacicus cela vitellinus	LSUMZ 163850	ND2, Cyt b	AY117732, AY117704	_
Cacicus haemorrhous	USNM 621068	ND2, Cyt b	AY117733, AY117705	_
	USNM 586489	COX1	JQ174230	_
Cacicus uropygialis uropygialis	LSUMZ B6093	ND2, Cyt b	AY117736, AY117708	_
Cacicus uropygialis microrhynchus	STRI PACUR-PC99	ND2, Cyt b	AY117738, AY117710	_
	USNM 608010	COX1	JQ174233	_
Cacicus uropygialis pacificus	ANSP 182884	ND2, Cyt b	AY117735, AY117707	_
Cacicus chrysopterus	USNM 620761	ND2, Cyt b	AY117740, AY117712	_
	MACN-Or-ct 987	COX1	FJ027255	_
Cacicus chrysonotus chrysonotus	LSUMZ 103278	ND2, Cyt b	AY117745, AY117717	_
Cacicus chrysonotus leucoramphus	ANSP 182883	ND2, Cyt b	AY117743, AY117715	_
Cacicus koepckeae	LSUMZ B48621	Cyt b	XX	_
Cacicus solitarius	FMNH 324089	ND2, Cyt b	AY117747, AY117719	XX, –, XX, –
	MACN-Or-ct 1403	COX1	FJ027264	_
Cacicus melanicterus	UWBM 52185	ND2, Cyt b	AY117749, AY117721	XX, –, XX, –

Amblycercus h. holosericeus	KUMNH 1928	ND2, Cyt b	AY117722, AY117750	_
	USNM 608009	COX1	JQ174007	_
Icterus icterus	LSUMZ B11328	ND2, Cyt b	AF099335, AF099296	_
	LSUMZ B48559	COX1	XX	_
Icterus croconotus	FMNH 324092	ND2, Cyt b	AF099336, AF089031	_
	USNM 632494	COX1	JQ175139	_
Icterus jamacaii	LGEMA 2742	COX1	JN801752	_
Icterus pectoralis	MMNH 42544	ND2, Cyt b	AF099348}, AF099304	_
	KUMNH 109733	COX1	DQ432954	_
Icterus graceannae	ANSP 181810	ND2, Cyt b	AF099329, AF089030	_
Icterus cayanensis	MPEG 40.357	ND2, Cyt b	AF099316, AF089027	_
	USNM 625332	COX1	JQ175135	_
Icterus chrysocephalus	FMNH 339734	ND2, Cyt b	AF099317, AF099279	_
	USNM 625748	COX1	JQ175138	_
Icterus pyrrhopterus	FMNH 334608	ND2, Cyt b	AF099319, AF099280	_
	USNM 614726	COX1	JQ175136	_
Icterus bonana	STRI MA-IBO2	ND2, COX1, ATP8–ATP6, Cyt b	AF109445, AF109429, AF109413, AF099277	_
Icterus laudabilis	STRI SL-ILA4	ND2, COX1, ATP8–ATP6, Cyt b	AF109455, AF109439, AF109423, AF099298	_
Icterus oberi	STRI MO-IOB4	ND2, COX1, ATP8–ATP6, Cyt b	AF109447, AF109431, AF109415, AF099303	_
Icterus dominicensis	AMNH NKK1112	Cyt b , CR	AY216867, AY211217	_
Icterus portoricensis	STRI PR-IDO1	ND2, COX1, ATP8–ATP6, Cyt b	AF109451, AF109435, AF109419, AF099288	_
Icterus melanopsis	MNHNCu 4/8/92	ND2, Cyt b	AF099324, AF099286	_
Icterus northropi	BNT REF024	ND2, Cyt b	AF099325, AF099287	_
Icterus prosthemelas	KUMNH MBR 4510	ND2, COX1, ATP8–ATP6	AF109448, AF109432, AF109416	_
	MMNH 42542	Cyt b	AY211213	_
Icterus spurius	NCSM	ND2	AF099352	_
	USNM 626504	COX1	DQ432955	_
	FMNH 381975	Cyt b, CR	AY211198, AY211230	_

Icterus fuertesi	MMNH 42538	ND2, Cyt b, CR	AF099351, AY211215, AY211219	–
Icterus wagleri	MZFC QRO-216	ND2, Cyt b	AF099353, AF099308	–
Icterus maculialatus	INIREB SRF-387	ND2, Cyt b	AF099340, AF099299	–
Icterus auricapillus	FMNH 261843	Cyt b	XX	–
Icterus chrysater	UWBM 69019	ND2, Cyt b	AF099321, AF099281	–
Icterus graduacauda	LSUMZ B-4023	ND2, Cyt b	AF099330, AF099291	–
Icterus galbula	UMMZ 226382	12S, ND2–COX1, cox2–ATP6	AF447237, AF447287, AF447337	–
	FMNH 350604	Cyt b, CR	AY607656, AY607621	–
	ROM 1B-131	COX1	EU525431	–
Icterus abeillei	MZFC 9657	ND2	AF099311	–
	MZFC keo-48	Cyt b, CR	AY607617, AY607602	–
Icterus bullockii	MBMC jk95-095	ND2, Cyt b	EF529839, EF529950	–
	UWBM 59056	COX1	DQ433689	–
	UWBM 55975	CR	AY611475	–
Icterus pustulatus	UWBM 52129	ND2, Cyt b	AF099349, AF099305	–
	MZFC keo38	CR	AY611477	–
Icterus leucopteryx	FMNH 331145	ND2, COX1, ATP8–ATP6	AF109443, AF109427, AF109411	–
	FMNH 33144	Cyt b	AF089032	–
Icterus auratus	UAM 7222	ND2, Cyt b	AF099312, AF099276	–
Icterus nigrogularis	STRI TR-INI1	ND2, COX1, ATP8–ATP6, Cyt b	AF109456, AF109440, AF109424, AF099302	–
	USNM 627066	COX1	JQ175145	–
Icterus gularis	FMNH	ND2, Cyt b	AF099332, AF099293	–
	UWBM 52191	COX1	DQ433697	–
Macroagelaius subalaris	LACM 40973	Cyt b	XX	–
Hypopyrrhus pyrohypogaster	ICN 33977	ND2, Cyt b	AY572450, AY572451	–
Agelasticus xanthophthalmus	FMNH 324094	mitogenome	JX516059	XX, –, XX, –
Chrysomus ruficapillus	FMNH 330775	mitogenome	JX516056	–
Molothrus aeneus	BB-73 James	mitogenome	JX516067	–

Molothrus bonariensis	LSUMZ 113963	12S, ND2, ND6	AF407090, AF407047, AF407133	–
	MACN-Or-ct 3062	COX1	FJ027842	–
	FMNH 334768	Cyt b	AF089043	–
	none	CR	DQ683553	–
Dives atroviolaceus	FMNH 375251	Cyt b	XX	–
Agelaius assimilis	MNHNCu	Cyt b	AF089004	–
Agelaius humeralis	none	ND2, Cyt b	AF109947, AF089006	–
Quiscalus lugubris lugubris	FMNH 339797	ND2, Cyt b	AF109952, AF089054	–
	USNM 627469	COX1	JQ176090	–
Quiscalus lugubris contrusus	USNM 612608	ND2, COX1, Cyt b	FJ389553, JQ176089, FJ389562	–
	STRI SV-QLU2125	ATP8–ATP6	AF132427	–
Quiscalus palustris	USNM 194170	Cyt b	FJ389557	–
Quiscalus mexicanus E[c]	MBMC JMD1014	ND2, Cyt b	FJ389555, FJ389564	–
	UWBM 52154	COX1	DQ434032	–
Quiscalus major	FMNH 341918	ND2, Cyt b	AF109953, AF089055	–
	USNM 626311	COX1	DQ433156	–
Quiscalus niger	FMNH 331153	ND2, Cyt b	AF109955, AF089057	–
Sturnella militaris	FMNH 339777	ND2, Cyt b	XX, XX	–
	USNM 625917	COX1	JQ176296	–
Sturnella bellicosa	ANSP 178118	ND2, Cyt b	FJ154660, AF089062	FJ154708, –, –, –
Sturnella defilippii	AMNH 816591	Cyt b	XX	–
Sturnella loyca	MACN-Or 68357	ND2, Cyt b	XX, XX	–
	AMNH DOT-13514	COX1, CR	FJ028336, JN417869	–, JN417982, –, –
Sturnella magna	FMNH 339780	Cyt b	AF089063	FJ154709, –, –, –
	UMMZ 227823	12S, ND2–COX1, COX2–ATP6	AF447257, AF447307, AF447357	–
	ROM SMM 88-1	COX1	AY666282	–
Sturnella lilianae	FMNH 393903	ND2, Cyt b	FJ154691, FJ154636	FJ154725, –, –, –
	ROM JCB5473	COX1	AY666267	–

[a] Following Gill and Donsker (2012) or Jaramillo and Burke (1999).
[b] When tissue and skin specimens are housed at different institutions, only one is listed here—see Table 3-1 for more complete specimen information.

[c] Eastern and western lineages of *Quiscalus mexicanus* (see Powell et al. 2008) are treated as separate taxa.

Table 3-3. Characteristics of data subsets used in phylogenetic analyses of DNA sequences of New World blackbirds (Icteridae). Parameter values given here are those estimated for the partitioned analysis of the 119-taxon dataset using the full mitochondrial alignment and four nuclear loci.

Data subset	Positions	Number of positions			Model[a]	Parameter values for ML analysis (r_{AC}, r_{AG}, r_{AT}, r_{CG}, r_{CT}, r_{GT}), (π_A, π_C, π_G, π_T), p_{iv}, α
		Total	Variable	Parsimony informative		
mtDNA 1	Codon 1st positions of ATP6, COX2, Cyt *b*, ND1–6	2958	716	507	TVM+I+G[b]	(0.54, 10.65, 0.70, 0.15, 10.65, 1), (0.30, 0.33, 0.18, 0.18,), 0.681, 0.622
mtDNA 2	Codon 2nd positions (all mitochondrial)	3620	251	138	TrN+I+G[b]	(4.12, 73.47, 1, 4.12, 17.65, 1), (0.19, 0.29, 0.12, 0.40,), 0.535, 0.870
mtDNA 3	Codon 3rd positions of all except ND1-2	3121	2601	2034	GTR+I+G	(0.22, 14.52, 0.58, 0.22, 9.90, 1), (0.43, 0.42, 0.04, 0.10), 1.994, 0.030
mtDNA 4	Codon 3rd positions of ND1-2	672	624	554	TrN+G[c]	(1, 43.95, 1, 1, 27.36, 1), (0.41, 0.37, 0.07, 0.15,), NA, 3.396
mtDNA 5	RNA paired positions; codon 1st positions of COX1 and COX3	2880	270	160	K81+I+G[b]	(1, 22.89, 0.22, 0.22, 22.89, 1), (0.25, 0.25, 0.25, 0.25), 0.627, 0.830
mtDNA 6	RNA unpaired positions; codon 2nd positions of ND6	2160	384	241	TIM+I+G[b]	(1, 8.34, 0.58, 0.58, 12.13, 1), (0.44, 0.23, 0.12, 0.21), 0.558, 0.667
mtDNA 7	Noncoding positions; codon 1st positions of ATP8	1290	390	267	HKY+I+G	(1, 6.11, 1, 1, 6.11, 1), (0.31, 0.30, 0.11, 0.28), 0.611, 0.534
nucDNA 1	ACO1-I9, FGB-I5	1673	425	147	HKY+G	(1, 3.76, 1, 1, 3.76, 1), (0.29, 0.17, 0.21, 0.33), NA, 1.044
nucDNA 2	MB-I2, codon 3rd positions of RAG1	1671	295	112	TrNef+I+G[d]	(1, 5.79, 1, 1, 9.21, 1), (0.25, 0.25, 0.25, 0.25) 0.978, 0.523
nucDNA 3	Codon 1st positions of RAG1	961	57	17	HKY+I	(1, 3.60, 1, 1, 3.60, 1), (0.32, 0.20, 0.30, 0.17), 0.892, NA
nucDNA 4	Codon 2nd positions of RAG1	961	35	14	HKY+I	(1, 8.20, 1, 1, 8.20, 1), (0.36, 0.19, 0.18, 0.27), 0.925, NA

[a] Abbreviations as used in PartitionFinder 1.0.1 (Lanfear et al. 2012)
[b] GTR+I+G model implemented in MrBayes.
[c] GTR+G model implemented in MrBayes.
[d] HKY+I+G model implemented in MrBayes.

Figure 3-1

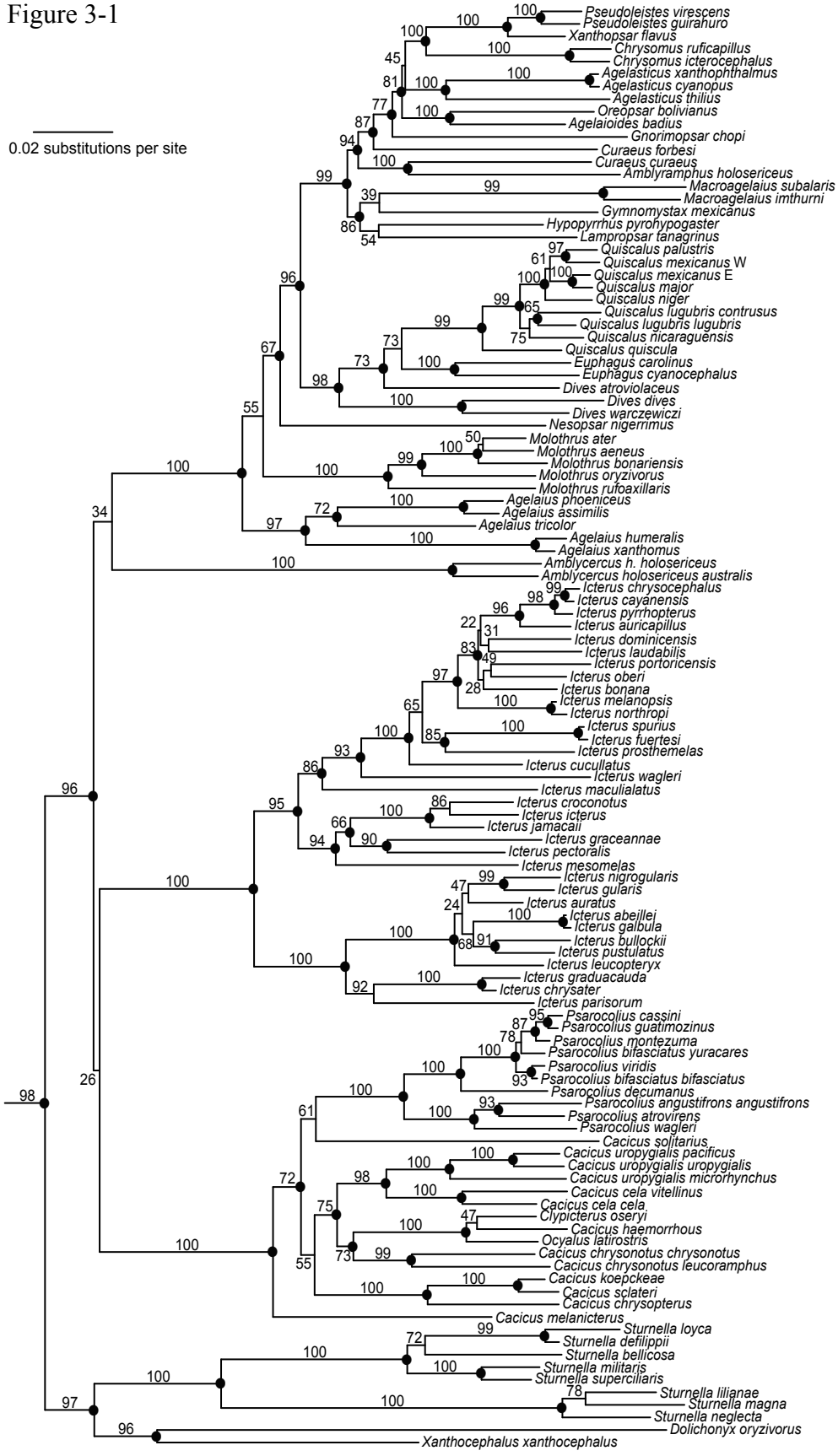
0.02 substitutions per site
Pseudoleistes virescens
Pseudoleistes guirahuro
Xanthopsar flavus
Chrysomus ruficapillus
Chrysomus icterocephalus
Agelasticus xanthophthalmus
Agelasticus cyanopus
Agelasticus thilius
Oreopsar bolivianus
Agelaioides badius
Gnorimopsar chopi
Curaeus forbesi
Curaeus curaeus
Amblyramphus holosericeus
Macroagelaius subalaris
Macroagelaius imthurni
Gymnomystax mexicanus
Hypopyrrhus pyrohypogaster
Lampropsar tanagrinus
Quiscalus palustris
Quiscalus mexicanus W
Quiscalus mexicanus E
Quiscalus major
Quiscalus niger
Quiscalus lugubris contrusus
Quiscalus lugubris lugubris
Quiscalus nicaraguensis
Quiscalus quiscula
Euphagus carolinus
Euphagus cyanocephalus
Dives atroviolaceus
Dives dives
Dives warczewiczi
Nesopsar nigerrimus
Molothrus ater
Molothrus aeneus
Molothrus bonariensis
Molothrus oryzivorus
Molothrus rufoaxillaris
Agelaius phoeniceus
Agelaius assimilis
Agelaius tricolor
Agelaius humeralis
Agelaius xanthomus
Amblycercus h. holosericeus
Amblycercus holosericeus australis
Icterus chrysocephalus
Icterus cayanensis
Icterus pyrrhopterus
Icterus auricapillus
Icterus dominicensis
Icterus laudabilis
Icterus portoricensis
Icterus oberi
Icterus bonana
Icterus melanopsis
Icterus northropi
Icterus spurius
Icterus fuertesi
Icterus prosthemelas
Icterus cucullatus
Icterus wagleri
Icterus maculialatus
Icterus croconotus
Icterus icterus
Icterus jamacaii
Icterus graceannae
Icterus pectoralis
Icterus mesomelas
Icterus nigrogularis
Icterus gularis
Icterus auratus
Icterus abeillei
Icterus galbula
Icterus bullockii
Icterus pustulatus
Icterus leucopteryx
Icterus graduacauda
Icterus chrysater
Icterus parisorum
Psarocolius cassini
Psarocolius guatimozinus
Psarocolius montezuma
Psarocolius bifasciatus yuracares
Psarocolius viridis
Psarocolius bifasciatus bifasciatus
Psarocolius decumanus
Psarocolius angustifrons angustifrons
Psarocolius atrovirens
Psarocolius wagleri
Cacicus solitarius
Cacicus uropygialis pacificus
Cacicus uropygialis uropygialis
Cacicus uropygialis microrhynchus
Cacicus cela vitellinus
Cacicus cela cela
Clypicterus oseryi
Cacicus haemorrhous
Ocyalus latirostris
Cacicus chrysonotus chrysonotus
Cacicus chrysonotus leucoramphus
Cacicus koepckeae
Cacicus sclateri
Cacicus chrysopterus
Cacicus melanicterus
Sturnella loyca
Sturnella defilippii
Sturnella bellicosa
Sturnella militaris
Sturnella superciliaris
Sturnella lilianae
Sturnella magna
Sturnella neglecta
Dolichonyx oryzivorus
Xanthocephalus xanthocephalus

Figure 3-1. Phylogeny of the New World blackbirds (Icteridae) inferred from mitochondrial DNA sequences of 119 taxa (outgroups not shown). The topology shown here is the single best tree ($-\ln L = 112546.18$) found under maximum likelihood (ML). Nonparametric bootstrap percentages from ML analysis appear immediately above or below branches. Filled circles indicate nodes with estimated posterior probabilities of ≥ 0.95 in Bayesian analyses of the same dataset.

Figure 3-2

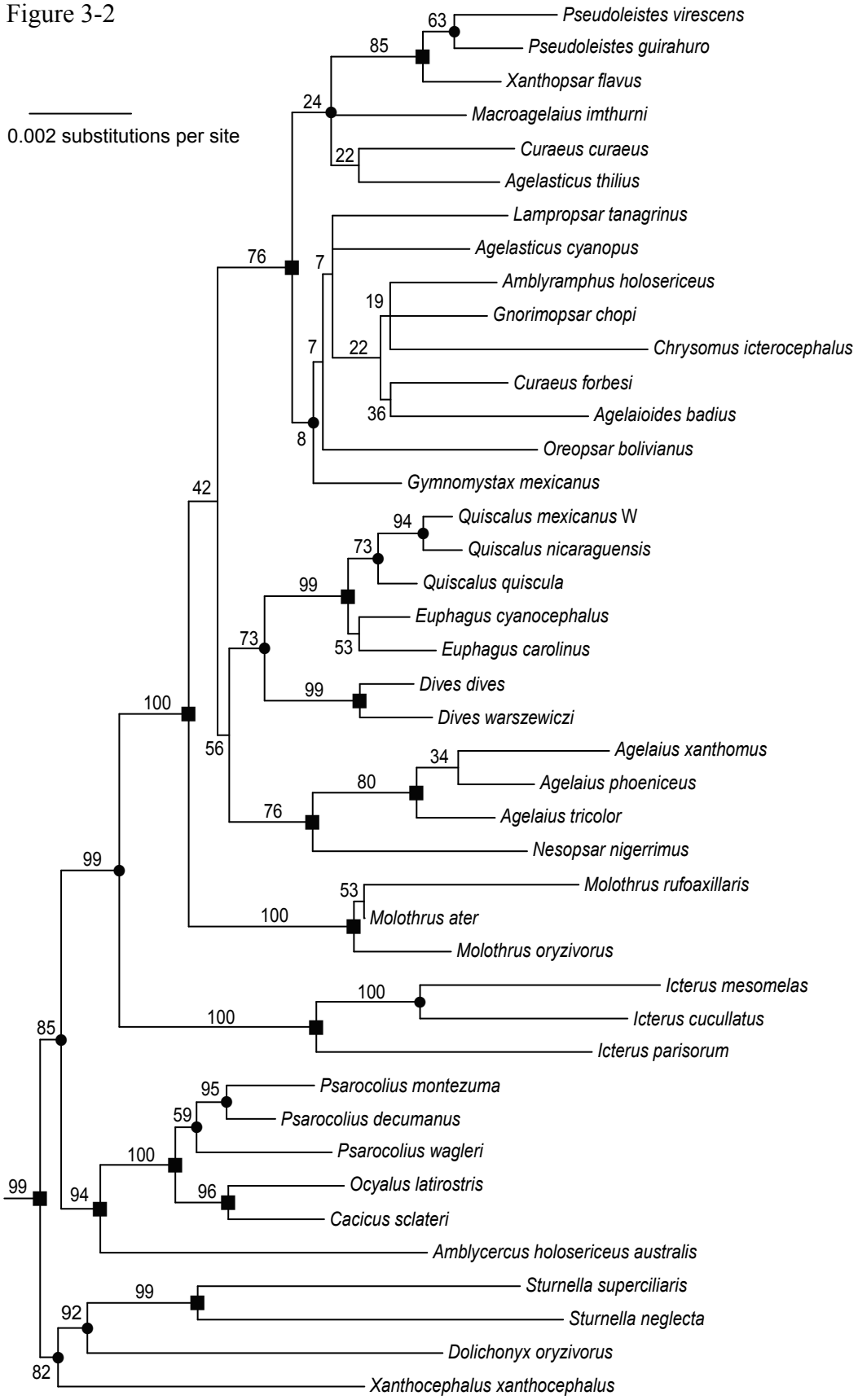
0.002 substitutions per site
Pseudoleistes virescens
Pseudoleistes guirahuro
Xanthopsar flavus
Macroagelaius imthurni
Curaeus curaeus
Agelasticus thilius
Lampropsar tanagrinus
Agelasticus cyanopus
Amblyramphus holosericeus
Gnorimopsar chopi
Chrysomus icterocephalus
Curaeus forbesi
Agelaioides badius
Oreopsar bolivianus
Gymnomystax mexicanus
Quiscalus mexicanus W
Quiscalus nicaraguensis
Quiscalus quiscula
Euphagus cyanocephalus
Euphagus carolinus
Dives dives
Dives warszewiczi
Agelaius xanthomus
Agelaius phoeniceus
Agelaius tricolor
Nesopsar nigerrimus
Molothrus rufoaxillaris
Molothrus ater
Molothrus oryzivorus
Icterus mesomelas
Icterus cucullatus
Icterus parisorum
Psarocolius montezuma
Psarocolius decumanus
Psarocolius wagleri
Ocyalus latirostris
Cacicus sclateri
Amblycercus holosericeus australis
Sturnella superciliaris
Sturnella neglecta
Dolichonyx oryzivorus
Xanthocephalus xanthocephalus

Figure 3-2. Phylogeny of the New World blackbirds (Icteridae) inferred from nuclear DNA sequences of 46 taxa (outgroups not shown). The topology shown here is the single best tree ($-\ln L = 14620.36$) found under maximum likelihood (ML). Nonparametric bootstrap percentages from ML analysis appear immediately above or below branches. Filled circles indicate nodes with estimated posterior probabilities of ≥ 0.95 in Bayesian analyses of the concatenated dataset, and filled squares indicate nodes that also received posterior probability estimates of ≥ 0.95 in species-tree analyses.

Figure 3-3

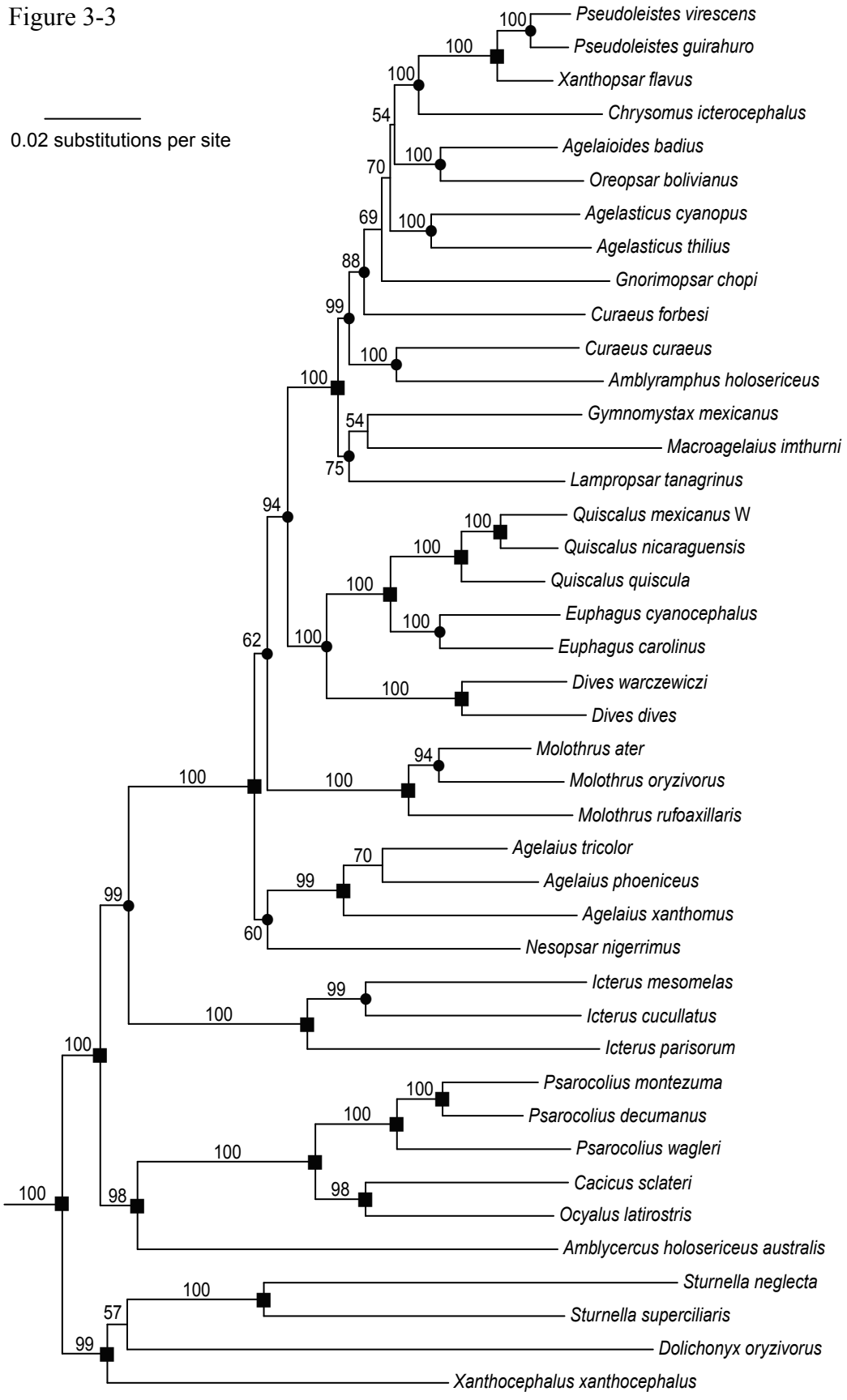
0.02 substitutions per site
Pseudoleistes virescens
Pseudoleistes guirahuro
Xanthopsar flavus
Chrysomus icterocephalus
Agelaioides badius
Oreopsar bolivianus
Agelasticus cyanopus
Agelasticus thilius
Gnorimopsar chopi
Curaeus forbesi
Curaeus curaeus
Amblyramphus holosericeus
Gymnomystax mexicanus
Macroagelaius imthurni
Lampropsar tanagrinus
Quiscalus mexicanus W
Quiscalus nicaraguensis
Quiscalus quiscula
Euphagus cyanocephalus
Euphagus carolinus
Dives warczewiczi
Dives dives
Molothrus ater
Molothrus oryzivorus
Molothrus rufoaxillaris
Agelaius tricolor
Agelaius phoeniceus
Agelaius xanthomus
Nesopsar nigerrimus
Icterus mesomelas
Icterus cucullatus
Icterus parisorum
Psarocolius montezuma
Psarocolius decumanus
Psarocolius wagleri
Cacicus sclateri
Ocyalus latirostris
Amblycercus holosericeus australis
Sturnella neglecta
Sturnella superciliaris
Dolichonyx oryzivorus
Xanthocephalus xanthocephalus

Figure 3-3. Phylogeny of the New World blackbirds (Icteridae) inferred from mitochondrial and nuclear DNA sequences of 46 taxa (outgroups not shown). The topology shown here is the single best tree ($-\ln L = 105577.92$) found under maximum likelihood (ML). Nonparametric bootstrap percentages from ML analysis appear immediately above or below branches. Filled circles indicate nodes with estimated posterior probabilities of ≥ 0.95 in Bayesian analyses of the concatenated dataset, and filled squares indicate nodes that also received posterior probability estimates of ≥ 0.95 in species-tree analyses.

Figure 3-4

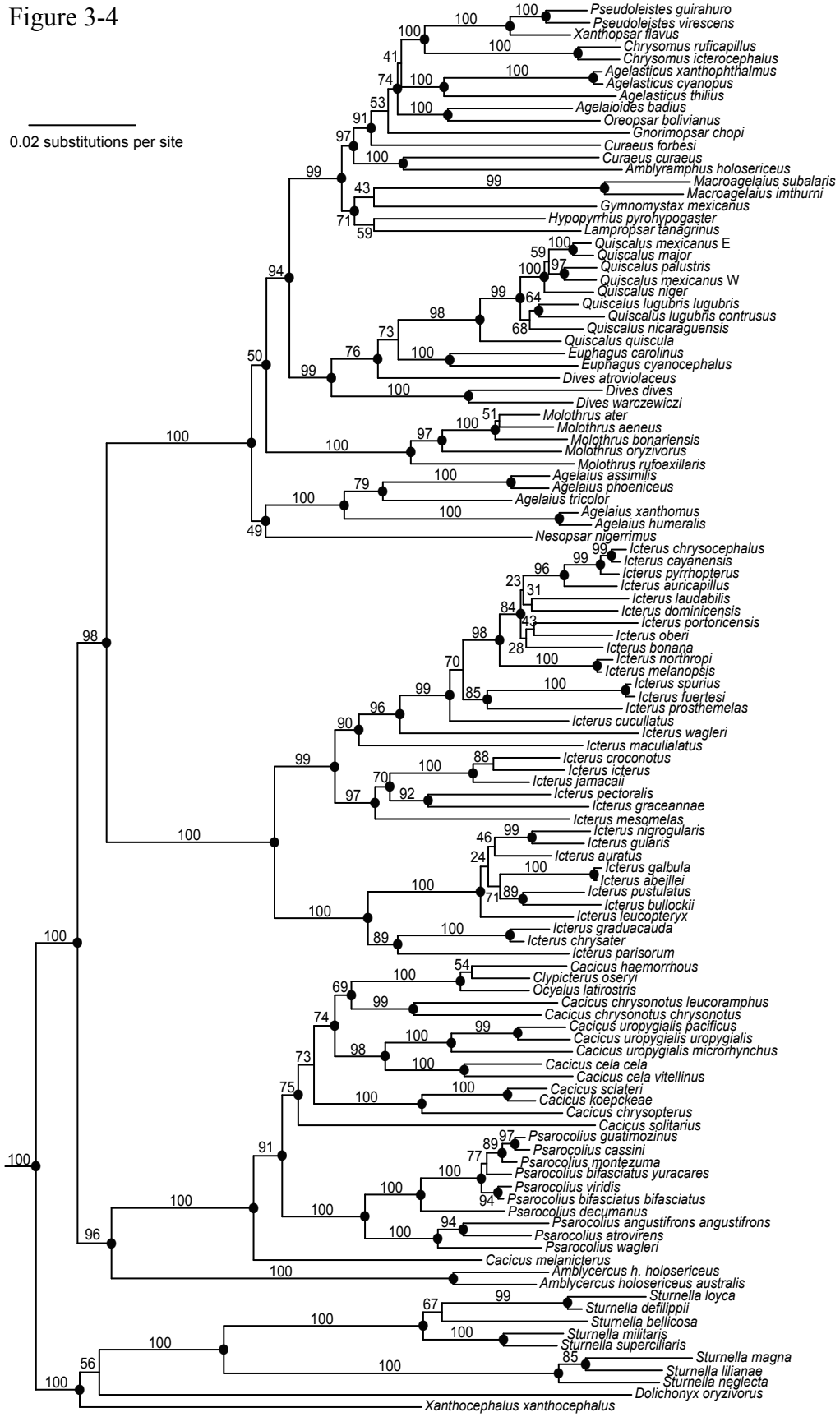
0.02 substitutions per site
Pseudoleistes guirahuro
Pseudolejstes virescens
Xanthopsar flavus
Chrysomus ruficapillus
Chrysomus icterocephalus
Agelasticus xanthophthalmus
Agelasticus cyanopus
Agelasticus thilius
Agelaioides badius
Oreopsar bolivianus
Gnorimopsar chopi
Curaeus forbesi
Curaeus curaeus
Amblyramphus holosericeus
Macroagelaius subalaris
Macroagelaius imthurni
Gymnomystax mexicanus
Hypopyrrhus pyrohypogaster
Lampropsar tanagrinus
Quiscalus mexicanus E
Quiscalus major
Quiscalus palustris
Quiscalus mexicanus W
Quiscalus niger
Quiscalus lugubris lugubris
Quiscalus lugubris contrusus
Quiscalus nicaraguensis
Quiscalus quiscula
Euphagus carolinus
Euphagus cyanocephalus
Dives atroviolaceus
Dives dives
Dives warczewiczi
Molothrus ater
Molothrus aeneus
Molothrus bonariensis
Molothrus oryzivorus
Molothrus rufoaxillaris
Agelaius assimilis
Agelaius phoeniceus
Agelaius tricolor
Agelaius xanthomus
Agelaius humeralis
Nesopsar nigerrimus
Icterus chrysocephalus
Icterus cayanensis
Icterus pyrrhopterus
Icterus auricapillus
Icterus laudabilis
Icterus dominicensis
Icterus portoricensis
Icterus oberi
Icterus bonana
Icterus northropi
Icterus melanopsis
Icterus spurius
Icterus fuertesi
Icterus prosthemelas
Icterus cucullatus
Icterus wagleri
Icterus maculialatus
Icterus croconotus
Icterus icterus
Icterus jamacaii
Icterus pectoralis
Icterus graceannae
Icterus mesomelas
Icterus nigrogularis
Icterus gularis
Icterus auratus
Icterus galbula
Icterus abeillei
Icterus pustulatus
Icterus bullockii
Icterus leucopteryx
Icterus graduacauda
Icterus chrysater
Icterus parisorum
Cacicus haemorrhous
Clypicterus oseryi
Ocyalus latirostris
Cacicus chrysonotus leucoramphus
Cacicus chrysonotus chrysonotus
Cacicus uropygialis pacificus
Cacicus uropygialis uropygialis
Cacicus uropygialis microrhynchus
Cacicus cela cela
Cacicus cela vitellinus
Cacicus sclateri
Cacicus koepckeae
Cacicus chrysopterus
Cacicus solitarius
Psarocolius guatimozinus
Psarocolius cassini
Psarocolius montezuma
Psarocolius bifasciatus yuracares
Psarocolius viridis
Psarocolius bifasciatus bifasciatus
Psarocolius decumanus
Psarocolius angustifrons angustifrons
Psarocolius atrovirens
Psarocolius wagleri
Cacicus melanicterus
Amblycercus h. holosericeus
Amblycercus holosericeus australis
Sturnella loyca
Sturnella defilippii
Sturnella bellicosa
Sturnella militaris
Sturnella superciliaris
Sturnella magna
Sturnella lilianae
Sturnella neglecta
Dolichonyx oryzivorus
Xanthocephalus xanthocephalus

Figure 3-4. Phylogeny of the New World blackbirds (Icteridae) inferred from mitochondrial and nuclear DNA sequences of 119 taxa (outgroups not shown). The topology shown here is the single best tree ($-\ln L = 127734.65$) found under maximum likelihood (ML). Nonparametric bootstrap percentages from ML analysis appear immediately above or below branches. Filled circles indicate nodes with estimated posterior probabilities of ≥ 0.95 in Bayesian analyses of the same

LITERATURE CITED

Abbott, C. L., M. C. Double, J. W. H. Trueman, A. Robinson, and A. Cockburn. 2005. An unusual source of apparent mitochondrial heteroplasmy: duplicate mitochondrial control regions in Thalassarche albatrosses. Molecular Ecology 14:3605–3613.

Allen, E. S. and K. E. Omland. 2003. Novel intron phylogeny supports plumage convergence in orioles (Icterus). Auk 120:961–969.

Alström, P., U. Olsson, F. Lei, H. Wang, W. Gao, and P. Sundberg. 2008. Phylogeny and classification of the Old World Emberizinae (Aves, Passeriformes). Molecular Phylogenetics and Evolution 47:960–973.

American Ornithologists' Union. 1983. Check-list of North American Birds. 6th ed. American Ornithologists' Union, Washington, DC.

American Ornithologists' Union. 1998. Check-list of North American Birds. 7th ed. American Ornithologists' Union, Washington, DC.

Anderson, S., M. H. L. de Bruijn, A. R. Coulson, I. C. Eperon, F. Sanger, and I. G. Young. 1982. Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. Journal of Molecular Biology 156:683–717.

Avise, J. C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology 7:371–379.

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. Auk 105:516–528.

Ballard, J. W. O. and M. C. Whitlock. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13:729–744.

Barker, F. K. 2004. Monophyly and relationships of wrens (Aves: Troglodytidae): a congruence analysis of heterogeneous mitochondrial and nuclear DNA sequence data. Molecular Phylogenetics and Evolution 31:486–504.

Barker, F. K., G. F. Barrowclough, and J. G. Groth. 2002. A phylogenetic hypothesis for passerine birds: taxonomic and biogeographic implications of an analysis of nuclear DNA sequence data. Proceedings of the Royal Society of London Series B-Biological Sciences 269:295–308.

Barker, F. K., K. J. Burns, J. Klicka, S. M. Lanyon, and I. J. Lovette. *in press*. Going to extremes: contrasting rates of diversification in a recent radiation of New World passerine birds. Systematic Biology: first published online 9 December 2012, doi:10. 1093/sysbio/sys094

Barker, F. K., A. J. Vandergon, and S. M. Lanyon. 2008. Assessment of species limits among yellow-breasted meadowlarks (*Sturnella* spp.) using mitochondrial and sex-linked markers. Auk 125:869–879.

Beecher, W. J. 1950. Convergent evolution in the American orioles. Wilson Bulletin 62:51–86.

Beecher, W. J. 1951. Adaptations for food-getting in the American blackbirds. Auk 68:411–440.

Berlin, S., N. G. C. Smith, and H. Ellegren. 2004. Do avian mitochondria recombine? Journal of Molecular Evolution 58:163–167.

Björklund, M. 1991. Evolution, phylogeny, sexual dimorphism and mating system in the grackles (*Quiscalus* spp.: Icterinae). Evolution 45:608–621.

Blake, E. R. 1968. Family Icteridae: American orioles and blackbirds, p. 138–202. *In* R. A. Paynter Jr. [ed.], Check-list of birds of the world: a continuation of the work of James L. Peters. Museum of Comparative Zoology, Cambridge, MA.

Bond, J. 1950. Check-list of Birds of the West Indies. 3rd ed. Academy of Natural Sciences of Philadelphia, PA.

Braun, E. L. and R. T. Kimball. 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. Systematic Biology 51:614–625.

Broughton, R. E., J. E. Milam, and B. A. Roe. 2001. The complete sequence of the zebrafish (Danio rerio) mitochondrial genome and evolutionary patterns in vertebrate mitochondrial DNA. Genome Research 11:1958–1967.

Broughton, R. E. and P. C. Reneau. 2006. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Molecular Biology and Evolution 23:1516–1524.

Burk, A., E. J. P. Douzery, and M. S. Springer. 2002. The secondary structure of mammalian mitochondrial 16S rRNA molecules: refinements based on a comparative phylogenetic approach. Journal of Mammalian Evolution 9:225–252.
Burns, K. J. 1997. Molecular systematics of tanagers (Thraupinae): Evolution and biogeography of a diverse radiation of Neotropical birds. Molecular Phylogenetics and Evolution 8:334–348.
Burns, K. J., S. J. Hackett, and N. K. Klein. 2002. Phylogenetic relationships and morphological diversity in Darwin's finches and their relatives. Evolution 56:1240–1252.
Burns, K. J., S. J. Hackett, and N. K. Klein. 2003. Phylogenetic relationships of Neotropical honeycreepers and the evolution of feeding morphology. Journal of Avian Biology 34:360–370.
Cadena, C. D., A. M. Cuervo, and S. M. Lanyon. 2004. Phylogenetic relationships of the Red-bellied Grackle (Icteridae: *Hypopyrrhus pyrohypogaster*) inferred from mitochondrial DNA sequence data. Condor 106:664–670.
Cannone, J. J., S. Subramanian, M. N. Schnare, J. R. Collett, L. M. D'Souza, Y. Du, B. Feng, N. Lin, L. V. Madabusi, K. M. Müller, N. Pande, Z. Shang, N. Yu, and R. R. Gutell. 2002. The Comparative RNA Web (CRW) Site: An Online Database of Comparative Sequence and Structure Information for Ribosomal, Intron, and Other RNAs. BioMed Central Bioinformatics 3:2 [Erratum: BioMed Central Bioinformatics 3:15]. Available at <http://www. rna. ccbb. utexas. edu> (Accessed 1 March 2010).
Cardiff, S. W. and J. V. Remsen, Jr. 1994. Type specimens of birds in the Museum of Natural Sciences, Lousiana State University. Occasional Papers of the Museum of Natural Sciences 68:1–33.
Castresana, J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Molecular Biology and Evolution 17:540–552.
Cameron, S. L., C. L. Lambkin, S. C. Barker, and M. F. Whiting. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology 32:40–59.
Cerasale, D. J., R. Dor, D. W. Winkler, and I. J. Lovette. 2012. Phylogeny of the Tachycineta genus of New World swallows: insights from complete mitochondrial genomes. Molecular Phylogenetics and Evolution 63, 64–71.
Christensen, A. F. 2000. The fifteenth- and twentieth-century colonization of the Basin of Mexico by the Great-tailed Grackle (*Quiscalus mexicanus*). Global Ecology and Biogeography 9:415–420.
Cummings, M. P., S. P. Otto, and J. Wakeley. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. Molecular Biology and Evolution 12:814–822.
DaCosta, J. M., W. Wehtje, and J. Klicka. 2008. Historic genetic structuring and paraphyly within the Great-tailed Grackle. Condor 110:170–177.
Darriba, D., G. L. Taboada, R. Doallo, and D. Posada. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nature Methods 9:772.
Desjardins, P. and R. Morais. 1990. Sequence and gene organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. Journal of Molecular Biology 212:599–634.
Dickerman, R. W. 1965. The juvenal plumage and distribution of *Cassidix palustris* (Swainson). Auk 82:268–270.
Drovetski, S. V., R. M. Zink, I. V. Fadeev, E. V. Nesterov, E. A. Koblik, Y. A. Red'kin, and S. Rohwer. 2004. Mitochondrial phylogeny of *Locustella* and related genera. Journal of Avian Biology 35:105–110.
Dugand, A., and E. Eisenmann. 1983. Rediscovery of, and new data on, *Molothrus armenti* Cabanis. Auk 100:991–992.
D'Horta, F. M., J. M. Cardosa da Silva, and C. C. Ribas. 2008. Species limits and hybridization zones in *Icterus cayanensis–chrysocephalus* group (Aves: Icteridae). Biological Journal of the Linnean Society 95:583–597.
Dickinson, E. C. (Ed). 2003. The Howard and Moore complete checklist of the birds of the world. Princeton University Press, Princeton, New Jersey.
Drummond, A. J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., Sturrock, S., Thierer, T., Wilson, A. 2010. Geneious v5. 1, Available at <http://www. geneious. com>.
Drummond, A. J., M. A. Suchard, D. Xie and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1. 7. Molecular Biology and Evolution 29:1969–1973.

Eaton, M. D. 2006. A phylogenetic perspective on the evolution of chromatic ultraviolet plumage coloration in grackles and allies (Icteridae). Auk 123:211–234.
Eaton, M. D., J. Soberón, and A. T. Peterson. 2008. Phylogenetic perspective on ecological niche evolution in american blackbirds (Family Icteridae). Biological Journal of the Linnean Society 94:869–878.
Edgar, R. C. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32:1792–1797.
Faith, J. J., and D. D. Pollock. 2003. Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomes. Genetics 165:735–745.
Fisher-Reid, M. C., and J. J. Wiens. 2011. What are the consequences of combining nuclear and mitochondrial data for phylogenetic analysis? Lessons from Plethodon salamanders and 13 other vertebrate clades. BMC Evolutionary Biology 11:300.
Fleischer, R. C., C. E. McIntosh, and C. L. Tarr. 1998. Evolution on a volcanic conveyor belt: using phylogenetic reconstructions and K–Ar based ages of the Hawaiian Islands to estimate molecular evolutionary rates. Molecular Ecology 7:533–545.
Fraga, R. M. 2008. Phylogeny and behavioral evolution in the family Icteridae. Ornitologia Neotropical 19:61–71.
Fraga, R. M. 2011. Family Icteridae (New World blackbirds). In: del Hoyo, J., A. Elliott, and D. A. Christie (Eds). Handbook of the Birds of the World. Vol. 16. Tanagers to New World Blackbirds. Lynx Edicions, Barcelona. pp. 684–807.
Freeman, S. and R. M. Zink. 1995. A phylogenetic study of the blackbirds based on variation in mitochondrial DNA restriction sites. Systematic Biology 44:409–420.
Friedman, N. R., C. M. Hofmann, B. Kondo, and K. E. Omland. 2009. Correlated evolution of migration and sexual dichromatism in the New World Orioles (Icterus). Evolution 63:3269–3274.
Friedman, N. R., L. M. Kiere, and K. E. Omland. 2011. Convergent gains of red carotenoid-based coloration in the New World blackbirds. Auk 128:678–687.
Galtier, N., B. Nabholz, S. Glémin, and G. D. D. Hurst. 2009. Mitochondrial DNA as a marker of molecular diversity: a reappraisal. Molecular Ecology 18:4541–4550.
Gibb, G. C., O. Kardailsky, R. T. Kimball, E. L. Braun, and D. Penny. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Molecular Biology and Evolution 24:269–280.
Gibson, A., V. Gowri-Shankar, P. G. Higgs, and M. Rattray. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Molecular Biology and Evolution 22:251–264.
Gill, F. and D. Donsker (Eds). 2011. IOC World Bird Names (version 2. 9). Available at <http://www. worldbirdnames. org/> (Accessed 29 July 2011).
Gill, F. and D. Donsker (Eds). 2012. IOC World Bird Names (version 3. 2). <http://www. worldbirdnames. org> (accessed 8 December 2012).
Gruber, K. F., R. S. Voss, and S. A. Jansa. 2007. Base-compositional heterogeneity in the RAG1 locus among Didelphid marsupials: implications for phylogenetic inference and the evolution of GC content. Systematic Biology 56:83–96.
Hackett, S. J. 1996. Molecular phylogenetics and biogeography of tanagers in the genus *Ramphocelus* (Aves). Molecular Phylogenetics and Evolution 5:368–382.
Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.
Haemig, P. D. 1978. Aztec emperor Auitzotl and the Great-tailed Grackle. Biotropica 10:11–17.
Härlid, A., A. Janke, and U. Arnason. 1999. The complete mitochondrial genome of Rhea americana and early avian divergences. Journal of Molecular Evolution 46:669–679.
Harrison, G. L., P. A. McLenachan, M. J. Phillips, K. E. Slack, A. Cooper, and D. Penny. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Molecular Biology and Evolution 21:974–983.
Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. Journal of Molecular Evolution 22:160–174.
Hellmayr, C. E. 1937. Catalogue of birds of the Americas and the adjacent islands. Part X. Icteridae. Field Museum of Natural History, Zoological Series 13.

Helm-Bychowski, K., and J. Cracraft. 1993. Recovering phylogenetic signal from DNA sequences: relationships within the corvine assemblage (Class Aves) as inferred from the complete sequences of the mitochondrial DNA cytochrome-*b* gene. Molecular Biology and Evolution 10:1196–1214.

Hofmann, C. M., T. W. Cronin, and K. E. Omland. 2006. Using spectral data to reconstruct evolutionary changes in coloration: carotenoid color evolution in New World orioles. Evolution. 60:1680–1691.

Hofmann, C. M., T. W. Cronin, and K. E. Omland. 2007. Melanin coloration in New World orioles II: ancestral state reconstruction reveals lability in the use of carotenoids and phaeomelanins. Journal of Avian Biology 38:172–181.

Hofmann, C. M., T. W. Cronin, and K. E. Omland. 2008. Evolution of sexual dichromatism. 1. Convergent losses of elaborate female coloration in New World orioles (Icterus spp). Auk 125:778–789.

Hofmann, C. M., T. W. Cronin, and K. E. Omland. 2008. Evolution of sexual dichromatism. 2. Carotenoids and melanins contribute to sexual dichromatism in New World orioles (Icterus spp). Auk 125:790–795.

Huelsenbeck, J. P. and F. Ronquist. 2001. MrBayes: Bayesian inference of phylogeny. Bioinformatics 17:754–755.

Irwin, R. E. 1994. The evolution of plumage dichromatism in the New World blackbirds: social selection on female brightness? American Naturalist 144:890–907.

IUCN [online]. 2008. *Quiscalus nicaraguensis*. 2008 IUCN Red List of Threatened Species. <http://www. iucnredlist. org/details/150421> (31 October 2008).

Jacobsen, F., N. R. Friedman, and K. E. Omland. 2010. Congruence between nuclear and mitochondrial DNA: Combination of multiple nuclear introns resolves a well-supported phylogeny of New World orioles (*Icterus*). Molecular Phylogenetics and Evolution 56:419–427.

Jacobsen, F. and K. E. Omland. 2011. Species tree inference in a recent radiation of orioles (Genus *Icterus*): Multiple markers and methods reveal cytonuclear discordance in the northern oriole group. Molecular Phylogenetics and Evolution 61:460–469.

Jaramillo, A. and P. Burke. 1999. New World Blackbirds: the Icterids. Princeton University Press, Princeton, NJ.

Johnson, K. and B. D. Peer. 2001. Great-tailed Grackle (*Quiscalus mexicanus*). *In* A. Poole and F. Gill [eds.], The birds of North America, No. 576. The Birds of North America, Inc., Philadelphia, PA.

Johnson, K. P. 2001. Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome b sequences. Systematic Biology 50:128–136.

Johnson, K. P. and S. M. Lanyon. 1999. Molecular systematics of the grackles and allies, and the effect of additional sequence (cytochrome *b* and ND2). Auk 116:759–768.

Johnson, K. P. and S. M. Lanyon. 2000. Evolutionary changes in color patches of blackbirds are associated with marsh nesting. Behavioral Ecology 11:515–519.

Kass, R. E. and A. E. Raftery. 1995. Bayes factors. Journal of the American Statistical Association 90:773–795.

Kerr, K. C. R., M. Y. Stoeckle, C. J. Dove, L. A. Weigt, C. M. Francis, and P. D. N. Hebert. 2007. Comprehensive DNA barcode coverage of North American birds. Molecular Ecology Notes 7:535–543.

Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Molecular Phylogenetics and Evolution 4:314–330.

Kjer, K. M. and R. L. Honeycutt. 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. BMC Evolutionary Biology 7:8.

Klein, N. K., K. J. Burns, S. J. Hackett, and C. S. Griffiths. 2004. Molecular phylogenetic relationships among the wood warblers (Parulidae) and historical biogeography in the Caribbean Basin. Journal of Caribbean Ornithology 17:3–17.

Klicka, J., K. Burns, and G. M. Spellman. 2007. Defining a monophyletic Cardinalini: A molecular perspective. Molecular Phylogenetics and Evolution 45:1014–1032.

Klicka, J., K. P. Johnson, and S. M. Lanyon. 2000. New World nine-primaried oscine relationships: constructing a mitochondrial DNA framework. Auk 117:321–336.

Klicka, J. and R. M. Zink. 1997. The importance of recent ice ages in speciation: a failed paradigm. Science 277:1666–1669.

Kondo, B. and K. E. Omland. 2007. Ancestral state reconstruction of migration: multistate analysis reveals rapid changes in New World orioles (*Icterus* spp.). Auk 124:410–419.
Krajewski, C., J. T. Sipiorski, and F. E. Anderson. 2010. Complete mitochondrial genome sequences and the phylogeny of cranes (Gruiformes: Gruidae). Auk 127:440–452.
Krishnan, N. M., H. Seligmann, S. Z. Raina, and D. D. Pollock. 2004. Detecting gradients of asymmetry in site-specific substitutions in mitochondrial genomes. DNA Cell Biology 23:707–714.
Lack, D. 1976. Island biology illustrated by the land birds of Jamaica. University of California Press, Berkeley, CA.
Lanfear, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Molecular Biology and Evolution 29:1695–1701.
Lansman, R. A., R. O. Shade, J. F. Shapira, and J. C. Avise. 1981. The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations. III. Techniques and potential applications. Journal of Molecular Evolution 17:214–226.
Lanyon, S. M. 1992. Interspecific brood parasitism in blackbirds (Icterinae): a phylogenetic perspective. Science 255:77–79.
Lanyon, S. M. 1994. Polyphyly of the blackbird genus *Agelaius* and the importance of assumptions of monophyly in comparative studies. Evolution 48:679–693.
Lanyon, S. M., and J. G. Hall. 1994. Reexamination of barbet monophyly using mitochondrial-DNA sequence data. Auk 111:389–397.
Lanyon, S. M., and K. E. Omland. 1999. A molecular phylogeny of the blackbirds (Icteridae): five lineages revealed by cytochrome-*b* sequence data. Auk 116:629–639.
Lerner, H. R. L., M. Meyer, H. F. James, M. Hofreiter, and R. C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Current Biology 21:1838–1844.
Li, C., G. Lu, and G. Ortí. 2008. Optimal data partitioning and a test case for ray-finned fishes (Actinopterygii) based on ten nuclear loci. Systematic Biology 57:519–539.
Lovette, I. J. and E. Bermingham. 2002. What is a wood-warbler? Molecular characterization of a monophyletic parulidae. Auk 119:695–714.
Lovette, I. J., E. Bermingham, and R. E. Ricklefs. 2001. Mitochondrial DNA phylogeography and the conservation of endangered Lesser Antillean Icterus orioles. Conservation Biology 13:1088–1096.
Lovette, I. J., J. L. Pérez-Emán, J. P. Sullivan, R. C. Banks, I. Fiorentino, S. Córdoba-Córdoba, M. Echeverry-Galvis, F. K. Barker, K. J. Burns, J. Klicka, S. M. Lanyon, and E. Bermingham. 2010. A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). Molecular Phylogenetics and Evolution 57:753–770.
Lovette, I. J., G. Seutin, R. E. Ricklefs, and E. Bermingham. 1999. The assembly of an island fauna by natural invasion: sources and temporal patterns in the avian colonization of Barbados. Biological Invasions 1:33–41.
Lowther, P. E. 2001. New name for the Bolivian Blackbird. Bulletin of the British Ornithologists' Club 121:280–281.
Lowther, P. E., R. Fraga, T. S. Schulenberg, and S. M. Lanyon. 2004. Nomenclatural solution for a polyphyletic *Agelaius*. Bulletin of the British Ornithologists' Club 124:171–175.
Markham, N. R. and M. Zuker. 2005. DINAMelt web server for nucleic acid melting prediction. Nucleic Acids Research 33:W577–W581.
Markham, N. R. and M. Zuker. 2008. UNAFold: software for nucleic acid folding and hybriziation, in: Keith, J. M. (Ed.), Bioinformatics, Volume II. Structure, Function and Applications, No. 453 in Methods in Molecular Biology. Humana Press, New Jersey, pp. 3–31.
Marshall, D. C. 2010. Cryptic failure of partitioned Bayesian phylogenetic analyses: lost in the land of long trees. Sytematic Biology 59:108–117.
Mears, J. A., M. R. Sharma, R. R. Gutell, A. S. McCook, P. E. Richardson, T. R. Caulfield, R. K. Agrawal, and S. C. Harvey. 2006. A structural model for the large subunit of the mammalian mitochondrial ribosome. Journal of Molecular Biology 358:193–212.
Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In: Proceedings of the Gateway Computing Environments Workshop

(GCE 2010), 14 November 2010, New Orleans, Louisiana. Institute of Electrical and Electronics Engineers. pp. 1–8.

Mindell, D. P., M. D. Sorenson, and D. E. Dimcheff, 1998. Multiple independent origins of mitochondrial gene order in birds. Proceedings of the National Academy of Sciences 95:10693–10697.

Mindell, D. P., M. D. Sorenson, D. E. Dimcheff, M. Hasegawa, J. C. Ast, and T. Yuri. 1999. Interordinal relationships of birds and other reptiles based on whole mitochondrial genomes. Systematic Biology 48:138–152.

Mindell, D. P., M. D. Sorenson, C. J. Huddleston, H. C. Miranda, A. Knight, S. J. Sawchuk, and T. Yuri. 1997. Phylogenetic relationships among and within select avian orders based on mitochondrial DNA, in: Mindell, D. P. (Ed.), Avian molecular evolution and systematics, Academic Press, San Diego, pp. 213–247.

Minin, V., Z. Abdo, P. Joyce, and J. Sullivan. 2003. Performance-based selection of likelihood models for phylogeny estimation. Systematic Biology 52:674–683.

Misof, B. and K. Misof. 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. Systematic Biology 58:21–34.

Moore, W. S. 1995. Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726.

Morgan-Richards, M., S. A. Trewick, A. Bartosch-Haerlid, O. Kardailsky, M. J. Phillips, P. A. McLenachan, and D. Penny. 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC Evolutionary Biology 8:20.

Mossman, J. A., T. R. Birkhead, and J. Slate, 2006. The whole mitochondrial genome sequence of the zebra finch (*Taeniopygia guttata*). Molecular Ecology Notes 6:1222–1227.

Mulcahy, D. G. and J. R. Macey. 2009. Vicariance and dispersal form a ring distribution in nightsnakes around the Gulf of California. Molecular Phylogenetics and Evolution 53:537–546.

Nabholz, B., E. D. Jarvis, and H. Ellegren, 2010. Obtaining mtDNA genomes from next-generation transcriptome sequencing: a case study on the basal Passerida (Aves: Passeriformes) phylogeny. Molecular Phylogenetics and Evolution 57:466–470.

Nylander, J. A. A., F. Ronquist, J. P. Huelsenbeck, and J. L. Nieves Aldrey. 2004. Bayesian phylogenetic analysis of combined data. Systematic Biology 53:47–67.

Okajima, Y. and Y. Kumazawa. 2009. Mitogenomic perspectives into iguanid phylogeny and biogeography: Gondwanan vicarience for the origin of Madagascan oplurines. Gene 441:28–35.

Omland, K. E. and S. M. Lanyon. 2000. Reconstructing plumage evolution in orioles (*Icterus*): repeated convergence and reversal in patterns. Evolution 54:2119–2133.

Omland, K. E., S. M. Lanyon, and S. J. Fritz. 1999. A molecular phylogeny of the New World orioles (*Icterus*): the importance of dense taxon sampling. Molecular Phylogenetics and Evolution 12:224–239.

Orians, G. H. 1985. Blackbirds of the Americas. University of Washington Press, Seattle.

Pacheco, M. A., F. U. Battistuzzi, M. Lentino, R. F. Aguilar, S. Kumar, and A. A. Escalante. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. Molecular Biology and Evolution 28:1927–1942.

Parker, T. A., III, and J. V. Remsen, Jr. 1987. Fifty-two Amazonian bird species new to Bolivia. Bulletin of the British Ornithologists' Club 107:94–107.

Paton, T., O. Haddrath, and A. J. Baker. 2002. Complete mitochondrial DNA genome sequences show that modern birds are not descended from transitional shorebirds. Proceedings of the Royal Society B: Biological Sciences 269:839–846.

Paxinos, E. E., H. F. James, S. L. Olson, M. D. Sorenson, J. Jackson, and R. C. Fleischer. 2002. mtDNA from fossils reveals a radiation of Hawaiian geese recently derived from the Canada goose (*Branta canadensis*). Proceedings of the National Academy of Sciences USA 99:1399–1404.

Peters, J. L. 1921. A review of the grackles of the genus *Holoquiscalus*. Auk 38:435–453.

Peterson, A. T. 1998. The distribution and type locality of the extinct Slender-billed Grackle *Quiscalus palustris*. Bulletin of the British Ornithologists' Club 118:119–212.

Peterson, A. T. and A. G. Navarro-Sigüenza. 2006. Hundred-year changes in the avifauna of the Valley of Mexico, Distrito Federal, Mexico. Huitzil 7:4–14.

Phillips, A. R. 1950. The Great-tailed Grackles of the Southwest. Condor 52:78–81.

Phillips, M. J. and D. Penny. 2003. The root of the mammalian tree inferred from whole mitochondrial genomes. Molecular Phylogenetics and Evolution 28:171–185.
Phillips, M. J., P. A. McLenachan, C. Down, G. C. Gibb, and D. Penny. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. Systematic Biology 55:122–137.
Phillips, M. J., G. C. Gibb, E. A. Crimp, and D. Penny. 2010. Tinamous and moa flock together: mitochondrial genome sequence analysis reveals independent losses of flight among ratites. Systematic Biology 59:90–107.
Podnar, M., W. Pinsker, and W. Mayer. 2009. Complete mitochondrial genomes of three lizard species and the systematic position of the Lacertidae (Squamata). Journal of Zoological Systematics and Evolutionary Research 47:35–41.
Posada, D. 2008. jModelTest: Phylogenetic Model Averaging. Molecular Biology and Evolution 25, 1253–1256.
Post, W., J. P. Poston, and G. T. Bancroft. 1996. Boat-tailed Grackle (*Quiscalus major*). *In* A. Poole and F. Gill [eds.], The birds of North America, No. 207. The Academy of Natural Sciences, Philadelphia, PA, and The American Ornithologists' Union, Washington, DC.
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. Condor 110:718–728.
Powell, A. F. L. A., F. K. Barker, and S. M. Lanyon. 2013. Empirical evaluation of partitioning schemes for phylogenetic analyses of mitogenomic data: an avian case study. Molecular Phylogenetics and Evolution 66:69–79.
Pratt, R. C., G. C. Gibb, M. Morgan-Richards, M. J. Phillips, M. D. Hendy, and D. Penny. 2009. Toward resolving deep Neoaves phylogeny: data, signal enhancement, and priors. Molecular Biology and Evolution 26:313–326.
Price, J. J. 2009. Evolution and life-history correlates of female song in the New World blackbirds. Behavioral Ecology 20: 967–977.
Price, J. J., N. R. Friedman, and K. E. Omland. 2007. Song and plumage evolution in the New World Orioles (Icterus) show similar lability and convergence in patterns. Evolution 61: 850–863.
Price, J. J. and S. M. Lanyon. 2002a. A robust phylogeny of the oropendolas: polyphyly revealed by mitochondrial sequence data. Auk 119:335–348.
Price, J. J. and S. M. Lanyon. 2002b. Reconstructing the evolution of complex bird song in the oropendolas. Evolution 56:1514–1529.
Price, J. J. and S. M. Lanyon. 2004a. Patterns of song evolution and sexual selection in the oropendolas and caciques. Behavioral Ecology 15:485–497.
Price, J. J. and S. M. Lanyon. 2004b. Song and molecular data identify congruent but novel affinities of the Green Oropendola (*Psarocolius viridis*). Auk 121:224–229.
Price, J. J., S. M. Lanyon, and K. E. Omland. 2009. Losses of female song with changes from tropical to temperate breeding in the New World blackbirds. Proceedings of the Royal Society B 276:1971–1980.
Price, J. J. and L. M. Whalen. 2009. Plumage evolution in the oropendolas and caciques: different divergence rates in polygynous and monogamous taxa. Evolution 63:2985–2998.
R Development Core Team. 2010. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at <http://www. R-project. org>.
Raikow, R. J. 1978. Appendicular myology and relationships of the New World nine-primaried oscines (Aves: Passeriformes). Bulletin of the Carnegie Museum of Natural History 7:1–43.
Rambaut, A. and A. J. Drummond. 2007. Tracer v1. 4. <http://beast. bio. ed. ac. uk/Tracer>.
Rambaut, A. and A. J. Drummond. 2009. Tracer v1. 5. <http://tree. bio. ed. ac. uk/software/tracer>.
Raposo do Amaral, F., F. H. Sheldon, and A. Wajntal. 2010. Towards an assessment of character interdependence in avian RNA phylogenetics: a general secondary structure model for the avian mitochondrial 16S rRNA. Molecular Phylogenetics and Evolution 56:498–506.
Raina, S. Z., J. J. Faith, T. R. Disotell, H. Seligmann, C-B. Stewart, and D. D. Pollock. 2005. Evolution of base-substitution gradients in primate mitochondrial genomes. Genome Research 15:665–673.
Remsen, J. V., Jr., Cadena, C. D., Jaramillo, A., Nores, M., Pacheco, J. F., Pérez-Emán, J., Robbins, M. B., Stiles, F. G., Stotz, D. F., Zimmer, K. J. Version 29 July 2011. A classification of the bird species

of South America. American Ornithologists' Union. Available at <http://www. museum. lsu. edu/~Remsen/SACCBaseline. html>.

Remsen, J. V., Jr., C. D. Cadena, A. Jaramillo, M. Nores, J. F. Pacheco, J. F. Pérez-Emán, M. B. Robbins, F. G. Stiles, D. F. Stotz, and K. J. Zimmer. Version 7 December 2012. A classification of the bird species of South America. American Ornithologists' Union. <http://www. museum. lsu. edu/~Remsen/SACCBaseline. html>

Reyes, A., M. Y. Yang, M. Bowmaker, and I. J. Holt. 2005. Bidirectional replication initiates at sites throughout the mitochondrial genome of birds. Journal of Biological Chemistry 280:3242–3250.

Ridgely, R. S., T. F. Allnutt, T. Brooks, D. K. McNicol, D. W. Mehlman, B. E. Young, and J. R. Zook. 2007. Digital distribution maps of the birds of the western hemisphere. Version 3. 0. NatureServe, Arlington, VA.

Ridgway, R. 1902. The birds of North and Middle America. Part II. Bulletin of the United States National Museum, No. 50.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Ronquist, F., M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, and J. P. Huelsenbeck. 2012. MrBayes 3. 2: efficient Bayesian phylogenetic inference and model choice across a large model space. Systematic Biology 61:539–542.

Rubinoff, D. and B. S. Holland. 2005. Between two extremes: mitochondrial DNA is neither the panacea nor the nemesis of phylogenetic and taxonomic inference. Systematic Biology 54:952–961.

Schwarz, G. 1978. Estimating the dimension of a model. Annals of Statistics 6:461–464.

Searcy, W. A., K. Yasukawa, and S. M. Lanyon. 1999. Evolution of polygyny in the ancestors of Red-Winged Blackbirds. Auk 116:5–19.

Selander, R. K. and D. R. Giller. 1961. Analysis of sympatry of Great-tailed and Boat-tailed Grackles. Condor 63:29–86.

Shawkey, M. D., M. E. Hauber, L. K. Estep, and G. E. Hill. 2006. Evolutionary transitions and mechanisms of matte and iridescent plumage coloration in grackles and allies (Icteridae). Journal of the Royal Society Interface 3:777–786.

Short, L. L., Jr., 1968. Sympatry of red-breasted meadowlarks in Argentina, and the taxonomy of meadowlarks (Aves: *Leistes*, *Pezites*, and *Sturnella*). American Museum Novitates 2349, 30 pp.

Short, L. L. and K. C. Parkes. 1979. The status of *Agelaius forbesi* Sclater. Auk 96:179–183.

Sibley, C. G. and B. L. Monroe Jr. 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven, CT.

Slack, K. E., F. Delsuc, P. A. McLenachan, U. Arnason, and D. Penny. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Molecular Phylogenetics and Evolution 42:1–13.

Sorenson, M. D., J. C. Ast, D. E. Dimcheff, T. Yuri, and D. P. Mindell. 1999. Primers for a PCR-based approach to mitochondrial genome sequencing in birds and other vertebrates. Molecular Phylogenetics and Evolution 12:105–114.

Sturge, R. J., F. Jacobsen, B. B. Rosensteel, R. J. Neale, and K. E. Omland. 2009. Colonization of South America from Caribbean islands confirmed by molecular phylogeny with increased taxon sampling. Condor 111:575–579.

Swofford, D. L. 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer, Sunderland, MA.

Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures on Mathematics in the Life Sciences 17:57–86.

Timura, K. and M. Nei. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Molecular Biology and Evolution 10:512–526.

Tordoff, H. B. 1954. Relationships in the New World nine-primaried oscines. Auk 71:273–284.

Webster, M. S. 1992. Sexual dimorphism, mating system and body size in New World Blackbirds. Evolution 46:1621–1641.

Webster, J. D. 2003. Skeletal characters and the genera of blackbirds (Icteridae). Condor 105:239–257.

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.

Weir, J. T. and D. Schluter. 2008. Calibrating the avian molecular clock. Molecular Ecology 17:2321–2328.

Werle, E., C. Schneider, M. Renner, M. Völker, and W. Fiehn. 1994. Convenient single-step, one tube purification of PCR products for direct sequencing. Nucleic Acids Research 22:4354–4355.

Wilgenbusch, J. C., D. L. Warren, and D. L. Swofford. 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference. <http://ceb. csit. fsu. edu/awty>.

Yang, Z. 1996. Maximum-likelihood models for combined analyses of multiple sequence data. Journal of Molecular Evolution 42:587–596.

Yang, Z. 2007. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Molecular Biology and Evolution 24:1586–1591.

Yang, S. Y. and R. K. Selander. 1968. Hybridization in the grackle *Quiscalus quiscula* in Louisiana. Systematic Zoology 17:107–143.

Yuri, T. and D. P. Mindell. 2002. Molecular phylogenetic analysis of Fringillidae, "New World nine-primaried oscines" (Aves: Passeriformes). Molecular Phylogenetics and Evolution 23:229–243.

Zink, R. M., W. L. Rootes, and D. L. Dittman. 1991. Mitochondrial DNA variation, population structure, and evolution of the Common Grackle (*Quiscalus quiscula*). Condor 93:318–329.

Zwickl, D. J. 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion. Ph. D. dissertation, University of Texas at Austin.

APPENDIX 1

References to publications used to describe methods currently employed in phylogenetic analyses of animal mitochondrial genomes. These publications were examined to determine what data partitioning schemes were applied and which portions of mitogenomic sequences were excluded from analyses. The first set (List A) constitutes a relatively unbiased sample of current practice, whereas the second set (List B) was assembled as a supplement to the first to further review the practices of researchers especially concerned with phylogeny estimation and thus most likely to employ exemplary phylogenetic methods.

A) Accession numbers associated with the 71 studies (one accession selected to reference each study) that included a phylogenetic analysis, of 91 studies published in conjunction with submissions of whole mitochondrial genome sequences of animals to GenBank during 2009–2010 reviewed by the authors.

NC_009629, NC_010197, NC_011816, NC_011826, NC_011921, NC_011944, NC_012043, NC_012051, NC_012060, NC_012218, NC_012308, NC_012312, NC_012323, NC_012366, NC_012374, NC_012376, NC_012387, NC_012420, NC_012421, NC_012430, NC_012453, NC_012566, NC_012573, NC_012618, NC_012647, NC_012681, NC_012694, NC_012705, NC_012727, NC_012739, NC_012761, NC_012768, NC_012821, NC_012827, NC_012837, NC_012845, NC_012889, NC_012928, NC_012975, NC_013063, NC_013070, NC_013074, NC_013136, NC_013186, NC_013239, NC_013240, NC_013246, NC_013249, NC_013270, NC_013272, NC_013435, NC_013436, NC_013479, NC_013480, NC_013555, NC_013561, NC_013568, NC_013582, NC_013635, NC_013706, NC_013723, NC_013806, NC_013848, NC_013863, NC_013975, NC_013993, NC_013996, NC_014051, NC_014058, NC_014180, NC_014182

B) List of additional publications (i. e. none repeated from list A) reporting phylogenetic analyses of mitogenomic data from various animal taxa, assembled opportunistically from searches of publication indices.

Cameron SL, Lambkin CL, Barker SC, Whiting MF. 2007. A mitochondrial genome phylogeny of Diptera: whole genome sequence data accurately resolve relationships over broad timescales with high precision. Systematic Entomology, 32:40–59.

Cerasale, DJ, R Dor, DW Winkler, IJ Lovette. 2012. Phylogeny of the *Tachycineta* genus of New World swallows: insights from complete mitochondrial genomes. Mol. Phylogenet. Evol. 63:64–71.

Cook CE, Yue QY, Akam M. 2005. Mitochondrial genomes suggest that hexapods and crustaceans are mutually paraphyletic. Proceedings of the Royal Society B-Biological Sciences, 272:1295–1304.

Douglas DA, Arnason U. 2009. Examining the utility of categorical models and alleviating artifacts in phylogenetic reconstruction of the Squamata (Reptilia). Mol. Phylogenet. Evol., 52:784–796.

Douglas DA, Gower DJ. 2010. Snake mitochondrial genomes: phylogenetic relationships and implications of extended taxon sampling for interpretations of mitogenomic evolution. BMC Genomics, 11:14.

Douglas DA, Janke A, Arnason U. 2006. A mitogenomic study on the phylogenetic position of snakes. Zoologica Scripta, 35:545–558.

Gibb GC, Kardailsky O, Kimball RT, Braun EL, Penny D. 2007. Mitochondrial genomes and avian phylogeny: complex characters and resolvability without explosive radiations. Mol. Biol. Evol. 24:269–280.

Gibson A, Gowri-Shankar V, Higgs PG, Rattray M. 2005. A comprehensive analysis of mammalian mitochondrial genome base composition and improved phylogenetic methods. Mol. Biol. Evol. 22:251–264.

Gotzek D, Clarke J, Shoemaker D. 2010. Mitochondrial genome evolution in fire ants (Hymenoptera: Formicidae). BMC Evolutionary Biology, 10:300.

Haddrath O, Baker AJ. 2001. Complete mitochondrial DNA genome sequences of extinct birds: ratite phylogenetics and the vicariance biogeography hypothesis. Proceedings of the Royal Society of London Series B-Biological Sciences, 268:939–945.

Harrison GL, McLenachan PA, Phillips MJ, Slack KE, Cooper A, Penny D. 2004. Four new avian mitochondrial genomes help get to basic evolutionary questions in the late Cretaceous. Mol. Biol. Evol. 21:974–983.

Inoue JG, Miya M, Lam K, Tay BH, Danks JA, Bell J, Walker TI, Venkatesh B. 2010. Evolutionary Origin and Phylogeny of the Modern Holocephalans (Chondrichthyes: Chimaeriformes): A Mitogenomic Perspective. Mol. Biol. Evol. 27:2576–2586.

Jackson JA, Baker CS, Vant M, Steel DJ, Medrano-Gonzalez L, Palumbi SR. 2009. Big and slow: phylogenetic estimates of molecular evolution in baleen whales (suborder Mysticeti). Mol. Biol. Evol. 26:2427–2440.

Jiang ZJ, Castoe TA, Austin CC, Burbrink FT, Herron MD, McGuire JA, Parkinson CL, Pollock DD. 2007. Comparative mitochondrial genomics of snakes: extraordinary substitution rate dynamics and functionality of the duplicate control region. BMC Evolutionary Biology 7:123.

Kjer KM, Honeycutt RL. 2007. Site specific rates of mitochondrial genomes and the phylogeny of eutheria. BMC Evolutionary Biology 7:8.

Krajewski C, Sipiorski JT, Anderson FE. 2010. Complete mitochondrial genome sequences and the phylogeny of cranes (Gruiformes: Gruidae). Auk 127:440–452.

Krause J, Unger T, Noçon A, Malaspinas A-S, Kolokotronis S-O, Stiller M, Soibelzon L, Spriggs H, Dear PH, Briggs AW, Bray SCE, O'Brien SJ, Rabeder G, Matheus P, Cooper A, Slatkin M, Pääbo S, Hofreiter M. 2008. Mitochondrial genomes reveal an explosive radiation of extinct and extant bears near the Miocene-Pliocene boundary. BMC Evol. Biol. 8:220.

Kurabayashi A, Sumida M, Yonekawa H, Glaw F, Vences M, Hasegawa M. 2008. Phylogeny, recombination, and mechanisms of stepwise mitochondrial genome reorganization in mantellid frogs from Madagascar. Mol. Biol. Evol. 25:874–891.

Lavoue S, Miya M, Nishida M. 2010. Mitochondrial phylogenomics of anchovies (family Engraulidae) and recurrent origins of pronounced miniaturization in the order Clupeiformes. Mol. Phylogenet. Evol. 56:480–485.

Lerner, H. R. L., M. Meyer, H. F. James, M. Hofreiter, and R. C. Fleischer. 2011. Multilocus resolution of phylogeny and timescale in the extant adaptive radiation of Hawaiian honeycreepers. Curr. Biol. 21:1838–1844.

Macey JR, Papenfuss TJ, Kuehl JV, Fourcade HM, Boore JL. 2004. Phylogenetic relationships among amphisbaenian reptiles based on complete mitochondrial genomic sequences. Mol. Phylogenet. Evol. 33:22–31.

Masta SE, Longhorn SJ, Boore JL. 2009. Arachnid relationships based on mitochondrial genomes: asymmetric nucleotide and amino acid bias affects phylogenetic analyses. Mol. Phylogenet. Evol. 50:117–128.

Mayden RL, Chen W-J, Bart HL, Doosey MH, Simons AM, Tang KL, Wood RM, Agnew MK, Yang L, Hirt MV, Clements MD, Saitoh K, Sado T, Miya M, Nishida M. 2009. Reconstructing the phylogenetic relationships of the earth's most diverse clade of freshwater fishes-order Cypriniformes (Actinopterygii: Ostariophysi): a case study using multiple nuclear loci and the mitochondrial genome. Mol. Phylogenet. Evol. 51:500–514.

Miya M, Satoh TR, Nishida M. 2005. The phylogenetic position of toadfishes (order Batrachoidiformes) in the higher ray-finned fish as inferred from partitioned Bayesian analysis of 102 whole mitochondrial genome sequences. Biological Journal of the Linnean Society 85:289–306.

Morgan-Richards M, Trewick SA, Bartosch-Haerlid A, Kardailsky O, Phillips MJ, McLenachan PA, Penny D. 2008. Bird evolution: testing the Metaves clade with six new mitochondrial genomes. BMC Evol. Biol. 8:20.

Mueller RL, Macey JR, Jaekel M, Wake DB, Boore JL. 2004. Morphological homoplasy, life history evolution, and historical biogeography of plethodontid salamanders inferred from complete mitochondrial genomes. Proc. Natl. Acad. Sci. U. S. A. 101:13820–13825.

Murata Y, Nikaido M, Sasaki T, Cao Y, Fukumoto Y, Hasegawa M, Okada N. 2003. Afrotherian phylogeny as inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 28:253–260.

Pacheco MA, Battistuzzi FU, Lentino M, Aguilar RF, Kumar S, Escalante AA. 2011. Evolution of modern birds revealed by mitogenomics: timing the radiation and origin of major orders. Mol. Biol. Evol. 28:1927–1942.

Phillips MJ, McLenachan PA, Down C, Gibb GC, Penny D. 2006. Combined mitochondrial and nuclear DNA sequences resolve the interrelations of the major Australasian marsupial radiations. Systematic Biology 55:122–137.

Podnar M, Pinsker W, Mayer W. 2009. Complete mitochondrial genomes of three lizard species and the systematic position of the Lacertidae (Squamata). Journal of Zoological Systematics and Evolutionary Research 47:35–41.

Pratt RC, Gibb GC, Morgan-Richards M, Phillips MJ, Hendy MD, Penny D. 2009. Toward resolving deep neoaves phylogeny: data, signal enhancement, and priors. Mol. Biol. Evol. 26:313–326.

San Mauro D, Gower DJ, Massingham T, Wilkinson M, Zardoya R, Cotton JA. 2009. Experimental design in caecilian systematics: phylogenetic information of mitochondrial genomes and nuclear rag1. Systematic Biology 58:425–438.

Shen YY, Liang L, Sun YB, Yue BS, Yang XJ, Murphy RW, Zhang YP. 2010. A mitogenomic perspective on the ancient, rapid radiation in the Galliformes with an emphasis on the Phasianidae. BMC Evol. Biol. 10:132.

Slack KE, Delsuc F, McLenachan PA, Arnason U, Penny D. 2007. Resolving the root of the avian mitogenomic tree by breaking up long branches. Mol. Phylogenet. Evol. 42:1–13.

Slack KE, Janke A, Penny D, Arnason U. 2003. Two new avian mitochondrial genomes (penguin and goose) and a summary of bird and reptile mitogenomic features. Gene 302:43–52.

Slack KE, Jones CM, Ando T, Harrison GL, Fordyce RE, Arnason U, Penny D. 2006. Early penguin fossils, plus mitochondrial genomes, calibrate avian evolution. Mol. Biol. Evol. 23:1144–1155.

Yamanoue Y, Miya M, Matsuura K, Katoh M, Sakai H, Nishida M. 2008. A new perspective on phylogeny and evolution of tetraodontiform fishes (Pisces : Acanthopterygii) based on whole mitochondrial genome sequences: Basal ecological diversification? BMC Evol. Biol. 8:212.

Yan J, Li HD, Zhou KY. 2008. Evolution of the mitochondrial genome in snakes: gene rearrangements and phylogenetic relationships. BMC Genomics 9:569.

Yu L, Li YW, Ryder OA, Zhang YP. 2007. Analysis of complete mitochondrial genome sequences increases phylogenetic resolution of bears (Ursidae), a mammalian family that experienced rapid speciation. BMC Evol. Biol. 7:198.

Zhang P, Wake DB. 2009. Higher-level salamander relationships and divergence dates inferred from complete mitochondrial genomes. Mol. Phylogenet. Evol. 53:492–508.

APPENDIX 2

Supplemental results and discussion relating to sequencing and characteristics of the mitogenomes of the grackles and allies subfamily of New World blackbirds (Icteridae). Section A2-1 expands upon Results section 2-3. 1. 1, and section A2-2 relates to Results section 2-3. 1. 3.

A2-1. Sequencing and mitogenome organization. —Although amplification and sequencing of mitogenomic fragments was unproblematic for most taxa, we were unable to sequence two successfully amplified sections from *Icterus* and one fragment from *Lampropsar* (Table 1). We found one instance of heteroplasmy for a single position in *Agelaius phoeniceus* (16S, base 1826, alignment position 1840), and possibly also in *Agelaiodes* (control region, base 16587, alignment position 16669) and *Pseudoleistes guirahuro* (control region, base 16245, alignment position 16324). All other samples yielded single invariant sequences with clean fragment overlap. Putative protein-coding genes from all final sequences yielded open reading frames with no internal stop codons, all structural RNA genes transcribed to apparently stable RNA structures with conserved motifs, and variation across the genome alignment was biased toward synonymous, non-coding and unpaired positions; these characteristics suggest a purely mitochondrial origin for the sequences reported here.

Once the mitogenomic sequences were aligned, we found gene overlaps and locations of intergenic spacers to be identical to *Taeniopygia* except that (1) *Quiscalus quiscula* has a 10 bp insertion at the juncture of its $tRNA^{Gln}$ and $tRNA^{Met}$ genes such that they do not overlap as in other taxa and are instead separated by 9 bp, (2) *Oreopsar* has a 3 bp insertion within (what is in other species) the third codon prior to the stop codon of COX2, which adds a leucine when translated, and (3) *Pseudoleistes guirahuro* has a substitution that converts the usual COX2 stop to a glutamine codon; presumably, the thymine that follows, and which is a 1 bp intergenic spacer in other species, serves as the stop codon after posttranscriptional polyadenylation. Most other interspecific variation in start and stop codons (Table S3) resulted from substitutions, but *Euphagus cyanocephalus, Nesopsar*, and *Quiscalus quiscula* have a single thymine insertion in their ND1 stop codon that converts it to TAG from AGG and that frameshifts the third position of the original stop codon to the subsequent intergenic spacer.

A2-2. Heterogeneity among taxa and across genomes. —Base composition of variable sites (n = 3798) differed slightly among taxa, ranging from 38. 9 to 41. 1% for A, 43. 2–46. 1% for C, 3. 5–4. 8% for G, and 9. 6–12. 8% for T. The base composition of *Dives* at variable sites was significantly different from the other species ($\chi^2 = 15.96$, df = 3, $P = 0.03$ after Bonferroni correction), primarily owing to its low T content, though it also shares with *Icterus* the highest C content seen among the sampled taxa. No spatial pattern of base substitution bias was apparent across the mitogenome of *Dives*, as judged from sliding window analyses of inferred changes (not shown). The proportionality of C to T in *Dives* appeared to be reflective of a more general pattern—we compared the percentages of those bases to one another among taxa and found them to be negatively correlated ($r = -0.818$, $P << 0.001$), as too were the percentages of A and G ($r = -0.658$, $P = 0.001$). Differences in base composition among taxa were not correlated with our inferences of phylogenetic relationships.

As reported in Results (section 3. 1. 3), we found no evidence of variation in base composition and evolutionary model parameter estimates across the mitogenomes that might relate to strand-asymmetric replication processes. Note that the strand-displacement model of mitogenome replication has not been confirmed in birds, and that it has been called into question, even for mammals (Bowmaker et al. 2003, Yasukawa et al. 2006, Holt 2010, Pohjoismäki et al. 2010; but see Bogenhagen and Clayton 2003), leaving the mechanistic basis for linear gradients, where they do occur, unknown.

Although we found no evidence for mitogenome-wide compositional gradients, cursory examination of base frequency estimates across the alignment (Fig. S1) suggests constraints on exchangeability, with an inverse relationship between local frequencies of A and C versus G and T. That apparent pattern is corroborated by correlations among parameter estimates from non-overlapping subsamples of our sliding window analyses: $r_{AG} = -0.55$, $r_{AT} = -0.74$, $r_{CG} = -0.72$, $r_{CT} = -0.74$ (all $P < 0.001$, corrected for multiple comparisons). However, although G and T content are positively correlated ($r_{GT} = 0.69$, $P < 0.001$), A and C frequencies vary independently of one another ($r_{AC} = 0.16$, $P = 0.36$).

Transitional biases also appear related to local variation in base frequencies, showing a negative correlation with frequencies of G and T ($r_{G\kappa} = -0.45$, $P < 0.01$; $r_{T\kappa} = -0.58$, $P < 0.001$).

References for Appendix 2

Bogenhagen, D. F., Clayton, D. A. 2003. The mitochondrial DNA replication bubble has not burst. Trends. Biochem. Sci. 28, 357–360.

Bowmaker, M., Yang, M. Y., Yasukawa, T., Reyes, A., Jacobs, H. T., Huberman, J. A., Holt, I. J. 2003. Mammalian mitochondrial DNA replicates bidirectionally from an initiation zone. J. Biol. Chem. 278, 50961–50969.

Holt, I. J. 2009. Mitochondrial DNA replication and repair: all a flap. Trends. Biochem. Sci. 34, 358–365.

Pohjoismäki, J. L. O., Holmes, J. B., Wood, S. R., Yang, M. Y., Yasukawa, T., Reyes, A., Bailey, L. J., Cluett, T. J., Goffart, S., Willcox, S., Rigby, R. E., Jackson, A. P., Spelbrink, J. N., Griffith, J. D., Crouch, R. J., Jacobs, H. T., Holt, I. J. 2010. Mammalian mitochondrial DNA replication intermediates are essentially duplex but contain extensive tracts of RNA/DNA hybrid. Journal of Molecular Biology 397, 1144–1155.

Yasukawa, T., Reyes, A., Cluett, T. J., Yang, M. Y., Bowmaker, M., Jacobs, H. T., Holt, I. J. 2006. Replication of vertebrate mitochondrial DNA entails transient ribonucleotide incorporation throughout the lagging strand. EMBO J. 25, 5358–5371.