doi:10.1111/j.1558-5646.2010.00954.x

HISTORICAL DIVERGENCE AND GENE FLOW: COALESCENT ANALYSES OF MITOCHONDRIAL, AUTOSOMAL AND SEX-LINKED LOCI IN *PASSERINA* BUNTINGS

Matthew D. Carling,[1,2,3,4,5] Irby J. Lovette,[3,4] and Robb T. Brumfield[1,2]

[1]*Museum of Natural Science, 119 Foster Hall, Louisiana State University, Baton Rouge, Louisiana 70803*

[2]*Department of Biological Sciences, Louisiana State University, Baton Rouge, Louisiana 70803*

[3]*Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, 159 Sapsucker Woods Road, Ithaca, New York 14850*

[4]*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853*

[5]*E-mail: mdc248@cornell.edu*

Received March 27, 2009
Accepted December 14, 2009

Quantifying the role of gene flow during the divergence of closely related species is crucial to understanding the process of speciation. We collected DNA sequence data from 20 loci (one mitochondrial, 13 autosomal, and six sex-linked) for population samples of Lazuli Buntings (*Passerina amoena*) and Indigo Buntings (*Passerina cyanea*) (Aves: Cardinalidae) to test explicitly between a strict allopatric speciation model and a model in which divergence occurred despite postdivergence gene flow. Likelihood ratio tests of coalescent-based population genetic parameter estimates indicated a strong signal of postdivergence gene flow and a strict allopatric speciation model was rejected. Analyses of partitioned datasets (mitochondrial, autosomal, and sex-linked) suggest the overall gene flow patterns are driven primarily by autosomal gene flow, as there is no evidence of mitochondrial gene flow and we were unable to reject an allopatric speciation model for the sex-linked data. This pattern is consistent with either a parapatric divergence model or repeated periods of allopatry with gene flow occurring via secondary contact. These results are consistent with the low fitness of female avian hybrids under Haldane's rule and demonstrate that sex-linked loci likely are important in the initial generation of reproductive isolation, not just its maintenance.

KEY WORDS: **Divergence population genetics, indigo bunting, isolation-with-migration, Lazuli Bunting.**

The increased ease of gathering DNA sequence data from multiple unlinked loci combined with the development of coalescent-based analytical methods to analyze such multilocus datasets allows researchers to investigate and quantify the role of gene flow and introgression in divergence and speciation. Although the allopatric speciation model, in which species originate in isolation without the exchange of genes, continues to be considered the predominate mode of speciation in most taxa (Coyne and Orr 2004), the application of "divergence population genetics (DPG)" methods has provided empirical evidence that isolation with gene flow may play a larger role in the speciation process than previously thought (Hey 2006; Nosil 2008). Although many DPG studies have uncovered divergence patterns consistent with postdivergence gene flow (e.g., Won and Hey 2005; Lee and Edwards 2008; Niemiller et al. 2008; Strasburg and Rieseberg 2008; Kane et al. 2009), most have not formally tested alternative models of strict allopatry and isolation with gene flow (but see Geraldes et al. 2008; Nadachowska and Babik 2009; Ross-Ibarro et al. 2009). Because the majority of DPG studies have not explicitly rejected an allopatric speciation model, it is difficult to assess

accurately whether speciation in the face of gene flow is a common phenomenon.

It is also possible to use DPG to investigate the role that sex-chromosomes play in the formation of new species over the course of divergence between two closely related taxa. Empirical studies of hybrid zones have demonstrated reduced introgression of sex-linked loci relative to autosomal loci (Hagen and Scriber 1989; Tucker et al. 1992; Saetre et al. 2003; Borge et al. 2005; Carling and Brumfield 2008a), but cline-based studies have two limitations. First, the differences in the effective population size of autosomal and sex-linked loci may contribute to observed differences in cline widths. Current methods of cline-based analyses do not consider differences in effective population size among loci. In contrast, differences in effective population size are readily accounted for in coalescent-based DPG analyses. Second, when applied to investigations of gene flow between divergent taxa, cline-based analyses are typically used to describe patterns of gene flow that result upon secondary contact between partially reproductively isolated taxa. Whether such introgression patterns are indicative of gene flow patterns occurring during divergence remains unknown.

Here, we apply the recently described "isolation-with-migration analytic" model (Hey and Nielsen 2007) to test two hypotheses related to the divergence process between two species of buntings in the genus *Passerina* (Aves: Cardinalidae). The first hypothesis is a strict isolation model whereby speciation resulted from divergence in allopatry without any subsequent gene flow (Fig. 1A). The second hypothesis is an isolation-with-migration model, in which speciation resulted despite the presence of gene flow between the two diverging lineages (Fig. 1B). It is important to note that we are investigating the role of postdivergence gene flow that occurred prior to current hybridization that has resulted from secondary contact, which likely arose within the last 6500 years (see below). We also use DPG to investigate the relative contributions of mitochondrial, autosomal, and sex-linked gene flow to the overall patterns of gene flow during divergence. If sex-linked loci play a disproportionately large role in reproductive isolation and speciation (Charlesworth et al. 1987; Coyne and Orr 1989), this should be evident in reduced gene flow estimates for these sex-linked loci relative to autosomal loci.

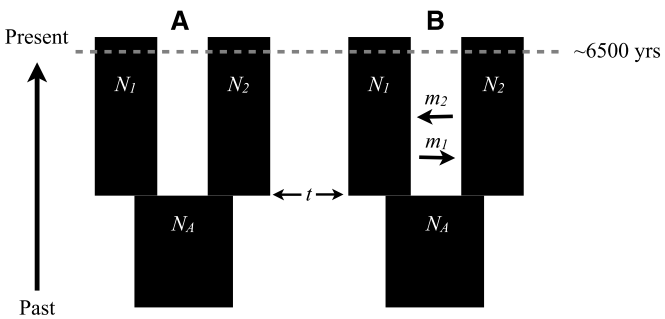

Figure 1. Schematic of strict allopatric speciation model (A) versus isolation-with-migration speciation model (B). The dashed line represents the approximate timing of secondary contact that has resulted in current hybridization. Based on available evidence, this secondary contact likely began no more than ~6500 years ago (Abrams 1992).

STUDY SYSTEM

Passerina amoena (Lazuli Bunting) and *P. cyanea* (Indigo Bunting) are sister species (Carling and Brumfield 2008b) that hybridize where their breeding ranges overlap, as a result of secondary contact, in the western Great Plains and eastern foothills of the Rocky Mountains of North America (Sibley and Short 1959; Emlen et al. 1975; Kroodsma 1975; Baker and Boylan 1999). Exactly when *P. amoena* and *P. cyanea* came into secondary contact is unclear, but breeding records indicate it has been at least 120 years (Sibley and Short 1959). An estimate of the upper limit can be derived from paleoecological data that suggest a warming trend between 6500 and 3500 years ago resulted in the expansion of oak (*Quercus*) savannas in the mid-west and central prairie regions of North America (Abrams 1992). Such oak savannas, along with associated undergrowth species, likely provided a suitable breeding habitat for *Passerina* buntings (Payne 1992; Greene et al. 1996). Taken together, current hybridization between *P. amoena* and *P. cyanea* likely began at least 120 years ago, but probably not more than 6500 years ago.

Previous studies of the *Passerina* hybrid zone have indicated that sex-linked loci likely play an important role in the continued maintenance of reproductive isolation between these species (Carling and Brumfield 2008a, 2009). Using population samples collected along a geographic transect spanning the contact zone, cline-based analyses of multiple loci showed that mean autosomal clines were significantly wider than mean sex-linked and mitochondrial clines (Carling and Brumfield 2008a). What is less clear is whether divergence prior to secondary contact (i.e., prior to 6500 years before present) between *P. amoena* and *P. cyanea* occurred in complete geographic isolation or whether it occurred in the face of ongoing gene flow at that time (Fig. 1). In addition, the importance of reduced gene flow of sex-linked loci during speciation is unknown, as is a robust estimate of the divergence time between these species. The goal of this study is to address these questions using DNA sequence data from 20 loci.

Methods

SAMPLING AND MOLECULAR METHODS

Samples were obtained from four populations, two *P. amoena* populations west of the contact zone (Washington and Wyoming) and two *P. cyanea* populations east of the contact zone (Illinois and Louisiana; Fig. 2). These populations were chosen specifically

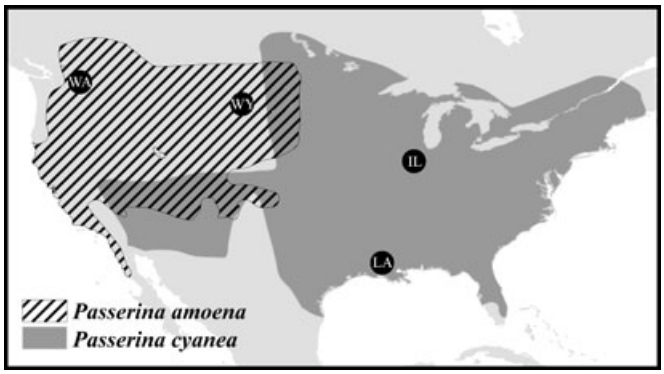

Figure 2. Breeding distributions and sampling localities for populations of *Passerina amoena* and *P. cyanea*. Digital maps (Ridgely et al. 2003) were downloaded from NatureServe (2006) and modified.

because they lie well outside the previously defined contact zone (Fig. 2); therefore divergence time and gene flow estimates are unlikely to be strongly influenced by ongoing hybridization between these species. Sample sizes for the Washington and Louisiana populations were 12 and 13 individuals, respectively. As only 10 individuals were available from Illinois, two individuals from Minnesota were added to the Illinois sample to bring the total to 12. Similarly, two individuals from Montana were added to the 10 individuals from Wyoming, bringing the total to 12 individuals. All of these individuals are vouchered (Table S1) and have been included in previous research on patterns of differential introgression across the *Passerina* hybrid zone (Carling and Brumfield 2008a, 2009).

We extracted genomic DNA from ~25 mg of pectoral muscle from all individuals using a DNeasy Tissue Kit (Qiagen, Valencia, CA) and amplified each individual at 20 loci (one mitochondrial, 13 nuclear autosomal and six sex-linked; Table 1) using standard PCR conditions (Table S2). In contrast to mammals, birds have ZW sex-determination in which females are the heterogametic sex. All of the sex-linked loci are located on the Z chromosome (Carling and Brumfield 2009), which is analogous to the X chromosome in mammals. Some of these loci were examined in previous studies (Table S2), but all were chosen based on ease of amplification. This strategy was chosen so as to not systematically bias parameter estimates that may result from only including loci that show strong differentiation between *P. amoena* and *P. cyanea*. PCR amplicons were purified using either 20% polyethylene glycol (PEG) or Exo-Sap and then cycle-sequenced in both directions using the Big Dye Terminator Cycle-Sequencing Kit version 3.1 (Applied Biosystems Inc., Foster City, CA). We edited and assembled all sequences using Sequencher version 4.7 (GeneCodes Corp., Ann Arbor, MI). For those sequences that contained more than one heterozygous site, we resolved haplotypes probabilistically using PHASE (Stephens et al. 2001; Stephens and Donnelly 2003), which can accurately infer haplotypes even when the number of individuals sampled is less than in this study (Harrigan et al. 2008). Only those individuals for which PHASE was able to assign haplotypes with a probability greater than 0.70 were used in subsequent analyses. We conservatively used a probability of 0.70 as our cutoff because it has been shown that haplotypes inferred by PHASE with a probability greater than 0.60 accurately reflect haplotypes determined through cloning (Harrigan et al. 2008). All sequences newly generated for this study have been deposited in GenBank (accession numbers GU197010 through GU197377, GU215039 through GU215074 and Table S3).

DATA ANALYSES

We tested for intralocus recombination using the four-gamete test as implemented in DnaSP version 4.10 (Rozas et al. 2003). When recombination was detected, we kept the longest independently segregating block of sequence for subsequent analyses. We also used DnaSP to calculate haplotype diversity (H_d) (Nei 1987), nucleotide diversity (π) (Nei and Li 1979), and Tajima's D, to test for selection (Tajima 1983) for each locus in each species (Table 1). We tested for disequilibrium between all pairs of loci and ran an analysis of molecular variance (AMOVA) on the mitochondrial data using Arlequin version 3.11 (Excoffier et al. 2005).

Divergence time (t), levels of gene flow (m_1 and m_2), and effective population sizes (q_1, q_2, and q_A) were estimated using the coalescent-based "isolation-with-migration analytic" model implemented in the program IMa (Hey and Nielsen 2007). Two general classes of IMa runs were conducted. The first included all sampled individuals and investigated the relative contributions of each marker class (mitochondrial, autosomal and sex-linked) to the overall speciation history of *P. amoena* and *P. cyanea*. Four separate IMa analyses were run to address this first question, using all individuals but different subsets of loci: (1) the complete dataset, which contained all loci, (2) the mitochondrial dataset, which contained only the two mitochondrial loci, (3) the autosomal dataset, which contained only the 13 autosomal loci, and (4) the z-linked dataset, which contained only the six z-linked loci. The second class of IMa runs was designed to explore the geographic structure of gene flow between these species by comparing divergence time and gene flow estimates between the two most distant populations (WA and LA) with the estimates between the two closer populations (WY and IL). Separate analyses were run for two datasets using all loci but only subsets of individuals: (1) WA/LA, which consisted of individuals from WA and LA, and (2) WY/IL, which included individuals from WY and IL.

In all IMa runs, only the longest independently segregating block of sequence (Table 1), as identified using the four-gamete test, was used for each locus. In addition, only assigned haplotypes with a posterior probability greater than 0.70 were used. Initial runs with wide priors (q_1, q_2, $q_A = 20$, m_1, $m_2 = 10$,

Table 1. Loci sampled and population genetic parameter estimates within *Passerina amoena* and *P. cyanea*.

Locus	Species	Length (bp)[1]	N^2	H_d	π (variance)	Tajima's *D*
mtDNA						
ND3+Cont. Reg.	*P. amoena*	524	24	0.958	0.053 (0.022)	−1.947
	P. cyanea		23	0.972	0.006 (0.002)	−1.401
Autosomal						
16214	*P. amoena*	252	40	0.583	0.003 (0.001)	−1.371
	P. cyanea		32	0.482	0.003 (0.001)	−1.471
17483	*P. amoena*	112	40	0.748	0.014 (0.005)	0.925
	P. cyanea		32	0.804	0.015 (0.005)	−0.189
18503	*P. amoena*	344	36	0.838	0.005 (0.002)	−0.275
	P. cyanea		30	0.710	0.003 (0.001)	−1.869
23361	*P. amoena*	123	40	0.644	0.008 (0.003)	−0.835
	P. cyanea		30	0.715	0.008 (0.003)	−1.043
24792	*P. amoena*	264	38	0.671	0.003 (0.001)	−0.611
	P. cyanea		32	0.391	0.002 (0.001)	−1.638
AETC	*P. amoena*	111	42	0.340	0.003 (0.001)	−1.027
	P. cyanea		32	0.286	0.003 (0.001)	−1.379
βACT3	*P. amoena*	128	36	0.560	0.005 (0.002)	−1.428
	P. cyanea		30	0.634	0.007 (0.003)	−1.960
GADPH	*P. amoena*	101	48	0.709	0.009 (0.003)	−0.690
	P. cyanea		48	0.537	0.006 (0.002)	−1.388
MC1R	*P. amoena*	202	48	0.197	0.001 (0.0003)	−1.459
	P. cyanea		42	0.391	0.002 (0.001)	−0.884
MYO2	*P. amoena*	113	40	0.468	0.005 (0.002)	−0.393
	P. cyanea		40	0.594	0.007 (0.002)	−0.373
RHO1	*P. amoena*	114	48	0.196	0.002 (0.001)	−0.714
	P. cyanea		48	0.159	0.001 (0.0003)	−1.157
TGFβ2	*P. amoena*	217	32	0.769	0.005 (0.002)	−1.271
	P. cyanea		48	0.532	0.004 (0.001)	−2.224
TROPO	*P. amoena*	286	18	0.446	0.002 (0.001)	−1.240
	P. cyanea		34	0.405	0.002 (0.001)	−1.853
Sex-linked						
24105	*P. amoena*	186	36	0.000	0.000	–
	P. cyanea		40	0.740	0.005 (0.002)	−2.082
ALDOB3	*P. amoena*	111	42	0.436	0.004 (0.001)	−0.686
	P. cyanea		39	0.588	0.006 (0.002)	−1.622
BRM15	*P. amoena*	300	40	0.235	0.001 (0.0003)	−1.518
	P. cyanea		32	0.698	0.004 (0.001)	−1.482
IQGAP2	*P. amoena*	70	40	0.806	0.018 (0.006)	0.731
	P. cyanea		40	0.312	0.005 (0.002)	−0.575
PPWD1	*P. amoena*	85	41	0.000	0.000	–
	P. cyanea		36	0.215	0.005 (0.002)	−0.141
VLDLR9	*P. amoena*	240	42	0.184	0.001 (0.0003)	−2.001
	P. cyanea		42	0.699	0.004 (0.001)	−1.284

[1]Length of longest independently segregating block of sequence.

[2]Number of individuals (mtDNA) or chromosomes (all other loci) sampled.

$t = 10$) were used to identify priors that encompassed the entire distribution of each parameter estimate for use in subsequent runs. Once appropriate priors were identified, we ran three replicate runs for each analysis that differed only in starting random number seed. In all runs, we used at least 25 Markov-coupled chains with a geometric heating scheme (g1 = 0.95, g2 = 0.80), a burn-in of 150,000 steps, and a run length of at least 5×10^6 steps. To ensure proper chain mixing and parameter convergence, all parameter trend lines were visually inspected and the three independent runs of each analysis, which differed only in starting

random seed, were compared. Replicate, independent runs of the mitochondrial dataset indicated no gene flow between *P. amoena* and *P. cyanea* (they are reciprocally monophyletic), so both migration parameters were removed for the final runs. Removing unnecessary parameters allows for slightly more robust estimates of included parameters (J. Hey, pers. comm.), but did not change any results in this study.

As indicated, we ran all IMa analyses (except the mitochondrial data) under the full model, which estimates six demographic parameters. Once runs had converged, we tested for the fit of the data to simpler demographic models using the nested model approach in the "Load-Trees" mode of IMa. This test calculates log-likelihood ratio statistics for all possible nested models, the significance of which can be assessed using a chi-square test (Hey and Nielsen 2007). Calculating the log-likelihood statistics of nested models allows for a statistical test of whether the data are consistent with a strict allopatric speciation model. Specifically, we tested the full model against a model in which both gene flow parameters (m_1, m_2) were set to zero. Because we could not reject the allopatric speciation model for the z-linked dataset, we also analyzed those data without including the gene flow parameters. Additionally, we created three replicate datasets that each contained a randomly chosen subset of six autosomal loci. These replicate datasets allowed us to investigate whether any differences in model rejection between the z-linked and autosomal datasets were due to differences in statistical power given the differences in number of loci in the full datasets (13 autosomal vs. six z-linked). IMa analyses of the replicate datasets were conducted using the same strategy as outlined above.

All parameter estimates are scaled to the neutral mutation rate (μ), which is approximately 10 times faster for mitochondrial loci than for nuclear loci (Graur and Li 2000). To convert the divergence time estimate (t) to years, we used a generation time of one year (Greene et al. 1996; Payne 2006) and a neutral mutation rate of 1.35×10^{-9} substitutions/site/year for the autosomal loci (Ellegren 2007) and 1.45×10^{-9} substitutions/site/year for the z-linked loci (Axelsson et al. 2004; Ellegren 2007) and 1.35×10^{-8} substitutions/site/year for the mitochondrial locus (autosomal rate $\times$ 10). Using these per-site mutation rates, we calculated a mean per-locus mutation rate (5.85×10^{-7} substitutions/locus/year) which is required when using IMa to estimate divergence time in years. We also calculated a mitochondrial divergence time using an avian-calibrated divergence rate of 2.1% ($\pm$0.1%) per million years for the cytochrome *b* gene (Weir and Schluter 2008) and previously published data on the mean-corrected cytochrome *b* pairwise divergence between *P. amoena* and *P. cyanea* (Klicka et al. 2001). We did not calculate a mitochondrial divergence time estimate using IMa because the mitochondrial haplotypes of the individuals of the two species included in this study are reciprocally monophyletic.

Results

The mean length of the longest independently segregating block of sequence data for all sampled nuclear loci was 177 bp. The mean length was 194 bp for all loci, including the mitochondrial sequences. As a result of the presence of multiple indels and the inability of PHASE to assign some haplotypes with a probability greater than 0.70, DNA sequence data were not analyzed for all individuals at all loci (Table 1). Across loci, levels of variation were highly variable, with both haplotype diversity and nucleotide diversity differing by more than one order of magnitude (Table 1).

No loci showed evidence of selection after a Bonferroni correction for multiple tests was applied (Table 1). Tests of pairwise disequilibrium indicated that all loci were effectively unlinked and could be treated as independent loci in the IMa analyses ($P > 0.05$ in all cases). Of the total molecular variance in the mitochondrial data, most (85.02%) was between *P. cyanea* and *P. amoena*, and nearly all of the rest (14.84%) was within populations of each species. Very little (0.13%) of the mitochondrial variation resulted from differences between populations within each species.

The strict allopatric speciation model was rejected for the complete (all loci, all individuals), all loci WA and LA, all loci WY and IL, and autosomal datasets (2LLR $>$ 106 in all tests, df $=$ 2, $P < 0.001$ for all tests; Table 2). Analyzing only a randomly chosen subset of six autosomal loci did not change the results; the strict allopatric speciation model was rejected for all three replicate datasets (Table 2). In contrast, the strict allopatric speciation model could not be rejected for the z-linked dataset (2LLR $=$ 4.35, df $= 2$, $P > 0.10$; Table 2). Tests of nested models were not performed for the dataset containing only mitochondrial data because migration parameters were not included in those IMa runs.

In all IMa analyses, the estimated effective population size of *P. cyanea* was larger than for *P. amoena* (Table 2). Divergence time estimates (with the exception of the estimate from the mitochondrial dataset) were similar across all analyses (range $=$ 0.28–0.64; Fig. 3; Table 3). The divergence time estimated from the mitochondrial dataset was 2.67 (Fig. 3; Table 3). Introgression from *P. cyanea* into *P. amoena* was higher than from *P. amoena* into *P. cyanea* (Fig. 3; Table 3). When just the autosomal data were analyzed, the difference in introgression was nearly one order of magnitude ($m_1 = 0.37$, $m_2 = 2.12$; Fig. 3; Table 3). All parameters estimated from the WA/LA and WY/IL datasets were similar (Fig. 3; Table 3).

The effective population size ($N_e = \theta/4\mu$, where the mean $\mu = 5.85 \times 10^{-7}$ substitutions/locus/year), calculated from the dataset with all loci and all individuals, for *P. cyanea* was $\sim$1,180,000 individuals, and was $\sim$400,000 individuals for *P. amoena*. Using the same dataset, *P. amoena* and *P. cyanea* diverged $\sim$1,040,000 years ago (90% highest posterior density:

Table 2. Tests of nested models for divergence between *Passerina amoena* and *P. cyanea*. All models are tested against the full six parameter (θ_1, θ_2, θ_A, m_1, m_2, t) model.

Analysis	Model (Θ)	log ($P'(\Theta \mid X)$)	df[1]	2LLR	P[2]
All loci, all individuals	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−215.43	2	439.89	<0.001
All loci, WY and IL	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−72.83	2	156.82	<0.001
All loci, WA and LA	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−47.18	2	106.78	<0.001
Autosomal loci, all individuals	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−262.71	2	527.95	<0.001
Six autosomal loci, all individuals—rep 1[3]	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−75.03	2	155.62	<0.001
Six autosomal loci, all individuals—rep 2[3]	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−54.56	2	115.52	<0.001
Six autosomal loci, all individuals—rep 3[3]	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	−166.30	2	333.40	<0.001
Z-linked loci, all individuals	$\theta_1, \theta_2, \theta_A, m_1=0, m_2=0$	1.63	2	4.35	>0.10

[1]Because all nested models include parameters fixed at a boundary (m_1=0, m_2=0), the chi-square distributions are mixtures of the multiple chi-square distributions (Hey and Nielsen 2007). Accounting for this mixture does not change any significance values.

[2]All *P*-values<0.001 are significant after a Bonferroni correction (α=0.05/8 tests=0.00625 corrected α).

[3]These datasets consisted of 6 (from a total of 13 loci) randomly chosen autosomal loci.

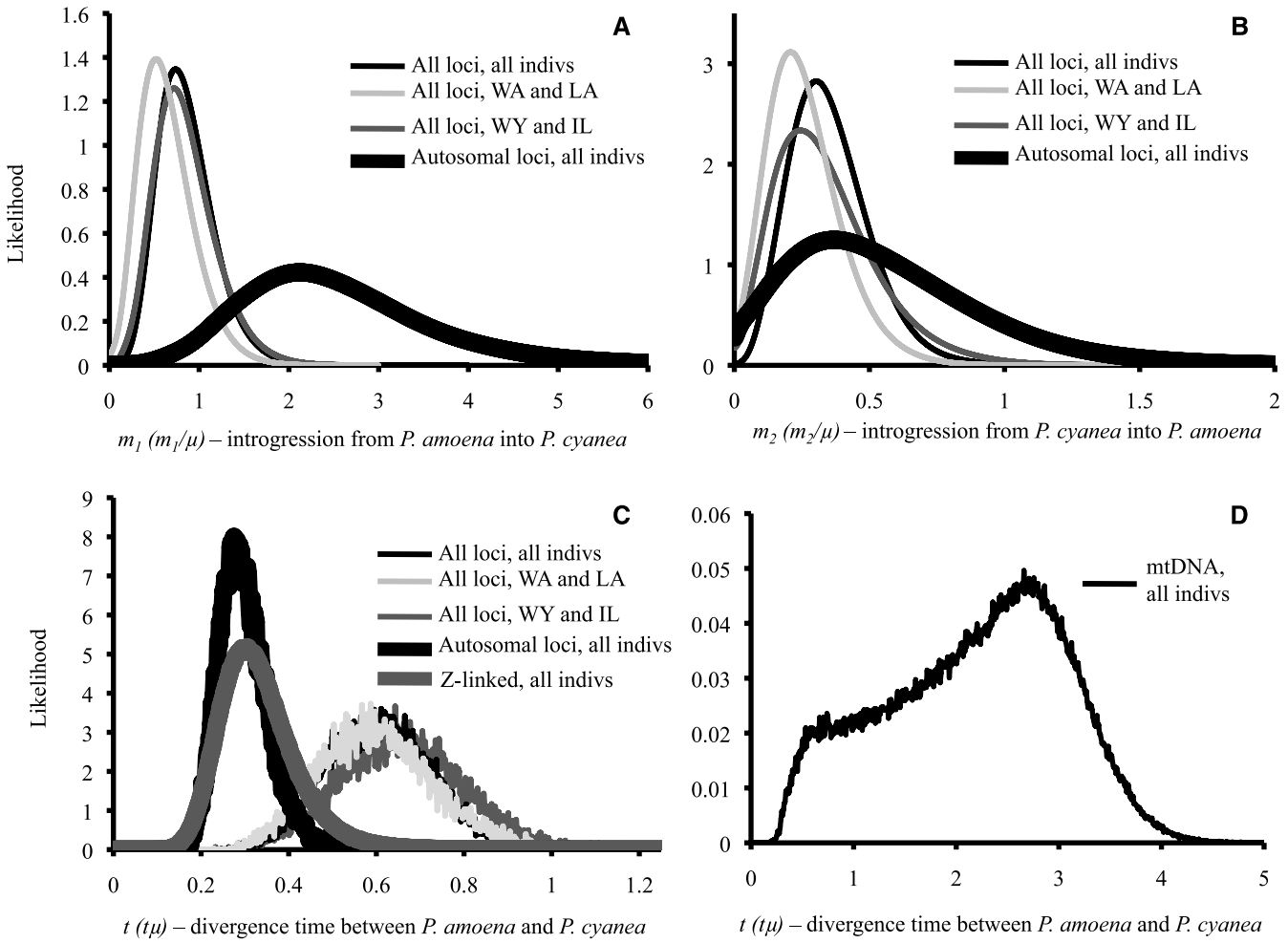

Figure 3. Population genetic parameter estimate distributions. All parameter estimates are scaled to the neutral mutation rate (μ). (A) Introgression from *P. amoena* into *P. cyanea*. (B) Introgression from *P. cyanea* into *P. amoena*. (C) Divergence time estimates from all analyses except the mitochondrial dataset. (D) Divergence time estimate from the mitochondrial dataset. Note the different axes scales.

Table 3. Maximum likelihood estimates and 90% highest posterior density (in parentheses) intervals of demographic parameter estimates for divergence between *Passerina amoena* and *P. cyanea*.

Analysis	$\theta_1=4N_1\mu$[1]	$\theta_2=4N_2\mu$[1]	$\theta_A=4N_A\mu$	$t=t\mu$	$m_1=m_1/\mu$[2]	$m_2=m_2/\mu$[2]
All loci, all individuals	2.76 (2.10–3.61)	0.93 (0.69–1.21)	0.01 (0.01–0.53)	0.61 (0.41–0.80)	0.31 (0.14–0.64)	0.74 (0.34–1.28)
All loci, WY and IL	1.76 (1.20–2.62)	0.98 (0.68–1.36)	0.01 (0.01–0.62)	0.64 (0.41–0.88)	0.25 (0.05–0.62)	0.72 (0.30–1.39)
All loci, WA and LA	2.13 (1.43–3.21)	0.75 (0.51–1.05)	0.06 (0.01–0.54)	0.59 (0.38–0.82)	0.21 (0.04–0.47)	0.52 (0.16–1.12)
Autosomal loci, all individuals	3.22 (2.10–4.89)	0.80 (0.52–1.19)	0.21 (0.07–0.48)	0.28 (0.21–0.39)	0.37 (0.01–1.05)	2.12 (0.87–4.20)
Z-linked loci, all individuals	4.21 (2.32–7.30)	0.40 (0.21–0.70)	0.03 (0.01–0.48)	0.36 (0.21–0.62)	0.16 (0.03–0.48)	0.00 (0.00–0.61)
Z-linked loci, all individuals[3]	4.75 (2.69–8.16)	0.46 (0.25–0.79)	0.50 (0.19–1.09)	0.31 (0.19–0.46)	–	–
Mitochondrial, all individuals	4.82 (2.77–8.36)	3.17 (1.68–5.73)	0.03 (0.03–47.21)	2.67 (0.58–3.40)	–	–

[1] θ_1=*P. cyanea*, θ_2=*P. amoena*.

[2] m_1=introgression from *P. amoena* into *P. cyanea*, m_2=introgression from *P. cyanea* into *P. amoena*.

[3] Parameter estimates when both gene flow parameters were eliminated from analyses.

~700,000–1,370,000 years). Estimated divergence times for the autosomal and sex-linked datasets were ~1,140,000 (90% highest posterior density: ~850,000–1,600,000 years) and ~1,290,000 (90% highest posterior density: ~790,000–1,920,000 years) years ago, respectively. The estimated mitochondrial divergence time between the species using mean corrected cytochrome *b* pairwise divergence data (8.9%) was 4.2 million years (±200,000 years).

Discussion

Quantifying the timing and levels of introgression between diverging populations or species that have not yet attained complete reproductive isolation requires information from multiple unlinked loci. In this study, we explored variation from 20 loci (one mitochondrial, 13 autosomal, and six z-linked) to investigate the speciation history of *P. amoena* and *P. cyanea*. Using the complete dataset, we were able to reject a strict allopatric speciation model in favor of a model that requires postdivergence gene flow between *P. amoena* and *P. cyanea* to explain the observed pattern of genetic variation. Additionally, the data suggest that the majority of introgression between these species has been at autosomal loci, with little or no introgression at z-linked or mitochondrial loci.

Recently, two studies made extensive use of simulated datasets to test the performance of IM and IMa when model assumptions are violated (Becquet and Przeworski 2009; Strasburg and Reiseberg 2009). Overall, IM (Becquet and Przeworski 2009) and IMa (Strasburg and Rieseberg 2010) analyses were robust to assumption violations, but we briefly discuss the impact of three violations: population structure, population growth, and gene flow following secondary contact.

Becquet and Przeworski (2009) found that the presence of geographic structure in the ancestral population causes IM (they did not test IMa) to overestimate ancestral population size. In contrast, Strasburg and Rieseberg (2010) found that the presence of population structure has little impact on demographic parameter estimates. The different findings may be due to the fact that Strasburg and Rieseberg (2010) investigated the impact of structure present in the sampled populations, whereas Becquet and Przeworski (2009) specifically address structure in the ancestral population. However, in both cases, the presence of population structure, in either ancestral or contemporary populations, did not significantly influence estimates of the parameters we were primarily concern with in our study: divergence time and gene flow.

The majority of Tajima's *D* values are negative (Table 1), which suggests recent population growth in both species, a violation of the IMa model. In their simulations, Strasburg and Rieseberg (2010) found that exponential and instantaneous population growth influenced the estimates of current effective population sizes, but not of any other parameters; similar analyses were not performed by Becquet and Przeworski (2009). Therefore, the fact that our datasets violated the constant population size assumption of IMa probably did not change the outcome of any of the likelihood ratio tests we performed.

Becquet and Przeworski (2009) also found that IM estimates nonzero migration (gene flow) under a divergence model characterized by a relatively long period of allopatry followed by a period of secondary contact and gene flow; a similar scenario

was not investigated by Strasburg and Rieseberg (2010). These findings are troubling given the increasing number of studies (including this one) rejecting the null model of allopatric divergence (Geraldes et al. 2008; Nadachowska and Babik 2009). It is unknown if IMa, which we used in this study, suffers from the same problems as IM. Both methods apply the "Isolation-with-Migration" model, but use different Markov chain Monte Carlo methods to estimate the model parameters (Nielsen and Wakeley 2001; Hey and Nielsen 2004, 2007).

The possibility that gene flow following secondary contact after a long period of allopatry can cause IM to produce nonzero gene flow estimates, which could lead to the erroneous rejection of allopatric divergence (Becquet and Przeworski 2009), is germane to our work. If, after an initial divergence time of 3.2N generations, the time since secondary contact was at least 0.8N generations, estimates of gene flow were significantly greater than zero, with an apparent trend toward greater estimates of gene flow with increasing time since secondary contact. If we assume that secondary contact between *P. amoena* and *P. cyanea* began ~6500 years ago (see above), then the ratio of time between the initial divergence and current secondary contact is 0.00625 (6500 years/1,040,000 years), which is considerably less than the ratio in the simulations (0.25 = 0.8N/3.2N). Although we cannot be sure without additional simulations, the available evidence suggests that secondary contact between *P. amoena* and *P. cyanea* has influenced the parameter estimates from IM to a much lesser degree than the patterns of secondary contact simulated by Becquet and Przeworski (2009). Additionally, in an attempt to guard against the possible influence of ongoing hybridization on the parameter estimates, we collected the individuals sampled for this study from well outside the current contact zone (Carling and Brumfield 2008b, 2009), and we were able to reject the strict allopatric speciation model in all cases in which all loci were included in the analysis. This was true even when the analyses involved only the populations (WA and LA) that are separated by more than 3000 km (Table 2).

Although we were able to reject a strict allopatric speciation model, it is more difficult to exclude the possibility of repeated bouts of allopatric isolation and divergence followed by gene flow during past periods of secondary contact, which may have tracked climate cycles during the divergence of these species. Unfortunately, currently available methods are unable to distinguish these more historically complex alternate scenarios and we can only conclude that *P. amoena* and *P. cyanea* speciated despite ongoing gene flow during divergence or despite periodic episodes of allopatric divergence and secondary contact.

Many studies have provided evidence for divergence in the face of gene flow using coalescent-based methods (Won and Hey 2005; Geraldes et al. 2008; Lee and Edwards 2008; Niemiller et al. 2008; Strasburg and Rieseberg 2008; Kane et al. 2009; Nadachowska and Babik 2009; Ross-Ibarro et al. 2009), so many in fact that some authors have argued that parapatric speciation may be common (Nosil 2008). However, few of those studies performed statistical tests to assess the fit of the data to alternate speciation models. On this basis alone, the evidence for parapatric speciation is less than convincing. Additionally, most studies have not adequately addressed the problems associated with the presence of geographic structure as discussed above. Lastly, it is likely that climatic cycles and corresponding habitat changes have provided the opportunity for repeated periods of allopatric divergence and secondary contact. This phenomenon may explain the gene flow patterns we found between *P. amoena* and *P. cyanea*, and could play an important role in the evolutionary history of many closely related taxa. Taken together, we suggest that many more studies that rigorously test alternate speciation models are required before we can estimate the relative frequency of sympatric, parapatric, and allopatric speciation.

Numerous cline-based studies between closely related species have supported a prediction of Haldane's rule (Haldane 1922) that sex-linked loci, and mitochondrial loci in taxa with heterogametic females (like birds), should introgress less than autosomal loci (Hagen and Scriber 1989; Tucker et al. 1992; Carling and Brumfield 2008a). These findings have contributed to the hypothesis that the sex-chromosomes play a disproportionately large role in reproductive isolation and speciation. Unfortunately, cline-based analyses are unable to account for the differences in effective population sizes between mitochondrial, autosomal, and sex-linked loci, which likely contribute to the differences in estimates of cline-shape parameters. As demonstrated in this study, coalescent-based analyses that account for the effective population size differences can be used to investigate patterns of differential introgression.

The speciation history of *P. amoena* and *P. cyanea* appears to be marked by introgression of autosomal loci in the absence of introgression at either mitochondrial or z-linked loci (Fig. 3; Table 3), as predicted by Haldane's rule. Outside of the contact zone, *P. amoena* and *P. cyanea* are reciprocally monophyletic at mitochondrial loci, suggesting no mitochondrial introgression. Further, we could not reject a strict allopatric divergence model for the z-linked loci, which is consistent with a lack of postdivergence gene flow via backcrossing female hybrids, as predicted under Haldane's rule. These results illustrate how coalescent-based methods can be used to elucidate the relative contributions of mitochondrial, autosomal, and sex-linked loci to the overall patterns of divergence and introgression between closely related species. To the best of our knowledge, this is the first study to specifically use a coalescent-based analytical method to test for differences in gene flow estimates across these different marker classes, so we presently cannot know whether this is a general phenomenon among closely related taxa.

These results have implications for our understanding of the role of the late Pleistocene on speciation in North American birds, a topic that has long been debated (Zink and Slowinski 1995; Klicka and Zink 1997; Avise et al. 1998; Johnson and Cicero 2004; Lovette 2005). Some empirical studies have supported the hypothesis that many speciation events were initiated before the Pliocene/Pleistocene boundary (~1,800,000 years ago) and others suggest many divergences were initiated much more recently (many even within the past 250,000 years). Although it is clear that reliable estimates of divergence time require data from many independent loci (Edwards and Beerli 2000), all of these past studies have used divergence times estimated using data from a single locus. By using a calibrated mutation rate of 5.85×10^{-7} substitutions/locus/year, the estimated divergence time (using the all loci, all individual dataset) between *P. amoena* and *P. cyanea* in absolute time is ~1,040,000 years, which is in the middle of the Pleistocene epoch. Divergence time estimates from the autosomal chromosomes and sex chromosomes were similar (~1,140,000 and ~1,290,000 years ago, respectively). Although these divergence time estimates are dependent on accurate calibration rates, taken at face value they suggest *P. amoena* and *P. cyanea* diverged more recently than the Pliocene/Pleistocene boundary. The single divergence time estimated here cannot address the general impact of the Pleistocene glaciations on North American bird speciation, but it provides a template for how such multilocus comparisons might be marshaled across many pairs of sister taxa to address this question about the general timing of avian speciation in North America.

We found that the divergence time estimated from the mitochondrial data was much older than the divergence time estimated from any of the other datasets (Fig. 2; Table 3), even after accounting for the 10-fold difference in mutation rates and the differences in effective population size. Because our mitochondrial data consists of a portion of the protein-coding gene ND3 as well as a portion of the noncoding mitochondrial control region, we did not calculate a mitochondrial divergence time from our data for comparison. However, we did calculate a mitochondrial divergence time using previously published cytochrome *b* sequence data for *P. cyanea* and *P. amoena* (Klicka et al. 2001) and a calibrated rate of cytochrome *b* divergence in birds ($2.1 \pm 0.1\%$ per million years; Weir and Schluter 2008). Using a mean corrected pairwise divergence of 8.9%, *P. cyanea* and *P. amoena* diverged 4.2 million years ago (±200,000 years), again, much older than the divergence time inferred from coalescent analyses of the multilocus dataset.

A similar pattern was found in a recent multilocus phylogeographic study of the red-backed fairy wren, *Malurus melanocephalus* (Lee and Edwards 2008). Likewise, whereas nearly all population-level mitochondrial studies of Neotropical birds have inferred pre-Pleistocene processes in diversification, Brumfield et al. (2008) concluded from a multilocus nuclear analysis that Pleistocene processes accounted for divergence times in the Neotropical bird *Manacus manacus*. In these studies, it is unclear if the discrepancy is due to errors in the autosomal or mitochondrial calibration rates, or even error in both calibrations. Until there are more multilocus investigations of divergence times it is difficult to know if this is a general issue in birds.

In summary, our results suggest divergence between *P. amoena* and *P. cyanea* was characterized by postdivergence gene flow. Whether this pattern is a result of parapatric divergence or repeated periods of allopatric divergence and secondary contact is unknown. Overall, the general pattern is likely driven by relatively high levels of gene flow at autosomal loci, as we could not reject a strict allopatric speciation model for the mitochondrial and sex-linked data. Additionally, our findings illustrate the utility of coalescent-based methods to elucidate the relative contributions of gene flow of mitochondrial, autosomal, and sex-linked loci during divergence and speciation.

ACKNOWLEDGMENTS

We thank S. Birks (University of Washington Burke Museum of Natural History and Culture), A. Capparella (Illinois State University), and M. Westberg (University of Minnesota Bell Museum) for generously providing tissue samples for this project. S. Carling and J. Maley were outstanding field assistants. B. Cramer, R. Dor, J. Feder, J. Maley, J. Petersen, D. Rabosky, A. Townsend, R. Vallender, K. Wagner, and five anonymous reviewers all provided helpful comments. Scientific collecting permits were granted by the Montana Fish, Wildlife and Parks, Wyoming Game and Fish, and the United States Fish and Wildlife Service. All individuals were collected in accordance with Louisiana State University Institutional Animal Care and Use Committee Protocol: 08–025. The map data were provided by NatureServe in collaboration with Robert Ridgely, James Zook, The Nature Conservancy—Migratory Bird Program, Conservation International—CABS, World Wildlife Fund—US, and Environment Canada—WILDSPACE. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University that is partially funded by the Microsoft Corporation. This work was supported by grants from NSF (DEB-0543562 and DBI-0400797 to RTB, DEB-0515981 and DEB-0814277 to IJL, and DEB-0808464 to MDC), as well as by grants from the AMNH Chapman Fund, AOU, Explorer's Club, LSUMNS Birdathon and Prepathon Funds, LSU Department of Biological Sciences, and Sigma-Xi. MDC is currently supported by a Fuller Postdoctoral Fellowship.

LITERATURE CITED

Abrams, M. 1992. Fire and the development of oak forests. Bioscience 42:346–353.

Avise, J. C., D. Walker, and G. C. Johns. 1998. Speciation durations and Pleistocene effects on vertebrate phylogeography. Proc. R. Soc. Lond. B 265:1707–1712.

Axelsson, E., N. G. C. Smith, H. Sundstrom, S. Berlin, and H. Ellegren. 2004. Male-biased mutation rate and divergence in autosomal, z-linked and w-linked introns of chicken and turkey. Mol. Biol. Evol. 21:1538–1547.

Baker, M. C., and J. T. Boylan. 1999. Singing behavior, mating associations and reproductive success in a population of hybridizing Lazuli and Indigo buntings. Condor 101:493–504.

Becquet, C., and M. Przeworski. 2009. Learning about modes of speciation by computation approaches. Evolution 63:2547–2562.

Borge, T., M. T. Webster, G. Andersson, and G. P. Saetre. 2005. Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two Ficedula flycatchers. Genetics 171:1861–1873.

Brumfield, R. T., L. Liu, D. E. Lum, and S. V. Edwards. 2008. Comparison of species tree methods for reconstructing the phylogeny of bearded manakins (Aves: Pipridae: *Manacus*) from multilocus sequence data. Syst. Biol. 57:719–731.

Carling, M. D., and R. T. Brumfield. 2008a. Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal and sex-linked loci across the *Passerina* bunting hybrid zone. Evolution 62:2600–2615.

———. 2008b. Integrating phylogenetic and population genetic analyses of multiple loci to test species divergence hypotheses in Passerina buntings. Genetics 178:363–377.

———. 2009. Speciation in *Passerina* buntings: introgression patterns of sex-linked loci identify a candidate gene region for reproductive isolation. Mol. Ecol. 18:834–847.

Charlesworth, B., J. A. Coyne, and N. Barton. 1987. The relative rates of evolution of sex chromosomes and autosomes. Am. Nat. 130:113–146.

Coyne, J. A., and H. A. Orr. 1989. Two rules of speciation. Pp. 180–207 *in* D. Otte and J. A. Endler, eds. Speciation and its consequences. Sinauer Associates, Inc., Sunderland, MA.

———. 2004. Speciation. Sinauer Associates, Inc., Sunderland, MA.

Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution 54:1839–1854.

Ellegren, H. 2007. Molecular evolutionary genomics of birds. Cytogenet. Genome Res. 117:120–130.

Emlen, S. T., J. D. Rising, and W. L. Thompson. 1975. Behavioral and morphological study of sympatry in Indigo and Lazuli buntings of the Great Plains. Wilson Bull. 87:145–179.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin ver 3.0: an integrated software package for population genetics data analysis. Evol. Bioinform. Online 1:47–50.

Geraldes, A., P. Basset, B. Gibson, K. L. Smith, B. Harr, H. T. Yu, N. Bulatova, Y. Ziv, and M. W. Nachman. 2008. Inferring the history of speciation in house mice from autosomal, X-linked, Y-linked and mitochondrial genes. Mol. Ecol. 17:5349–5363.

Graur, D., and W. H. Li. 2000. Fundamentals of molecular evolution. Sinauer Associations, Inc., Sunderland, MA.

Greene, E., V. R. Muehter, and W. Davison. 1996. Lazuli Bunting. *The Birds of North America Online* (A. Poole, ed.) Ithaca: Cornell Lab of Ornithology; Retrieved from The Birds of North America Online database: http://bna.bird.cornell.edu.

Hagen, R. H., and M. Scriber. 1989. Sex-linked diapause, color, and allozyme loci in *Papilo glaucus*: linkage analysis and significance in a hybrid zone. J. Hered. 80:179–185.

Haldane, J. B. S. 1922. Sex ratio and unisexual sterility in animal hybrids. J. Genet. 12:101–109.

Harrigan, R. J., M. E. Mazza, and M. D. Sorenson. 2008. Computation vs. cloning: evaluation of two methods for haplotype determination. Mol. Ecol. Res. 8:1239–1248.

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.

———. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA 104:2785–2790.

Johnson, N. K., and C. Cicero. 2004. New mitochondrial DNA data affirm the importance of Pleistocene speciation in North American birds. Evolution 58:1122–1130.

Kane, N. C., M. G. King, M. S. Barker, A. Raduski, S. Karrenberg, Y. Yatabe, S. J. Knapp, and L. H. Reiseberg. 2009. Comparative genomic and population genetic analyses indicate highly porous genomes and high levels of gene flow between divergent *Helianthus* species. Evolution 63:2061–2075.

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

Klicka, J., A. J. Fry, R. M. Zink, and C. W. Thompson. 2001. A cytochrome-b perspective on *Passerina* bunting relationships. Auk 118:611–623.

Kroodsma, R. L. 1975. Hybridization in buntings (*Passerina*) in North-Dakota and eastern Montana. Auk 92:66–80.

Lee, J. Y., and S. V. Edwards. 2008. Divergence across Australia's Carpentarian barrier: statisitical phylogeography of the Red-Backed Fairy Wren (*Malurus melanocephalus*). Evolution 62:3117–3134.

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

Nadachowska, K., and W. Babik. 2009. Divergence in the face of gene flow: the case of two newts (Amphibia: Salamandridae). Mol. Biol. Evol. 26:829–841.

NatureServe. 2006. NatureServe Explorer: an Online Encyclopedia of Life, Version 6.1. http://www.natureserve.org/explorer.

Nei, M. 1987. Molecular evolutionary genetics. Columbia Univ. Press, New York City.

Nei, M., and W. H. Li. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA 76:5269–5273.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158:885–896.

Niemiller, M. L., B. M. Fitzpatrick, and B. T. Miller. 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol. Ecol. 17:2258–2275.

Nosil, P. 2008. Speciation with gene flow could be common. Mol. Ecol. 17:2103–2106.

Payne, R. B. 2006. Indigo Bunting. *The Birds of North America Online* (A. Poole, ed.) Ithaca: Cornell Lab of Ornithology; Retrieved from The Birds of North America Online database: http://bna.birds.cornell.edu.

Ridgely, R. S., T. F. Allnutt, T. Brooks, D. K. McNicol, D. W. Mehlman, B. E. Young, and J. R. Zook. 2003. Digital distribution maps of the birds of the Western hemisphere, version 1.0. NatureServe, Arlington, VA.

Ross-Ibarro, J., M. Tenaillon, and B. S. Gaut. 2009. Historical divergence and gene flow in the genus *Zea*. Genetics 181:1399–1413.

Rozas, J., J. C. Sanchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Saetre, G. P., T. Borge, K. Lindroos, J. Haavie, B. C. Sheldon, C. Primmer, and A. C. Syvanen. 2003. Sex chromosome evolution and speciation in *Ficedula* flycatchers. Proc. R. Soc. Lond. B 270:53–59.

Sibley, C. G., and L. L. J. Short. 1959. Hybridization in the buntings (*Passerina*) of the Great Plains. Auk 76:443–463.

Stephens, M., and P. Donnelly. 2003. A comparison of Bayesian methods for haplotype reconstruction from population genotype data. Am. J. Hum. Genet. 73:1162–1169.

Stephens, M., N. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

Strasburg, J. L., and L. H. Rieseberg. 2008. Molecular demographic history of the annual sunflowers *Helianthus annus* and *H. petiolarus*-–large effective population sizes and rates of long-term gene flow. Evolution 62:1936–1950.

———. 2010. How robust are "Isolation with Migration" analyses to violations of the IM model? A simulation study. Mol. Biol. Evol. 27:297–310.

Tajima, F. 1983. Evolutionary relationships of DNA sequences in finite populations. Genetics 105:437–460.

Tucker, P. K., R. D. Sage, J. Warner, A. C. Wilson, and E. M. Eicher. 1992. Abrupt cline for sex-chromosomes in a hybrid zone between two species of mice. Evolution 46:1146–1163.

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

Won, Y. J., and J. Hey. 2005. Divergence population genetics of chimpanzees. Mol. Biol. Evol. 22:297–307.

Zink, R. M., and J. B. Slowinski. 1995. Evidence from molecular systematics for decreased avian diversification in the Pleistocene Epoch. Proc. Natl. Acad. Sci. USA 92:5832–5835.

Associate Editor: J. Feder

Supporting Information

The following supporting information is available for this article:

Table S1. Sampling localities for all individuals.

Table S2. PCR primers and amplification conditions for all loci amplified.

Table S3. GenBank Accession numbers for sequences previously deposited.

Supporting Information may be found in the online version of this article.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.