

Nuclear locus divergence at the early stages of speciation in the Orchard Oriole complex

Rachel J. Sturge¹, M. Nandadevi Cortés-Rodríguez¹, Octavio R. Rojas-Soto² & Kevin E. Omland¹

¹Department of Biological Sciences, University of Maryland, Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland 21052

²Red de Biología Evolutiva, Instituto de Ecología (INECOL), Carretera antigua a Coatepec 351, Xalapa 91070, Mexico

Keywords

Icterus fuertesi, *Icterus spurius*, mtDNA
paraphyly, Orchard Oriole complex,
phylogeography, recent speciation.

Correspondence

Rachel J. Sturge, Department of Ecology and
Evolutionary Biology, Cornell University,
Ithaca, NY 14853.
E-mail: rjs522@cornell.edu

Funding Information

US National Science Foundation CAREER
Grant (DEB – 0347083) and an National
Science Foundation Phylogenetic Systematics
Grant (DEB – 1119506).

Received: 14 August 2015; Revised: 30
March 2016; Accepted: 1 April 2016

**Ecology and Evolution 2016; 6(13): 4307–
4317**

doi: 10.1002/ece3.2168

Abstract

As two lineages diverge from one another, mitochondrial DNA should evolve fixed differences more rapidly than nuclear DNA due to its smaller effective population size and faster mutation rate. As a consequence, molecular systematists have focused on the criteria of reciprocal monophyly in mitochondrial DNA for delimiting species boundaries. However, mitochondrial gene trees do not necessarily reflect the evolutionary history of the taxa in question, and even mitochondrial loci are not expected to be reciprocally monophyletic when the speciation event happened very recently. The goal of this study was to examine mitochondrial paraphyly within the Orchard Oriole complex, which is composed of *Icterus spurius* (Orchard Oriole) and *Icterus fuertesi* (Fuertes' Oriole). We increased the geographic sampling, added four nuclear loci, and used a range of population genetic and coalescent methods to examine the divergence between the taxa. With increased taxon sampling, we found evidence of clear structure between the taxa for mitochondrial DNA. However, nuclear loci showed little evidence of population structure, indicating a very recent divergence between *I. spurius* and *I. fuertesi*. Another goal was to examine the genetic variation within each taxon to look for evidence of a past founder event within the *I. fuertesi* lineage. Based on the high amounts of genetic variation for all nuclear loci, we found no evidence of such an event – thus, we found no support for the possible founding of *I. fuertesi* through a change in migratory behavior, followed by peripheral isolates speciation. Our results demonstrate that these two taxa are in the earliest stages of speciation, at a point when they have fixed differences in plumage color that are not reflected in monophyly of the mitochondrial or nuclear DNA markers in this study. This very recent divergence makes them ideal for continued studies of species boundaries and the earliest stages of speciation.

Introduction

For the last 25 years, neutral molecular markers, especially mitochondrial DNA (mtDNA), have been used to examine closely related species – elucidating their evolutionary histories and looking for evidence of gene flow. Studies of closely related taxa often address species delimitation by looking for reciprocal monophyly in mitochondrial genes (Moritz 1994; Zink and McKittrick 1995; Funk and Omland 2003; Hebert et al. 2004; Zink and Barrowclough 2008). Mitochondrial DNA has two key advantages that make it useful in this regard: (1) mtDNA has a rapid mutation rate, allowing for increased chances of

accumulating variation; (2) mtDNA also has faster sorting (genetic drift) due to its maternal inheritance and generally lower effective population sizes (ESSs; Avise 1994; Zink and Barrowclough 2008).

However, there are many examples where species delimitation based solely on mtDNA would disagree with boundaries delimited using other characters (Baker et al. 2003; Funk and Omland 2003; Olsson et al. 2005; Omland et al. 2006; Knowles and Carstens 2007; Joseph and Omland 2009). There are several possible reasons for this disagreement. For instance, individual gene trees of closely related taxa can have unpredictable amounts of variation due to stochasticity in mutation, genetic drift,

and sampling error (Rosenberg and Nordborg 2002; Knowles and Carstens 2007). A major factor that limits the utility of mtDNA is that because it is haploid and maternally inherited, different mitochondrial genes are inherited as a group, sharing a single evolutionary history that may or may not accurately represent the evolutionary history of the species of interest (Knowles and Carstens 2007). A second problem with the approach mentioned above is its focus on one criterion – the presence of reciprocal monophyly. Recent studies have shown that a substantial time lag exists between species divergence and the development of reciprocal monophyly in gene trees (Hey 2006; Omland *et al.* 2006; Knowles and Carstens 2007). Neutral genetic changes can diverge slower than traits under selective pressure, which may result in genetically indistinct taxa with fixed differences in phenotypes that play a role in fitness or sexual selection (e.g., Joseph and Omland 2009; Ross 2014). Species delimited by genetic markers thus can disagree with those delimited using other characteristics, such as morphology. Thus, recently diverged species will most likely not have any fixed differences in their mtDNA due to this time lag, as is evidenced by the multiple examples of known gene tree parphyly in Animalia (see the following papers for a detailed explanation and for examples: Edwards and Beerli 2000; Funk and Omland 2003; Joseph and Omland 2009) and the many well-recognized species that lack fixed mtDNA differences (Ross 2014). Our study focuses on a pair of taxa that represent a well-documented case of mitochondrial parphyly: the Orchard Oriole complex.

The Orchard Oriole complex is comprised of two taxa: *Icterus spurius* (Orchard Oriole), which breeds from southeastern Canada, across the eastern United States and into north-central Mexico, and *Icterus fuertesi* (Fuertes' Oriole), which breeds in a geographically restricted region along the coastal lowlands of southern Tamaulipas and Veracruz, Mexico (Howell and Webb 1995). These two taxa are considered separate species by the IOC World Bird List (Gill and Donsker 2015; also Navarro and Peterson 2004). Clements World Checklist (e.g., Clements 2007) formerly considered them separate species, but they were recently reclassified as subspecies within *I. spurius*, which is also followed by the American Ornithologist's Union (Chesser *et al.* 2015). In this article, we follow the taxonomy of IOC (Gill and Donsker 2015). The controversy regarding the separation of Orchard Oriole and Fuertes' Oriole into different species is based on a variety of studies carried out over the last 10 years. The two taxa differ in a number of ways. Adult males show a fixed difference in color (Fig. 1), with no overlap in color variation when their plumage is analyzed through spectrophotometric comparisons (Hofmann *et al.* 2007; Kiere *et al.* 2007). Adult males of the two taxa are easily

distinguished visually in the field (pers. obs.). They also differ in migratory behavior, as *I. spurius* migrates whereas *I. fuertesi* is a short-distance migrant or partial migrant (Jaramillo and Burke 1999; Tobón-Sampedro and Rojas-Soto 2015). These taxa breed at different latitudes, which has resulted in bioclimatic niche differentiation for their breeding distributions (Martin and Omland 2011). Lastly, studies of their vocalizations have found no evidence of song differentiation (Hagemeyer *et al.* 2012), yet a call that appears to play a role in territory defense differs significantly in a number of different measures of frequency, duration, and amplitude (Sturge *et al.* 2016).

In addition, Baker *et al.* (2003) sampled these two taxa across their ranges and sequenced two regions of mtDNA (cytochrome *b* and control region). They found that the two species were paraphyletic in their mtDNA with intermixed haplotypes (haplotypes that are shared and/or show paraphyletic relationships with another species; Joseph and Omland 2009). *Icterus spurius* and *I. fuertesi* show both these patterns of intermixing in their mtDNA. However, the frequencies of different mitochondrial haplotypes differ significantly, indicating little or no mitochondrial gene flow between the two taxa ($F_{ST} = 0.608$, $P < 0.00001$, Baker *et al.* 2003). Baker *et al.* (2003) suggested that retained ancestral polymorphisms could be sufficient to explain this mitochondrial parphyly between the two taxa, but they did not test it with rigorous statistical analyses (i.e., coalescent methods using multiple loci).

The mitochondrial parphyly in the Orchard Oriole complex has two possible explanations, as with other



Figure 1. Adult male *Icterus spurius* (left) and *I. fuertesi* (right). Adult females and yearling males of the two taxa are predominantly olive yellow; the two taxa cannot be distinguished in these plumages. Painting by J. C. Anderton.

cases. First, the split between the two taxa could have happened so recently that the haplotype variation within each of these taxa has not yet coalesced; thus, they could share haplotypes due to retained ancestral polymorphisms. Second, it is possible that gene flow has been occurring since the two taxa diverged, which could explain the observed patterns of haplotype and lineage sharing. In avian species, hybridization and gene introgression are common even among more distantly related species (Peters *et al.* 2007; Price 2008; Rheindt and Edwards 2011), and species divergence can occur even in the face of ongoing gene flow (Hey 2006).

Additionally, the results from Baker *et al.* (2003) suggested that the divergence that led to the formation of these two taxa could have been the result of a founder event within the lineage leading to *I. fuertesi*. This was based on two mitochondrial genes (cytochrome *b* and control region), for which *I. fuertesi* showed lower haplotype diversity than *I. spurius* – which is the predicted pattern if the lineage leading to *I. fuertesi* was founded from a few individuals whose genetic diversity only represented a portion of the available diversity within the ancestral population (Harrison 1991; Baker *et al.* 2003; Funk and Omland 2003). The two taxa within the Orchard Oriole complex exhibit different migratory behavior, as mentioned above – with *I. spurius* migrating long distances while *I. fuertesi* is a short-distance migrant that overwinters in the southern portion of its breeding range, where it likely overlaps with overwintering *I. spurius* (Howell and Webb 1995; Jaramillo and Burke 1999; Kondo and Omland 2007; Tobón-Sampedro and Rojas-Soto 2015). As migratory behavior is highly labile in orioles (Winker 2000; Kondo and Omland 2007), a loss or decrease in migratory behavior within a subset of the ancestral population could have resulted in the founding of the current breeding grounds of *I. fuertesi*, followed by peripheral isolates speciation in allopatry (Mayr 1942, 1963; West-Eberhard 2005). Examining Figure 2, *I. fuertesi* looks like it could be a classic example of peripheral isolates speciation based on its much smaller breeding range, which is geographically isolated from the closest *I. spurius* breeding populations by the Sierra Madre Oriental Mountains. However, the lack of genetic variation within *I. fuertesi* reported by Baker *et al.* (2003) could also have resulted from differences in sampling effort – there were roughly twice as many *I. spurius* individuals as *I. fuertesi* included in their study.

We addressed these questions regarding the mitochondrial paralogy between these taxa, and the potential founder event involving *I. fuertesi* by sequencing multiple loci (four nuclear introns) and through an increase in sample size, adding nine *I. fuertesi* samples. We sequenced multiple nuclear loci to obtain a broader picture of the

genomic divergence between these two taxa. Using multiple loci, we attempted to avoid either stochastic or deterministic processes that might have only affected a part of the genome or one type of marker (e.g., mtDNA; Edwards and Beerli 2000; Peters *et al.* 2008). We also used population coalescent methods, including IM (isolation with migration, Hey and Nielsen 2004; Hey 2006) to examine polyphyly and shared haplotypes to see whether it was possible to differentiate between retained ancestral polymorphisms and gene flow as possible explanations for these patterns. Our goals were to address the following questions: (1) Do nuclear loci also support very recent divergence, or does mtDNA reflect recent introgression between two well-diverged lineages? (2) What is the nature of the nuclear intron variation between these two taxa? Do they support two distinct groups? (3) Was divergence the result of a founder event and peripheral isolates speciation? (4) Is it possible to determine whether these two taxa are currently exchanging genes?

Methods

DNA sequencing

In this study, we had 32 samples already available to us and we collected nine additional *I. fuertesi* samples from multiple localities in Veracruz, Mexico, to increase our sample size. Not all of the old samples would amplify; as a result, we included samples from 25 *I. spurius* and 14 *I. fuertesi* from across their breeding ranges (Fig. 2, Table S1). For each individual, we sequenced a total of five loci: four nuclear introns and one mitochondrial gene. The four nuclear introns were located on four different chromosomes of the zebra finch genome; three were autosomal loci (GADPH11 397 b.p., TGFB2 574 b.p. and RDP2 298 b.p.) and one was on the Z sex chromosome (SLC9 429 b.p.). The fifth locus was the mitochondrial control region (CR – domain I; 344 b.p.). Accession numbers on GenBank for the genes we sequenced are as follows: GADPH11 (KU903008–KU903079), C.R. (KU903080–KU903118), SLC9 (KU903119–KU903152), RDP2 (KU903153–KU903221), and TGFB2 (KU903222–KU903282). For additional information about the five loci, see Table 1.

We extracted DNA from muscle tissue of all samples using a DNeasy Blood and Tissue Kit (QIAGEN, Valencia, CA). PCRs were then performed on each locus using a Gene Amp PCR System 9700 (Applied Biosystems, Foster City, CA), and we verified our product using 1% agarose gels with ethidium bromide. We cleaned PCR product using QIAquick PCR Purification Kits (QIAGEN). The amplified loci were sequenced using ABI's BigDye v. 3 Terminator Cycle Sequencing Kit on the Gene Amp PCR

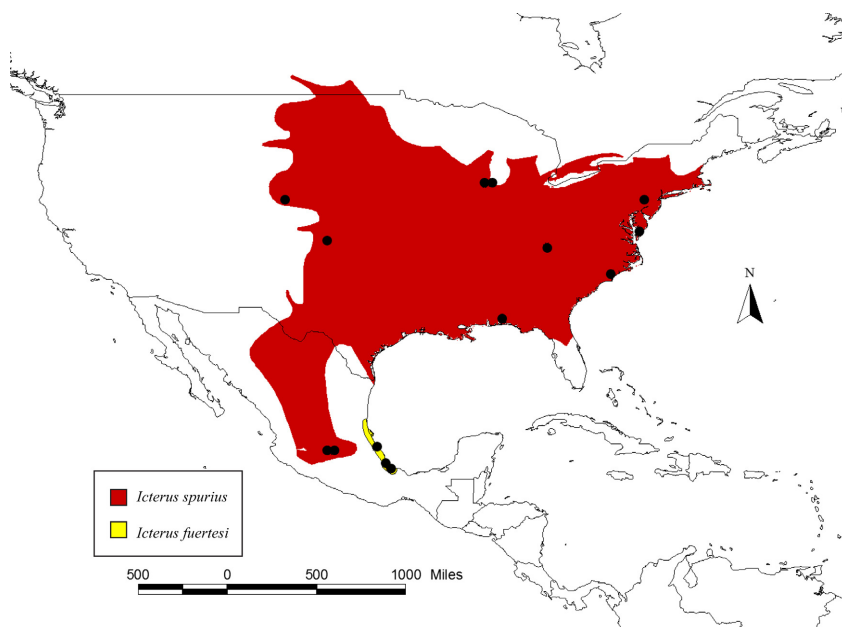


Figure 2. Map showing sampling locations (black dots) for *Icterus spurius* and *I. fuertesi*. All the dots for *I. fuertesi* represent localities where we collected multiple samples. This map also shows the breeding ranges for both species. The map was built using ArcView 3.2 (Environmental Systems Research Institute, Redlands, CA) and the distributional ranges from Nature Serve (www.natureserve.org) and by adding each locality point where the samples were collected.

Table 1. Information on the loci (one mitochondrial locus, three autosomal introns, and one Z-linked intron) used in this study.

Locus	Chromosome	Length	<i>N</i>		Annealing temperature (°C)	Primer source
			<i>Icterus spurius</i>	<i>Icterus fuertesi</i>		
CR-Domain I	Mitochondrial	344	25	16	52	Kondo et al. (2004)
GADPH11	1	397	48	28	55	Primmer et al. (2002)
TGFB2	2	574	44	28	65	Bureš et al. (2002)
RDP2	12	298	44	28	56	Waltari and Edwards (2002)
SLC9	Z chromosome	429	35	21	56	Bandelt et al. (1999)

Chromosome locations were determined using BLAST searches of a reference genome (Zebra Finch). The number of alleles for each taxon are given as *N*.

System 2400 (Applied Biosystems). We used EDTA-ethanol precipitation to remove excess dye terminators. Chromatograms for each sequence were produced using the University of Maryland, Baltimore County's ABI3100 DNA Sequencer (Applied Biosystems). We edited and aligned the chromatograms using Sequencher 4.1 (Gene Codes Corporation, Ann Arbor, MI).

Individuals within our dataset that contained more than one polymorphic site within a locus were phased using PHASE 2.1.1 (Stephens et al. 2001; Stephens and Scheet 2005). We set the burn-in rate to 1000 and the thinning interval to 1. Each run included 10,000 iterations, and we repeated each analysis ten times with different random numbers as the starting seed. We selected the haplotype states suggested by the runs with the highest log-likelihood values and included all haplotypes that

resolved at a probability of at least 0.5 (for justification of this approach and cutoff value, see Jacobsen and Omland 2012). All of the SNP phasing in this study met this cutoff, so no further action was necessary. TOPALi 2.5 (Milne et al. 2004) was used to look for evidence of intralocus recombination – none of our loci had a significant DSS peak and so all were included in our dataset. We constructed and edited haplotype networks within NETWORK v. 4.6.1.1 (Bandelt et al. 1999) using the median joining algorithm.

Population statistics and structure

The number of haplotypes (*h*), the haplotype diversity (*H_d*), the nucleotide diversity (π), the average number of nucleotide differences per site (*p*), and the number of

nucleotide polymorphisms (θ) were all calculated using DnaSP (Rozas et al. 2003; for results, see Table 2). We also used DnaSP to calculate Tajima's D (Tajima 1989), Fu and Li's F^* (Fu and Li 1993), Ramos-Onsins & Roza's neutrality state (R_2 , Ramos-Onsins and Rozas 2002), and mismatch distributions (with associated Tau and Theta values). To look for significant population structure within the nuclear loci, we used PGDSpider to produce a dataset that could be analyzed within Structure by assigning each haplotype a unique number (Lischer and Excoffier 2012). One of the assumptions of Structure is that the different loci included are unlinked. As mentioned previously, our loci were located on different chromosomes of the Zebra Finch genome, so our approach should not violate this assumption. Note that this method does not consider the similarity between different haplotypes when it assigns numerical designations. We analyzed our data using Structure v. 2.3.4 (Pritchard et al. 2000; Hubisz et al. 2009). We used the admixture model with correlated allele frequencies and the LOCPRIOR flag activated, and set both the Dirichlet parameter for degree of admixture (α) and the allelic frequency parameter (k) to be inferred from the dataset, rather than being fixed. All runs had a burn-in of 50,000, followed by 250,000 iterations to collect data to estimate posterior probabilities. We tested K values from 1 to 10, and replicated each run 20 times for all K values.

We used Arelquin v3 3.5.1.2 (Excoffier and Lischer 2010) to run AMOVAs on each locus to test for significant population structure within and between the two taxa (Schneider et al. 2000; Excoffier and Lischer 2010). We calculated Φ_{ST} values in Arelquin as well, which are analogous to Wright's F_{ST} values, but take into account the number of mutations between each haplotype within the sample (Excoffier and Lischer 2010).

Coalescent analyses

To determine whether or not we could estimate how much of the shared DNA variation between these two taxa is due to retained ancestral polymorphisms versus gene flow (either historic or ongoing), we analyzed mtDNA, nuclear DNA, and the combined mtDNA and nuclear DNA datasets using both the programs IM and Ima (isolation with migration, Hey 2006, 2007). IM estimates seven demographic parameters: ESSs of the ancestral (θ_A) and daughter populations (θ_1 and θ_2), migration rates between the two daughter populations (m_1 and m_2), and the time the two daughter populations diverged from each other (t), and the contribution of the ancestral population to each daughter population (splitting parameter, s) (Hey and Nielsen 2004). We used the infinite sites model and set the inheritance scalars based on the modes

Table 2. Estimates of the number of haplotypes (h), the haplotype diversity (Hd), and the nucleotide diversity (π) for the loci included in this study.

Locus	<i>Icterus spurius</i> (Orchard Oriole)			<i>Icterus fuertesi</i> (Fuertes' Oriole)		
	h	Hd	π	H	Hd	π
CR- mtDNA	9	0.597	0.00276	3	0.242	0.01410
GADPH11	14	0.840	0.00531	11	0.807	0.00655
TGFB2	16	0.862	0.00509	12	0.889	0.00441
RDP2	12	0.852	0.00682	7	0.794	0.00507
SLC9	10	0.724	0.00285	8	0.886	0.00433

For each locus, the highest haplotype diversity between the two species is shown in bold.

of inheritance for the loci (autosomal were set to 1, Z linked to 0.75 and mitochondrial to 0.25). In these analyses, we included multiple heated chains (30 chains) and examined the ESS estimates to insure the Markov chains were mixing appropriately (Hey and Nielsen 2004). We also included 1,000,000 generations and sampled every 100 generations so we could monitor output. We completed the run multiple times, using different starting seeds for each run. We also varied the following parameters in an effort to reach convergence: the maximum migration rates for both populations (m_1 and m_2), the maximum time of population splitting (t), the duration of the burn (b) and the duration of the run (l), the heating parameters (g_1 and g_2), the number of chains (n) and the number of chain swap attempts per step (k), and the estimates of maximum ancestral and daughter population sizes (q_A , q_1 , and q_2).

Results

Both taxa were polymorphic with respect to all loci included in this study, as evidenced by the allele networks (Fig. 3). None of the nuclear introns show species monophyly, and there are many shared alleles between taxa. In contrast, the mitochondrial CR is much more segregated between the taxa, with only one shared haplotype and very little intermixing – however, two *I. spurius* haplotypes are only one base pair different than the closest *I. fuertesi* haplotype, while they are either two or three base pairs different than the closest *I. spurius* haplotypes.

Both taxa show high levels of haplotype diversity (Hd) for all of the nuclear loci, ranging from 0.724 (SLC9 for *I. spurius*) to 0.889 (TGFB2 for *I. fuertesi*) (Table 2). However, both show lower levels of diversity for the mitochondrial gene CR, with 0.597 for *I. spurius* and 0.242 for *I. fuertesi*. Note that *I. spurius* has more than two times the haplotype diversity of *I. fuertesi* for CR, yet the taxa show similar amounts of diversity for all four

nuclear loci (Table 2). In terms of nucleotide diversity (π), three of the five loci (CR, TGFB2, and RDP2) have higher nucleotide diversity for *I. spurius* and two (GADPH11 and SLC9) have higher diversity for *I. fuertesi*.

The majority of the loci for both taxa show negative values for Tajima's D and for Fu and Li's F^* , with a few exceptions (Table 3). However, none of these values differed significantly from zero once Bonferroni corrections were applied, indicating that there is no evidence of selection within the loci included in this study. The sizes of the Ramos-Onsins & Rozas R_2 values indicate that both taxa may have undergone recent population expansions (Ramos-Onsins and Rozas 2002).

The results from STRUCTURE analysis of the four nuclear loci provide no evidence of population structure. The highest log-likelihood probability for $\ln P(D)$ was detected for $K = 1$ (Fig. 4), and structure was unable to group individuals into clusters that matched taxon boundaries (Fig. 5). The AMOVA results for these loci

generally agree with this finding, as they show that the majority of the haplotype variation can be explained by within population divergence (see Table 4). Thus, there is generally a lack of nuclear intron structure between the two taxa, with a percent of variance explained by species boundaries that ranges from -6.3% (SLC9) to 5.6% (RDP2), see Table 4 for additional values. These findings are further supported by three of the four nuclear loci, which had nonsignificant Φ_{ST} values, with the vast majority of the variation being explained by within population variation. The remaining autosomal locus, TGFB2, however, has an important difference; whereas the majority of variation within this locus is again explained by within population variation, TGFB2's Φ_{ST} value was significant, with a value of 0.056 ($P < 0.00001$).

For the IM analyses of our nuclear loci, the IM runs consistently failed to converge even after more than twenty runs with the autosomal loci – the last five of which included the Z-linked intron SLC9 – and using multiple random number seeds. (The failure to converge could be

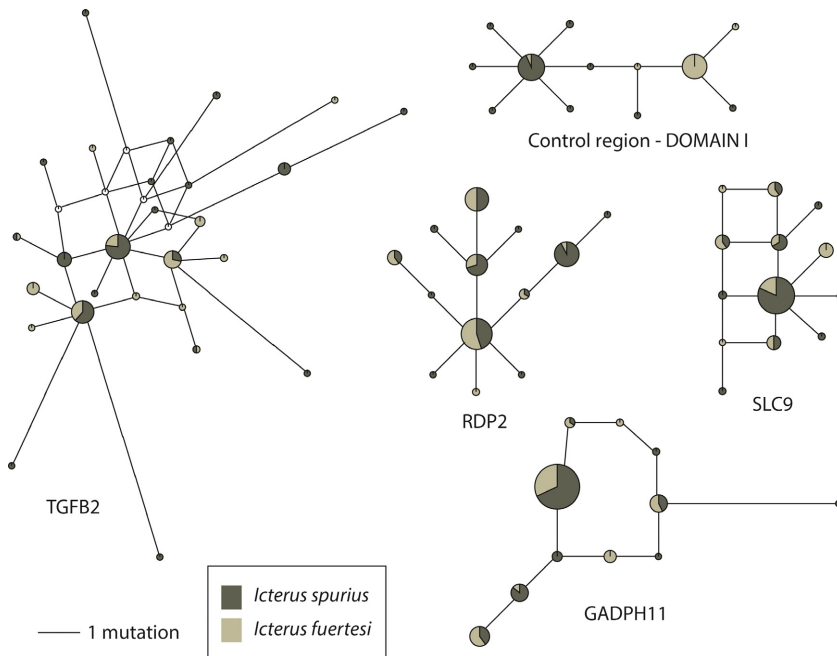


Figure 3. Haplotype networks for four nuclear introns (three autosomal and one Z-linked) show extensive haplotype sharing between *Icterus spurius* and *I. fuertesi*. In contrast, the haplotype network for the mitochondrial control region shows very little intermixing and only one shared haplotype between the two taxa. Circle sizes correspond to sample sizes of each allele.

Table 3. Results of tests for neutrality and constant populations sizes for *Icterus spurius* and *I. fuertesi*.

Locus	Tajima's D		Fu and Li's F^*		Ramos-Onsins and Rozas R_2	
	<i>I. spurius</i>	<i>I. fuertesi</i>	<i>I. spurius</i>	<i>I. fuertesi</i>	<i>I. spurius</i>	<i>I. fuertesi</i>
CR- mtDNA	-1.965*	-1.349	-2.271	-1.351	0.054	0.166
GADPH11	-0.046	1.0921	-0.637	1.375	0.101	0.177
TGFB2	-1.276	-1.008	-2.264	-2.098	0.067	0.091
RDP2	-0.052	-0.059	-0.471	0.354	0.109	0.124
SLC9	-1.965	1.014	0.152	1.309	0.087	0.186

Significant value ($0.05 > P > 0.01$) is indicated by *. None of the below values remained significant after Bonferroni corrections (adjusted at a level of 0.01).

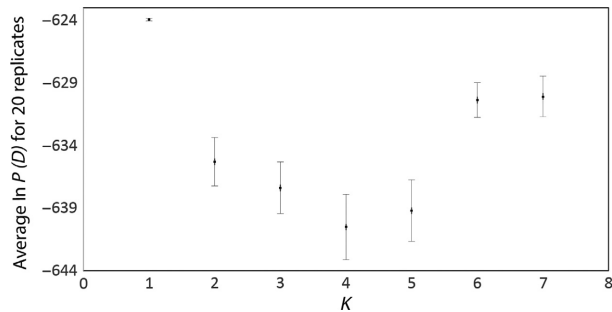


Figure 4. Mean $\ln P(D)$ with SE for $K = 1$ to 10, with 20 replicates for each K , for nuclear loci for *Icterus spurius* and *I. fuertesi* (Structure v. 2.3.4).

caused by recent divergence; other recent publications from our laboratory used similar methods and resulted in convergence; see Discussion). As a result, we will not present the results from our IM analyses in this article.

The results of the AMOVA for mtDNA CR are a clear contrast to the nuclear DNA (Table 4). These results show statistically significant structure between the two taxa, with 71.3% of genetic variation being explained by species boundaries. Additionally, the Φ_{ST} value for this locus is 0.717 ($P < 0.00001$), indicating there is evidence of a great deal of structure between the two taxa for this locus. With the addition of more *I. fuertesi*, the mtDNA haplotype frequencies are clearly statistically significantly different.

Discussion

Do nuclear DNA support very recent divergence?

The majority of the nuclear loci included in this study show very little evidence of structure between the two taxa. One nuclear locus, TGFB2, is the exception as it has a small, but significant Φ_{ST} value ($\Phi_{ST} = 0.056$, $P < 0.00001$). The amount of variation explained at the among-species level in the AMOVA for TGFB2 is 5.1%, indicating that a small portion of the genetic variation can be explained by species boundaries. If the mtDNA paraphyly first described by Baker et al. (2003) was the result of a recent mitochondrial introgression between two well-diverged lineages, we would expect to find a great deal of divergence between the nuDNA loci that would not be reflected in the mtDNA (Jacobsen and Omland 2012). Instead, these two lineages appear to be very early in the process of diverging – as evidenced by the lack of structure for the majority of nuclear loci, as well as by the lack of reciprocal monophyly in mtDNA.

With increased sampling, we found significant differences in mitochondrial haplotype frequencies

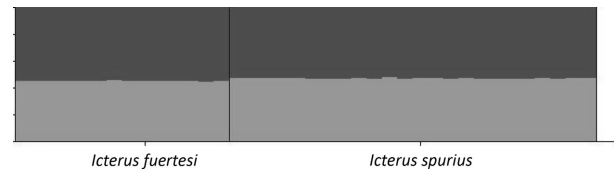


Figure 5. Results of Structure clustering analysis for nuclear loci for $K = 2$ using Structure v. 2.3.4.

Table 4. AMOVA results and Φ_{ST} values for all loci included in this study.

Locus	Φ_{ST}	Percentage of variation among species	Percentage of variation among populations within species	Percentage of variation within populations
CR- mtDNA	0.717**	71.3	1.1	27.7
GADPH11	-0.009	1.2	-6.9	105.8
TGFB2	0.056**	5.1	1.3	93.6
RDP2	0.044	5.6	-2.9	97.4
SLC9	0.086	-6.3	22.1	84.2

Significance value ($P < 0.01$) indicated by ** (P -values Bonferroni corrected).

between the two populations that confirms gene flow between the two taxa may be restricted (Baker et al. 2003). For the mitochondrial gene CR, there is clear evidence of significant population structure between the two taxa, with a Φ_{ST} value of 0.717 ($P < 0.00001$). Even with the inclusion of eight more *I. fuertesi* samples, only one haplotype is shared between the two taxa for this locus, and only two haplotypes are intermixed and showing mtDNA paraphyly. Two haplotypes that we found only in *I. spurius* are only one base pair different from *I. fuertesi* haplotypes (Fig. 3, CR), yet are at least two base pairs different than the closest *I. spurius* haplotypes. Thus, with increased sampling effort, CR seems to be nearing reciprocal monophyly between groups. These lineages show a classic pattern of intermediate divergence in which the only haplotype shared between the taxa is a central, likely ancestral haplotype. Omland et al. (2006) discussed this as an early step on the road to monophyly (“neotypy”) indicating that these two taxa are clearly showing evidence of divergence within the faster sorting mtDNA.

Does nuclear DNA support two distinct groups?

In comparison, the nuclear DNA included in this study show very little evidence of structure between the two taxa. The IM analyses failed to converge even though we tried a range of strategies to produce convergence. We

have published a number of recent studies from our laboratory on other species of recently diverged *Icterus* using similar methods, which resulted in convergence in those cases (Jacobsen and Omland 2012; Cortés-Rodríguez et al. 2013; Cortés-Rodríguez and Omland 2016). Furthermore, in the Orchard species complex, all four nuclear loci show a great deal of intermixing of alleles (Fig. 3). Additionally, the results of the AMOVAs indicates that almost all of the variation is found within the populations of each taxon (Table 4), further supporting a lack of divergence within the slower sorting nuclear loci included in the study. This is what we would expect to find if the divergence between these two taxa was at the very earliest stages of speciation.

Although the majority of variation within the nuclear loci exists at the level of variation within populations, one autosomal locus showed a small amount of variation that could be explained by species boundaries. TGFB2 had roughly 5% of its variation being explained by variation within each taxon ($P < 0.00001$). Thus, one of the nuclear loci shows some evidence of segregation, indicating that the nuclear genomes of the two taxa are beginning to diverge.

Our inability to find much structure for the nuclear loci does not necessarily mean the taxa are not evolutionarily discrete units. There are many examples of known mitochondrial paraphyly within ornithological literature, with a large portion of it being attributed to retained ancestral polymorphisms (Funk and Omland 2003; McKay and Zink 2010). Neutral nuclear markers, with their larger ESSs and slower sorting rates, are predicted to lag even further behind (Price 2008; Zink and Barrowclough 2008; Joseph and Omland 2009). Taxa that fall into this category may still be discrete in other ways that indicate they are evolutionarily distinct units, for example, for traits under sexual selection such as song or plumage coloration (Edwards et al. 2005; Marthinsen et al. 2008; Price 2008; Joseph and Omland 2009; McKay and Zink 2010). A study examining a recent divergence in two species of crows, *Corvus (corone) corone* and *C. (corone) cornix*, showed significant divergence in the genes that controlled for plumage coloration of the two taxa, yet a lack of divergence in neutral genetic markers due to hybridization and widespread introgression, showing that assortative mating and sexual selection can lead to genetic differentiation in regions under selection (Polestra et al. 2014). Similar to the above study, research in our laboratory has previously shown that *I. spurius* and *I. fuertesi* have fixed differences in adult male coloration (Hofmann et al. 2007; Kiere et al. 2007). In a parallel study to this one, we found fixed differences between the taxa in a call that seems to play a role in territory defense, thus might be under the influence of sexual selection (Sturge et al.

2016). Any genes that could be contributing to either plumage coloration or this call thus have the potential to have diverged between these two taxa.

Did *I. fuertesi* result from a founder event?

If the divergence event that separated *I. fuertesi* and *I. spurius* resulted from a few individuals of the migratory common ancestor founding a new, nonmigratory or short-distance migratory population, followed by peripheral isolates speciation (Mayr 1942, 1963; West-Eberhard 2005), we would predict that *I. fuertesi* would have significantly less genetic diversity than *I. spurius*. Looking at both the haplotype networks (Fig. 3) and the population statistics for each locus (Table 2), it is clear that both *I. spurius* and *I. fuertesi* have similar levels of genetic diversity for all nuclear loci included in this study (although the mitochondrial locus still indicates less genetic diversity within *I. fuertesi*). Based on these results, we found no evidence of a founder event in *I. fuertesi*'s recent past. This result is surprising because the distribution of *I. fuertesi* suggested that it might have evolved through this mechanism. Instead, both taxa show a great deal of allelic diversity, indicating relatively large historic population sizes in spite of the very restricted range of *I. fuertesi*. Nevertheless, it is conceivable that *I. fuertesi* was established through a large number of founders, and/or that subsequent gene flow from *I. spurius* has added a lot of genetic diversity through introgression. As the IM analyses did not converge, we are unable to completely eliminate this possibility without including many additional loci. However, a classic splitting of a single more widespread population clearly fits our data. This taxon pair would be ideally suited to next generation sequencing approaches to study this and other aspects of their divergence (e.g., testing for genes affecting plumage coloration and ecological tolerances). Whole genome comparisons would most likely provide more insight into both the divergence between these two taxa, and any potential gene flow that has occurred since this divergence. The statistical results of the Ramos-Onsins and Rozas R_2 for both taxa (Table 3) does seem to support recent and rapid population expansion for both taxa, which may have contributed to problems with IM (Hey and Nielsen 2007). Nevertheless, both taxa also have a great deal of both nucleotide and haplotype diversity, as well as unique haplotypes, supporting the existence of large historic ESSs.

Is it possible to determine whether the two taxa are currently exchanging genes?

Based on the lack of convergence in IM, due largely to the lack of structure within the nuclear loci, we are

unable to determine how much of the shared variation is the result of retained ancestral polymorphisms or is the result of gene flow. The control region shows far more evidence of structure, with only one shared haplotype and two intermixed haplotypes within all individuals included in this study. There are two possible explanations: (1) significantly different haplotype frequencies in mtDNA between the taxa indicates gene flow is likely greatly reduced or nonexistent and is not being captured at the nuDNA level due to retained ancestral polymorphisms, or (2) haplotype frequency differences for only the mtDNA could also be explained by Haldane's rule in the face of ongoing gene flow (Mckay and Zink 2010; Peters *et al.* 2012). Haldane's rule predicts that hybrids of the heterogametic sex will have reduced fitness or be sterile – so that mtDNA in birds would be less likely to introgress from one species to another (Coyne 1985; Peters *et al.* 2012). As a consequence of this, as well as the failure of the IM analyses to converge, it is difficult if not impossible to determine whether gene flow is currently occurring between these two taxa, or has historically occurred since their divergence. Although their populations breed in allopatry, gene flow could occur as migratory *I. spurius* pass through *I. fuertesi*'s breeding range in early spring. However, the significant population structure, not only within the mtDNA (which is less likely to introgress), but also within the nuclear locus TGFB2, indicates that *I. spurius* and *I. fuertesi* are in the process of diverging genetically from one another. Based on this and other evidence such as coloration and migratory patterns, it is likely that these two taxa have become evolutionarily distinct groups, regardless of any potential gene flow.

Conclusions – Model Example of Early Divergence

Icterus spurius and *I. fuertesi* are two taxa that are in the very earliest stages of speciation. Their mtDNA show only a few shared and intermixed haplotypes, having almost achieved reciprocal monophyly for this locus. The four nuclear loci included in this study show very little population structure between the two taxa – yet one locus (TGFB2) shows a small, but significant Φ_{ST} divergence between the two taxa, confirming the formation of two distinct genome pools. Based on these molecular findings, as well as the fixed differences in plumage, migratory behavior, and bioclimatic niches, there is strong evidence that these two taxa are just beginning to diverge from one another, making them ideal study organisms for the early stages of speciation and the formation of species boundaries (Hofmann *et al.* 2007; Kiere *et al.* 2007; Kondo and Omland 2007; Martin and Omland 2011).

Acknowledgments

We thank all of the following institutions for providing many of the tissue samples used in this study: Academy of Natural Sciences, Philadelphia, Drexel University; Field Museum of Natural History; J. F. Bell Museum of Natural History, University of Minnesota; Louisiana State University Museum of Zoology; Museo de Zoología, Universidad Nacional Autónoma de México (UNAM); Museum of Vertebrate Zoology, University of California, Berkeley; North Carolina State Museum; and University of Kansas Natural History Museum. We further thank UNAM, especially Adolfo Navarro-Sigüenza, Blanca Hernández-Baños, and Samuel López, for help collecting fresh tissues of *Icterus fuertesi*. During the course of this project, Kevin Omland was funded by a US National Science Foundation CAREER Grant (DEB – 0347083) and an National Science Foundation Phylogenetic Systematics Grant (DEB – 1119506). Finally, we thank Frode Jacobsen for helpful discussions and technical assistance.

Conflict of Interest

None declared.

References

- Awise, J. C. 1994. Molecular markers, natural history and evolution. Chapman and Hall, London, U.K.
- Bandelt, H. J., P. Forster, and A. Röhl. 1999. Median-joining networks for inferring intraspecific phylogenies. *Mol. Biol. Evol.* 16:37–48.
- Baker, J. M., E. López-Medrano, A. G. Navarro-Sigüenza, O. R. Rojas-Soto, and K. E. Omland. 2003. Recent speciation in the Orchard Oriole group: divergence of *Icterus spurius* and *Icterus spurius fuertesi*. *Auk* 120:848–859.
- Bureš, S., P. Nádvořník, and G. P. Sætre. 2002. Hybridization and apparent hybridization between meadow pipit (*Anthus pratensis*) and water pipit (*A. spinoletta*). *Hereditas* 136:254–256.
- Chesser, R. T., R. C. Banks, K. J. Burns, C. Cicero, J. L. Dunn, A. W. Kratter, I. J. Lovette, A. G. Navarro-Sigüenza, P. C. Rasmussen, J. V. Remsen Jr, J. D. Rising, D. F. Stotz, and K. Winker. 2015. Fifty-sixth supplement to the American Ornithologists' Union check-list of North American birds. *Auk* 132:748–764.
- Clements, J. F. 2007. The Clements checklist of birds of the world. Cornell Univ. Press, Ithaca, NY.
- Cortés-Rodríguez, M. N., and K. E. Omland. 2016. Morphological and genetic variation of the Yellow-backed Oriole (*Icterus chrysater*) across its widely disjunct distribution in Central America. *Wilson J. Ornithol.* in press.
- Cortés-Rodríguez, N., F. Jacobsen, B. E. Hernandez-Baños, A. G. Navarro-Sigüenza, J. L. Peters, and K. E. Omland. 2013.

- Coalescent analyses show isolation without migration in two closely related tropical orioles: the case of *Icterus graduacauda* and *Icterus chrysater*. *Ecol. Evol.* 3:4377–4387.
- Coyne, J. A. 1985. The genetic basis for Haldane's rule. *Nature* 314:736–738.
- Edwards, S. V., and P. Beerli. 2000. Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Edwards, S. V., S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson, et al. 2005. Speciation in birds: genes, geography, and sexual selection. *Proc. Natl Acad. Sci. USA* 102(Suppl. 1):6550–6557.
- Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* 10:564–567.
- Fu, Y., and W. Li. 1993. Statistical tests of neutrality of mutations. *Genetics* 133:693–709.
- Funk, D. J., and K. E. Omland. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Evol. Syst.* 34:397–423.
- Gill, F., and D. Donsker (Eds). 2015. IOC world bird list (v. 5.4). <http://www.worldbirdnames.org>. doi: 10.14344/IOC.ML.5.4.
- Hagemeyer, N. D. G., R. J. Sturge, K. E. Omland, and J. J. Price. 2012. Incomplete song divergence between recently diverged taxa: syllable sharing by Orchard and Fuertes' orioles. *J. Field Ornithol.* 83:362–371.
- Harrison, R. G. 1991. Molecular changes at speciation. *Annu. Rev. Ecol. Syst.* 22:281–308.
- Hebert, P. D. N., M. Y. Stoeckle, T. S. Zemlak, and C. M. Francis. 2004. Identification of birds through DNA barcodes. *PLoS Biol.* 2:1657–1663.
- Hey, J. (2007). Using the IMA program. Distributed by the author: <http://lifesci.rutgers.edu/~heylab/heysoftware.htm>.
- Hey, J. 2006. Recent advances in assessing gene flow between diverging populations and species. *Curr. Opin. Genet. Dev.* 16:592–596.
- Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* 167:747–760.
- Hey, J., and R. Nielsen. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proc. Natl Acad. Sci. USA* 104:2785–2790.
- Hofmann, C. M., K. J. McGraw, T. W. Cronin, and K. E. Omland. 2007. Melanin coloration in New World orioles I: carotenoid masking and pigment dichromatism in the orchard oriole complex. *J. Avian Biol.* 38:163–171.
- Howell, S. N. G., and S. Webb. 1995. A guide to the birds of Mexico and northern Central America. Oxford Univ. Press, Oxford, U.K.
- Hubisz, M. J., D. Falush, M. Stephens, and J. K. Pritchard. 2009. Inferring weak population structure with the assistance of sample group information. *Mol. Ecol. Resour.* 9:1322–1332.
- Jacobsen, F., and K. E. Omland. 2012. Extensive introgressive hybridization within the northern oriole group (Genus *Icterus*) revealed by three-species isolation with migration analysis. *Ecol. Evol.* 2:2413–2429.
- Jaramillo, A., and P. Burke. 1999. New world blackbirds: the Icterids. Princeton Univ. Press, Princeton, NJ.
- Joseph, L., and K. Omland. 2009. Phylogeography: its development and impact in Australo-Papuan ornithology with special reference to paraphyly in Australian birds. *Emu* 109:1–23.
- Kiere, L. M., C. M. Hofmann, I. E. Tracy, T. W. Cronin, J. Leips, and K. E. Omland. 2007. Using color to define species boundaries: quantitative analysis in the orchard oriole complex supports the recognition of two species. *Condor* 109:692–697.
- Knowles, L. L., and B. C. Carstens. 2007. Delimiting species without monophyletic gene trees. *Syst. Biol.* 56:887–895.
- 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.
- Kondo, B., J. M. Baker, and K. E. Omland. 2004. Recent speciation between the Baltimore Oriole and the Black-backed Oriole. *Condor* 106:674–680.
- Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics* 28:298–299.
- Marthinsen, G., L. Wennerberg, and J. Lifjeld. 2008. Low support for separate species within the redpoll complex (*Carduelis flammea-hornemanni-cabaret*) from analyses of mtDNA and microsatellite markers. *Mol. Phylogenet. Evol.* 47:1005–1017.
- Martin, M., and K. Omland. 2011. Environmental niche modeling reveals climatic differences among breeding ranges of Orchard Oriole subspecies. *Am. Midl. Nat.* 166:404–414.
- Mayr, E. 1942. Systematics and the origin of species from the viewpoint of a zoologist. Columbia Univ. Press, New York.
- Mayr, E. 1963. Animal species and evolution. Belknap Press of Harvard University, Cambridge, MA.
- Mckay, B., and R. Zink. 2010. The causes of mitochondrial DNA gene tree paraphyly in birds. *Mol. Phylogenet. Evol.* 54:647–650.
- Milne, I., F. Wright, G. Rowe, D. Marshall, D. Husmeier, and G. McGuire. 2004. TOPALi: software for automatic identification of recombinant sequences within DNA multiple alignments. *Bioinformatics* 20:1806–1807.
- Moritz, C. 1994. Defining 'Evolutionarily Significant Units' for conservation. *Trends Ecol. Evol.* 9:373–375.
- Navarro-Sigüenza, A. G., and A. Townsend Peterson. 2004. An alternative species taxonomy of the birds of Mexico. *Biota Neotrop.* 4:1–32.

- Olsson, U., P. Alström, P. G. P. Ericson, and P. Sundberg. 2005. Non-monophyletic taxa and cryptic species – evidence from a molecular phylogeny of leaf-warblers (*Phylloscopus*, Aves). *Mol. Phylogenet. Evol.* 36:261–276.
- Omland, K. E., J. M. Baker, and J. L. Peters. 2006. Genetic signatures of intermediate divergence: population history of Old and New World Holarctic ravens (*Corvus corax*). *Mol. Ecol.* 15:795–808.
- Peters, J. L., Y. N. Zhuravlev, I. Fefelov, A. Logie, and K. E. Omland. 2007. Nuclear loci and coalescent methods support ancient hybridization as cause of mitochondrial paraphyly between gadwall and falcated duck (*Anas* spp.). *Evolution* 61:1992–2006.
- Peters, J. L., Y. N. Zhuravlev, I. Fefelov, E. M. Humphries, and K. E. Omland. 2008. Multilocus phylogeography of a Holarctic duck: colonization of North America from Eurasia by gadwall (*Anas strepera*). *Evolution* 62:1469–1483.
- Peters, J., K. Bolender, and J. Pearce. 2012. Behavioural vs. molecular sources of conflict between nuclear and mitochondrial DNA: the role of male-biased dispersal in a Holarctic sea duck. *Mol. Ecol.* 21:3562–3575.
- Polestra, J. W., N. Vijay, C. M. Bossu, H. Lantz, B. Ryll, I. Müller, et al. 2014. The genomic landscape underlying phenotypic integrity in the face of gene flow in crows. *Science* 344:1410–1414.
- Price, T. D. 2008. *Speciation in birds*. Roberts and Company, Greenwood Village, CO.
- Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. *Genetics* 155:945–959.
- Primmer, C. R., T. Borge, J. Lindell, and G. P. Sætre. 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol. Ecol.* 11:603–612.
- Ramos-Onsins, S., and J. Rozas. 2002. Statistical properties of new neutrality tests against population growth. *Mol. Biol. Evol.* 19:2092–2100.
- Rheindt, F. E., and S. V. Edwards. 2011. Genetic introgression: an integral but neglected component of speciation in birds. *Auk* 128:620–632.
- Rosenberg, N. A., and M. Nordborg. 2002. Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* 3:380–390.
- Ross, H. A. 2014. The incidence of species-level paraphyly in animals: a reassessment. *Mol. Phylogenet. Evol.* 76:10–17.
- Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19:2496–2497.
- Schneider, S., J. M. Keuffer, D. Roessli, and L. Excoffier. 2000. ARLEQUIN: a software package for population genetics. Genetics & Biometry Laboratory, Dept. Of Anthropology, University of Geneva, Geneva, Switzerland.
- Stephens, M., and P. Scheet. 2005. Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. *Am. J. Hum. Genet.* 76:449–462.
- Stephens, M., N. J. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* 68:978–989.
- Sturge, R. J., K. E. Omland, J. J. Price, and B. Lohr. 2016. Divergence in calls but not songs in the Orchard Oriole complex: *Icterus spurius* and *Icterus fuertesi*. *J. Avian Biol.* 47:109–120.
- Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595.
- Tobón-Sampedro, A., and O. R. Rojas-Soto. 2015. The geographic and seasonal potential distribution of the little known Fuertes's Oriole *Icterus fuertesi*. *Bird Conserv. Int.* 25:489–502.
- Waltari, E., and S. V. Edwards. 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in Archosaurs. *Am. Nat.* 160:539–552.
- West-Eberhard, M. 2005. Developmental plasticity and the origin of species differences. *Proc. Natl Acad. Sci. USA* 102:6543–6549.
- Winker, K. 2000. Evolution – migration and speciation. *Nature* 404:36.
- Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. *Mol. Ecol.* 17:2107–2121.
- Zink, R. M., and M. C. McKittrick. 1995. The debate over species concepts and its implications for ornithology. *Auk* 112:701–719.

Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article:

Table S1. Specimen numbers and sampling locations for tissue samples.