ORIGINAL ARTICLE

DIVERGENCE ACROSS AUSTRALIA'S CARPENTARIAN BARRIER: STATISTICAL PHYLOGEOGRAPHY OF THE RED-BACKED FAIRY WREN (*MALURUS MELANOCEPHALUS*)

June Y. Lee[1,2] and Scott V. Edwards[1]

[1]*Museum of Comparative Zoology and Department of Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138*

[2]*E-mail: jylee@fas.harvard.edu*

Multilocus analysis of phylogeography and population history is a powerful tool for understanding the origin, dispersal, and geographic structure of species over time and space. Using 36 genetic markers (29 newly developed anonymous nuclear loci, six introns and one from mitochondrial DNA, amounting to over 15 kb per individual), we studied population structure and demographic history of the red-backed fairy wren *Malurus melanocephalus*, a small passerine distributed in the northern and eastern part of Australia across the Carpentarian barrier. Analysis of anonymous loci markers revealed large amounts of genetic diversity ($\pi = 0.016 \pm 0.01$; average number of SNPs per locus = 48; total number of SNPs = 1395), and neither nuclear nor mitochondrial gene trees showed evidence of reciprocal monophyly among Cape York (CY), Eastern Forest (EF), and Top End (TE) populations. Despite traditional taxonomy linking TE and CY populations to the exclusion of EF, we found that the CY population is genetically closer to the EF population, consistent with predicted area cladograms in this region. Multilocus coalescent analysis suggests that the CY population was separated from the other two regions approximately 0.27 million years ago, and that significant gene flow between the ER and the CY populations (~2 migrants per generation) suggests geographic continuity in eastern Australia. By contrast, gene flow between the CY and the TE populations has been dampened by divergence across the Carpentarian barrier.

KEY WORDS: Anonymous loci, Carpentarian Barrier, historical demography, *Malurus melanocephalus*, phylogeography.

Understanding the geographic distribution of genealogical lineages and the underlying processes generating these patterns has been of great interest for evolutionary biologists. Phylogeography focuses on these distributions within and among closely related species (Avise et al. 1987; Arbogast and Kenagy 2001), combining both spatial and phylogenetic components, and thereby contributes to the study of micro- and macroevolution along with fields such as systematics and paleontology (Avise et al. 1987; Avise 1998, 2000; Hewitt 2001). Since the advent of the polymerase chain reaction (PCR), the use of DNA sequence variation to reconstruct phylogeographic histories has accelerated the development of the field, relying heavily in its first two decades on mitochondrial DNA (mtDNA) because of its high mutation rate, small effective population size (compared with autosomal nuclear DNA), haploidy, and its putative lack of recombination, although reports of mitochondrial recombination have become more common recently (Hagelberg et al. 2000; Pakendorf and Stoneking 2005; Tsaousis et al. 2005). However, because mtDNA is usually inherited maternally, it may not exhibit genealogical patterns that are representative of the entire population history, especially

when there is a sex bias in fitness or dispersal behavior (Hare 2001). More importantly, such single-locus approaches are liable to errors due to stochastic lineage sorting (Kuo and Avise 2005) and we cannot estimate the coalescent variance for parameters such as gene flow, divergence times and population sizes with a single locus (Edwards and Beerli 2000). In addition, natural selection has been increasingly shown to influence the trajectory of mtDNA lineages, in some cases placing a ceiling on diversity that appears to be eroded repeatedly by selective sweeps (Bazin et al. 2006). Although mtDNA will surely continue to be a mainstay of phylogeography, the relative power of nuclear versus mitochondrial variation for unraveling phylogeographic histories, as well as the consistency of scenarios presented by the two genomic compartments in natural populations, is still debated (Edwards et al. 2005; Brito and Edwards 2008; Zink and Barrowclough 2008).

Recently, single nucleotide polymorphisms (SNPs) and "resequencing" of nuclear DNA have arisen as important methods that may help refine comparative phylogeography and historical demography of populations (Nickerson et al. 1997; Di Rienzo et al. 1998; Wakeley et al. 2001; Brumfield et al. 2003). One marker type that has received little attention in studies of nonhuman animals is anonymous nuclear loci (Karl and Avise 1993; Brito and Edwards 2008). Anonymous nuclear markers are noncoding regions of the genome, randomly collected and presumably dispersed across the chromosomes. Because methods for developing anonymous loci can yield a nearly unlimited number of random segments of nuclear DNA, they provide a better representation of variation across the whole genome than does any one marker. In addition, via resequencing approaches they can minimize ascertainment bias, and help evaluate the extent of coalescent variation in datasets (Karl and Avise 1993; Carstens et al. 2005; Jennings and Edwards 2005). Even though nuclear DNA is known to have a lower mutation rate than mitochondrial DNA, in fact, few studies have compared levels of noncoding mitochondrial and nuclear DNA within species to check the almost universal assumption that mtDNA exhibits higher variation than nuclear DNA, and some studies that have done so have found lower levels of mtDNA variation, presumably a result of selection (Bazin et al. 2006; Bensch et al. 2006). The multiplicity of gene histories available in the nuclear genome can provide a strong signal for inferring demographic histories, even if the gene histories themselves appear heterogeneous (Carstens and Knowles 2007). Furthermore, the development of increasingly powerful coalescent-based methods makes it straightforward to test hypotheses and infer demographic parameters from multilocus data (Nielsen and Wakeley 2001; Hey and Machado 2003). Such methods, combined with multilocus data and frequently with simulation approaches designed to delimit the parameter space of a pri-

ori phylogeographic hypotheses, mark a new phase of multilocus phylogeographic research known as "statistical phylogeography" (Knowles and Maddison 2002; Knowles 2004).

BIOGEOGRAPHY OF AUSTRALIAN BIRDS

Due to its long history of isolation as a continent (Cracraft 2001), Australia has provided abundant opportunities for studying biogeography. By comparing general geographic distributions of Australian birds, early studies identified multiple codistributed songbirds ranges and areas of endemism around the continent (Keast 1961; Ford 1974, 1987a; Cracraft 1986). Some geographic patterns across the continent have been explained by selective responses to environmental gradients without geographic isolation (Wooller et al. 1985; Hughes et al. 2001, 2002; Toon et al. 2003). In addition, various combinations of differentiation in isolation and population expansion also have been proposed for populations now found in the center of the continent as well as for populations around the periphery (Keast 1961; Ford 1974, 1987a; Schodde 1982; Joseph and Wilke 2007). Many species show congruent distributions with major or minor barriers that have generally been assumed to be formed in the Pleistocene (Cracraft 1986; Ford 1987b; Jennings and Edwards 2005), but still we do not have a solid time scale for many of these events. Determining further time scales for the Australian biota is important for understanding global patterns of bird biogeography, particularly in light of the controversy surrounding the role of the Pleistocene in the evolution of the North American bird fauna (Avise and Walker 1998; Klicka and Zink 1999; Johnson and Cicero 2004).

One biogeographic barrier that has played an important role in Australian biotic history is the Carpentarian Barrier, a prominent barrier in northern Australia extending from the southernmost shore of the Gulf of Carpentaria. This barrier is generally assumed to have arisen as a result of the fluctuations of sea level over time as well as a steep climate and habitat gradient across this area during the last glacial maxima (Nix and Kalma 1972; Smart 1977; Cracraft 1986). The Carpentarian barrier is roughly 150-km wide depending on the taxon being considered and is a semi-arid region extremely poor in vegetation. Using cladistic analysis of morphology of seven avian lineages, Cracraft (1986) showed that in many lineages the northern and eastern biota were separated first by this barrier and then by secondary barriers along the northern, eastern, and western coasts (see Fig. 1). Several studies of various animal taxa across this barrier are consistent with the above explanation (Keast 1961; MacDonald 1969; Ford 1977; Edwards 1993; Ford and Blair 2005; Jennings and Edwards 2005), although the timing of divergence across this barrier has not been widely addressed. Using anonymous loci, Jennings and Edwards (2005) estimated divergence of phenotypically well-diverged grassfinches (*Poephila*) across the Carpenterian barrier
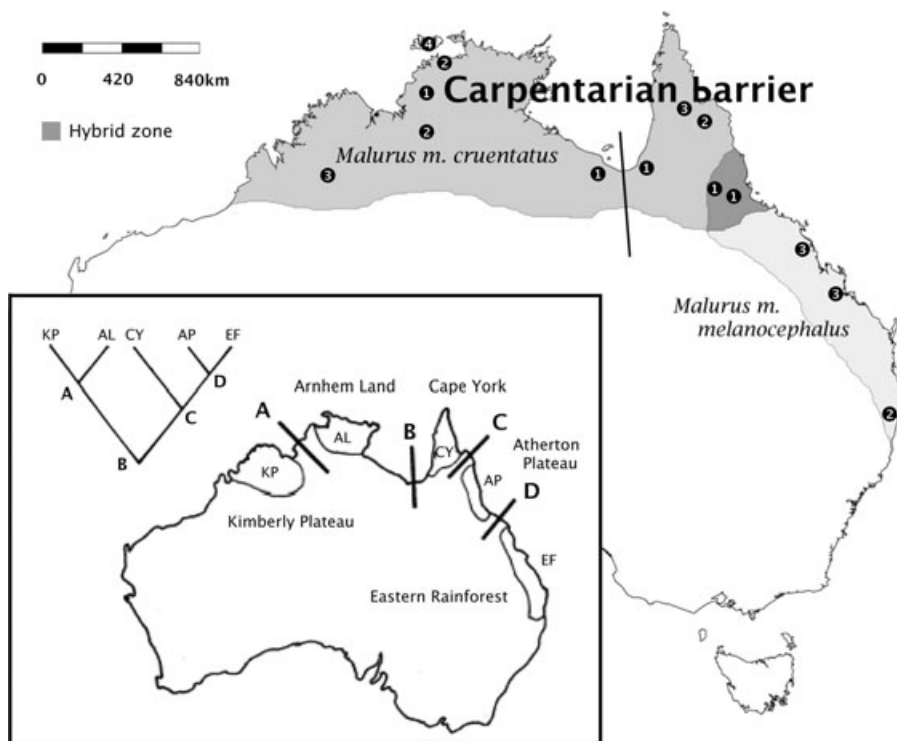
Figure 1. Range map of two red-backed fairy wren subspecies and location of the Carpentarian Barrier. The inset shows the historical interrelationships of five areas of endemism in the northern and eastern part of Australia as summarized by Cracraft (1986). Black circles represent the location of samples, and numbers on circles show the number of samples collected at that location.

to be about 600–700 kya, a value well within the Pleistocene and one that contrasts with the older divergences claimed for North America (Klicka and Zink 1999; Johnson and Cicero 2004).

The red-backed fairy wren (*Malurus melanocephalus*) is a common, sedentary, and territorial passerine in the genus *Malurus*, which is well known for its high degree of sexual dimorphism, promiscuity, and prevalence of cooperative breeding (Karubian 2002; Webster et al. 2008). As the smallest member of the genus (weight ranging from 6 to 8 g), the red-backed fairy wren is distributed across the northern part (north of latitude 20°S) as well as the east coast of Australia, primarily inhabiting open forest and savannah (Fig. 1). Given its large distribution and its poor capability of long-distance flight, the red-backed fairy wren is a good organism to study the effects of geographic barriers on population structure. Two subspecies, *M. m. cruentatus* and *M. m. melanocephalus*, were identified as sharing their boundaries on the southern part of the Atherton Plateau (Schodde and Mason 1999). This taxonomy is based mainly on male plumage color on the back: scarlet and crimson for *M. m. melanocephalus* and *M. m. cruentatus*, respectively. Such a geographic pattern suggests a role for a minor Pleistocene-originated geographic barrier, but contradicts the primary function of the Carpentarian Barrier as a major cause of shaping species distributions in northern Australia as predicted by area cladograms (see Fig. 1). Here we employ multilocus genetic tools to test this morphological taxonomy, and

use our multilocus dataset to infer the population structure, timing, and demographic history of this charismatic species.

Materials and Methods
SAMPLING
Our sampling and data-collecting efforts emphasized loci over individuals, and we focused on broad regions within Australia rather than detailed geographic sampling. We obtained 30 red-backed fairy wren samples from two institutions: 10 specimens from the Burke Museum, Seattle, WA (collected over several expeditions lead by SVE), and 20 specimens from the ANWC (Australian National Wildlife Collection, Commonwealth Scientific and Industrial Research Organization in Canberra), Australia (Supporting Table S1). Each sample was assigned to one of three major biogeographic regions according to their specific localities. All samples from the western side of the Carpentarian Barrier were assigned to Top End (TE: 14 individuals), and the rest of the samples were assigned either to Cape York (CY: 8 individuals) or to Eastern Forest (EF: 8 individuals). Because no hybrids have been reported from south of the Burdekin Barrier in northeast Australia (barrier D in Fig. 1), we used this barrier as the subspecies boundary that separates the CY and EF populations. Among eight samples assigned to the CY population, two individuals (sample #12 and #20) were collected from the

putative hybrid zone (Supporting Table S1). Eight white-winged fairy wrens (*M. leucopterus*) were included in the analysis so as to understand how frequently genes are shared between fairy wren sister species (Christidis and Schodde 1997). We also used three splendid fairy wrens *M. splendens*, three variegated fairy wrens *M. lamberti*, three striated grasswrens *Amytornis striatus*, and two rufous-crowned emu-wrens *Stipiturus ruficeps* as outgroups to root reconstructed gene trees, focusing on whichever taxon provided reliable sequence of a given locus (Supporting Table S1).

DEVELOPMENT AND EVALUATION OF ANONYMOUS NUCLEAR LOCI

We applied the same technique used by Jennings and Edwards (2005) to develop nuclear anonymous markers. We first used a female red-backed fairy wren specimen (number 60736, Burke Museum) as a source of genomic DNA. Genomic DNA was extracted from frozen pectoral muscles using a standard genomic DNA extraction kit (Qiagen, Valencia, CA) and sheared into small ($1 \sim 2kb$) fragments with a Hydroshear (Genomic Solution, Ann Arbor, MI). Sheared DNA was electrophoresed through a 1% agarose gel, and fragments within target range ($1 \sim 2kb$) were excised to obtain fragments that would ultimately minimize the number of sequencing reactions required to sequence across each locus, yet maximize sequence available for reconstructing gene trees. After blunting the ends of the sheared DNA and dephosphorylating them using a TOPO-shotgun cloning kit (Invitrogen, Carlsbad, CA), these fragments were ligated into pUC19 plasmid vectors and then transformed into chemically competent *Escherichia coli* cells by the heat-shock method. Clones were plated on agar plates containing ampicillin and kept for 18 h in an incubator at 37°C. Clones were picked randomly with toothpicks, and then sequenced using M13 vector primers. Because we aimed for markers that would reflect the variability in the noncoding subgenome, all sequences were screened bioinformatically to minimize capture of coding sequences. First, we searched for any significant matches in BLAST (Basic Local Alignment Search Tool) for each sequence. We also checked to determine if they contain any open reading frames (ORFs) using the NCBI ORF finder (http://www.ncbi.nlm.nih.gov/projects/gorf/).

Because we developed anonymous loci markers from random segments of nuclear DNA, it is possible that some of loci might represent paralogs, in which case we would expect to see much higher genetic diversity. To detect possible paralogs, we blasted sequences from all anonymous loci to the trace archive database of zebra finch *Taeniopygia guttata* (http://www.ncbi.nlm.nih.gov/Traces), which is the closest species whose genome is currently available. Because the zebra finch genome was sequenced sixfold, we would expect to see a similar number of matches in our search. Most of our anonymous loci had on average eight matches, but two loci (*Mame*-AL17 and *Mame*-AL30) matched with 100 and 71 clones in zebra finch trace archive. As candidate paralogous loci, these loci were cloned using one individual, and 30 clones for each locus were picked randomly. Whereas only two alleles were found in *Mama*-AL17 locus as expected, we found six different allele types in *Mame*-AL30. This suggests that this segment of DNA may exist in more than one place in the genome or may result from high recombination within the locus. Thus, we decided to exclude this locus in our further analyses. In addition, the fact that we did not find any double bands in PCR products further reduces the probability of paralogs. After satisfying ourselves that our loci did not encode known paralogs or protein-coding regions, we designed PCR primers flanking each sequence (Supporting Table S2).

MITOCHONDRIAL AND INTRON MARKERS AND DETERMINATION OF HAPLOTYPES

In addition to anonymous nuclear markers, six intron markers (AB4, aldolase B intron 4; α-globin2, α-globin intron 2; GAPDH, glyceraldehyde-3-phosphate dehydrogenase intron 11; GTP, phosphenolpyruvate carboxykinase intron 9; RI2, rhodopsin intron 2; TGFβ2, Transforming Growth Factor-β2 intron 5) were developed using previously published primers (Waltari and Edwards 2002; Sorenson et al. 2004; McCracken and Sorenson 2005). Part of the mitochondrial DNA (ND2, NADH dehydrogenase subunit 2) was also analyzed using available primers (Sorenson et al. 1999) to compare with characteristics of the nuclear loci markers (Supporting Table S2). After amplifying with these primers, the set of amplicons for 30 red-backed fairy wrens and eight white-winged fairy wrens was sequenced directly in both directions. We found a substantial number of indels in our anonymous loci, often in heterozygous condition (see Results). Sequences containing indels were separated manually by visually subtracting chromatogram peaks upstream of the indel in the reverse primer sequence from the double peaks downstream of the indel in the forward primer sequences. Both sequences were then reverse-complemented, and the subtraction was repeated in the alternative direction as described (Dolman and Moritz 2006). However, because this method was not applicable to sequences that contain more than one indel due to the complexity of their chromatograms, we excluded nine sequences that showed any evidence of multiple indels (e.g., overlapped double peaks in chromatograms). Analysis of indels using this approach has the additional advantage of allowing determination of allele phase directly from the chromatograms, rather than being inferred statistically. To estimate haplotypes of loci, the software PHASE2.1.1 (Stephens et al. 2001) was used to resolve haplotypes of genotypes that were heterozygous at multiple sites. Sequences from outgroups were aligned with these estimated haplotypes using MacClade 4.0 (Maddison and Maddison 2000).

ESTIMATION OF GENE TREES

Because patterns of substitution may differ between different regions of the genome, we used the program MrModeltest 2.2 (Nylander 2004) to infer adequate DNA substitution models for each locus. Based on these estimated DNA substitution models, gene trees were reconstructed using MrBayes version 3.0 (Huelsenbeck and Ronquist 2001), with which we obtained Bayesian posterior probabilities from five million MCMC cycles with a sample frequency of 100 and a burn-in period of 100,000 generations.

POPULATION STRUCTURE

Given that the number of distinct alleles (inferred haplotypes) per locus was high (33 ± 11) relative to the number of allele copies sequenced per locus (56 ± 4), F_{ST} was measured in Arlequin 2.0 (Excoffier and Schneider 2005) not using allele frequencies alone but with a genetic distance that was estimated for each locus, based on DNA substitution models we obtained from MrModeltest. We also estimated another statistic (S_{NN}) that was proposed as a more powerful method for handling loci with high haplotype diversity (Hudson 2000). S_{NN}, referred to as the nearest-neighbor statistic, is a measure of how often the "nearest neighbors" of sequences are from the same locality or geographic area. If populations are strongly structured, S_{NN} is expected to be close to one; when three subpopulations are unstructured (as in our case), S_{NN} is expected to be approximately one-third. The significance of these two statistics was evaluated by 1000 simulations in which the locality of samples was assigned randomly, but in the same proportions as our original dataset. In addition, we used the reconstructed gene trees to evaluate the degree of genetic differentiation. After the geographic locality of samples was assigned to the tips of trees and then the minimum number of character state changes (known as Slatkin's s) was counted by parsimony, treating each locality as a multi-state unordered character (Slatkin and Maddison 1989). To check the significance of our observed value of s, we calculated the statistic for 1000 randomly generated gene trees with the same proportions among sampled localities and compared the resulting distribution with our data. This test was performed to detect genetic differentiation within red-backed fairy wren as well as to assess how frequently genes are being shared between the red-backed and white-winged fairy wrens. To further assess genetic structure within the red-backed fairy wren, we employed the genetic clustering program STRUCTURE (Pritchard et al. 2000), which can detect genetic groups without a priori geographic information by assigning individuals to populations within which there is approximate Hardy–Weinberg equilibrium.

DEMOGRAPHIC HISTORY

To infer historical demography of the red-backed fairy wren, we used the computer program IM, which assumes an "isolation with migration" model (Hey and Nielsen 2004). We employed this program for two different divergence events (TE vs. CY and CY vs. EF), given that the TE population is not geographically adjacent to the EF population. For each divergence event, six population parameters scaled by mutation rate (μ, geometric mean of mutation rate per year per locus) were estimated: the neutral mutation parameter for the ancestral population (θ_A, $\theta = 4N\mu$) and for the two daughter populations (θ_1, θ_2); the divergence time between the two descendant populations (t); and forward and backward migration rates between daughter populations (m_{12}, m_{21}). The major assumption underlying this program is that genetic markers must be independent of selection and recombination, and that mating within ancestral and descendant populations is random. Because recombination and high mutation rate (especially multiple point-mutations at one site) can cause similar patterns in DNA sequence, we decided to apply the infinite site model to our nuclear markers so that we can exclude the effect of recombination, rather than the HKY model in which sequences that result from recombination would be interpreted as result of high mutation rate. To see if our markers satisfy the first of these assumptions, we calculated Tajima's D for neutrality and performed the four-gamete test to detect evidence of recombination, using DNAsp 4.0 (Rozas and Sanchez 2003). For loci that did not pass the four-gamete test, the locus was divided into a number of sequence segments such that any evidence of recombination was not detected. Then, we selected the longest sequence segment from each of these sets so as to maximize the information from each locus. The average length of each locus was 170 bp (151 bp) for anonymous nuclear markers (introns), which contained 13 (8) segregating sites on average. After a few short preliminary runs (five million steps) were made to optimize prior boundaries for the six parameters, the final simulation for each divergence was carried out with a different random seed, four different chains, 20 million steps, and burn-in of two million steps.

To understand how the number of loci affects the uncertainty of demographic parameters in the IM program, we plotted peak posterior distribution estimates as well as the lower and upper bounds of estimated 95% highest posterior densities (HPD) for each demographic parameter versus the number of loci. For a given number of loci, we ran the IM program with five randomly selected subsets of loci with five million steps and a burn-in of 500,000. Given that the six demographic parameters were not mutually independent, we discarded a particular run whenever any of the six parameters failed to converge to stationary distribution. Because none of the runs with fewer than four loci satisfied this condition, we included only subsets with more than five loci. All valid estimates were then averaged over the number of valid runs. To avoid biases due to sampling artifacts, we repeated this process five times and calculated the mean (over 5) for each number of loci.

Results

CHARACTERISTICS OF ANONYMOUS LOCI

Among 60 sequenced clones from our genomic library, seven sequences had significant matches (e value $< 10 \times 10^{-10}$) with deposited sequences in GenBank (*M. splendens* 12S rRNAs (2), *Malurus cyaneus* cytochrome b (1), *Cyanopica cyanus* control region of mtDNA (1), *Gallus gallus* Stearoyl-CoA Desaturase (SCD) mRNA (1), *G. gallus* Serum- & Glucocorticoid-induced kinase mRNA (1), and Avian Androgen Receptor genomic region (1)). However, no sequences were predicted to contain the ORFs. Discarding four more loci that did not amplify well, we randomly selected 30 loci but we had to exclude an additional locus that appeared to be paralogous (see details in Methods). Thus, we worked with 29 anonymous markers, amounting to 12,831 bp per individual. When we include the six introns (total 2173 bp) and a part of ND2 region (467 bp), in total we sequenced 15,471 bp from each individual (Table 1). The anonymous loci contained 48 SNPs on average (standard deviation = 21) and exhibited substantial diversity, yielding an average nucleotide diversity (π) of 0.016 ± 0.010 (Table 1) which is 10 times the typical level found, for example, in noncoding regions of the human genome (Zhao et al. 2000). This nucleotide diversity as well as its variance was significantly higher than introns ($\pi = 0.009 \pm 0.002$; $t_{33} = -3.384$, $P = 0.0018$, Fig. 2, Table 1). Except for two anonymous loci (*Mame*-AL10 and *Mame*-AL14), indels were found in all loci (mean frequency per locus across all anonymous loci = 4 ± 2; range: 0–11; Fig. 3A) with an average size of 6 bp (standard deviation = 10 bp, Fig. 3B). In addition, we found a positive correlation between the number of indels per locus and the number of SNPs (Pearson's $r = 0.66$, $N = 29$, $P < 0.01$; Fig. 3C).

RECONSTRUCTION OF GENE GENEALOGIES AND POPULATION STRUCTURE

Once we obtained two successfully converging runs from Mr-Bayes, gene trees were visualized in PAUP4.0 (Swofford 2003). Figure 4 shows some examples of gene trees we reconstructed using anonymous loci markers and the mitochondrial ND2 marker. None of the gene trees, including that for mtDNA, showed reciprocal monophyly of the three red-backed fairy wren populations, although the ND2 gene tree did possess the lowest value of Slatkin's s (s = 6; Fig. 4). For all but three loci, F_{ST} among the three populations was substantial, ranging from 0.028 to 0.499 for nuclear loci (mean = 0.119), and 0.553 for mtDNA (Fig. 5A). S_{NN} (mean across loci = 0.571 ± 0.118, Fig. 5B) also supported this result because a value above 0.333 is predicted only if there is detectable genetic structure among three populations. Figure 5C presents the distributions of Slatkin's s obtained from 36 reconstructed gene trees and from 1000 simulated random gene trees. The distribution of observed values was significantly smaller than that of the simulations ($t_{127} = -16.0149$, $P < 0.001$), suggesting that our observed gene trees are more structured than random gene trees. When we conducted this test between red-backed and white-winged fairy wrens, we found much greater differentiation than within species ($t_{126} = -22.0581$, $P < 0.0001$, Fig. 5D). Strong clustering by geographic region within red-backed fairy wrens was detected by STRUCTURE, which suggested that two genetically distinguishable populations separated by the Carpentarian Barrier was the best explanation of the data (Fig. 5E).

POPULATION DEMOGRAPHY

Figure 6 and Table 2 present the results for two IM runs (TE vs. CY and CY vs. EF). To convert model parameter estimates from the IM program into demographic quantities (population size, migration rate, and divergence time), we started with a previously suggested substitution rate of 1.5×10^{-9} substitutions/site/year for avian introns (Ellegren 2007). However, because the genetic diversity of anonymous loci in our study was significantly higher than that for introns, we applied an adjusted substitution rate of 2.56×10^{-9} substitutions/site/year to anonymous loci, given the amount of difference in genetic diversity (π) between anonymous loci and introns ($1.5 \times (0.015903/0.009292) \approx 2.56$). From both runs, the effective population size (N_e) of the CY population was estimated to be 260,000, assuming that the generation time of red-backed fairy wrens is 2 years (Rowley and Russell 1997). The N_e of the EF population ($\approx$275,000) was similar to that of the CY population, whereas the N_e of the TE population ($\approx$420,000) was estimated to be much larger than the other two populations. The two estimates of divergence time were similar to each other (Table 2), occurring between 271,000 and 281,000 years ago. Even though the IM estimated nonzero migration from CY to the other two populations (0.32, 0.16 migrants per generation to EF and to TE, respectively), the estimates from the TE and EF populations into the CY population were higher (0.81 and 2.06 migrants per generation from TE and EF populations, respectively).

SHIFTS IN RESOLUTION WITH NUMBERS OF LOCI

We wanted to determine how the uncertainty of our estimates of parameters in the population history changed with the number of loci sampled. To minimize any bias that could be caused by violating one of the IM program assumptions, namely that there should not be any other population exchanging genes with the sampled populations, we merged the CY population and EF populations, and then ran the IM program with the TE population. The estimates of two divergence events were similar, and so we proceeded to run IM repeatedly with different numbers of loci randomly sampled from our dataset. The results indicate that the variance for all six parameters decreases rapidly from five to 15 loci, and the rate of decrease is reduced thereafter (Fig. 7). Intriguingly, most of the parameter estimates remained stable regardless of the number of

Table 1. Descriptive statistics for the 36 loci used in our study. Sequence length (L) includes alignment gaps, and the number of segregating sites (S) is shown with its percentage given the length of the locus. Statistical significance of Tajima's *D* test (*D*) is indicated with an asterisk at the *P<0.05 level. The lengths of indels are listed as they occurred in each locus.

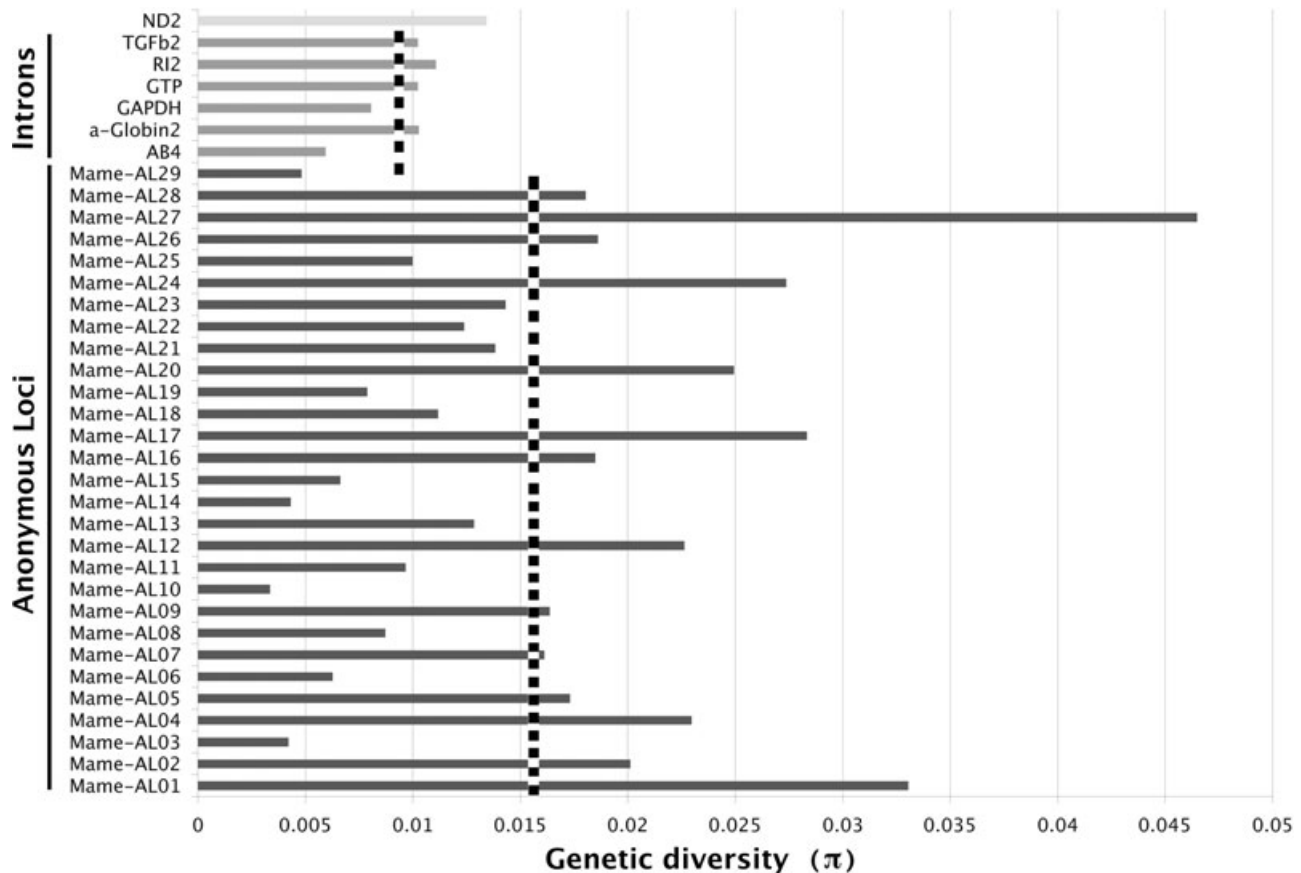| Locus	L	S(%)	π	Top End		Cape York		Eastern Forest		Indel length
				π	D	π	D	π	D	
Anonymous Loci										
Mame-AL01	337	60 (17.8)	0.033	0.031	0.939	0.035	0.940	0.035	0.341	1, 1
Mame-AL02	451	65 (14.4)	0.020	0.022	−0.383	0.016	−0.362	0.015	0.165	1, 2, 1, 1
Mame-AL03	453	22 (4.9)	0.004	0.004	−0.737	0.005	−0.488	0.003	0.058	1,4
Mame-AL04	579	73 (12.6)	0.023	0.022	−0.070	0.023	0.847	0.020	1.266	48, 4, 1, 4
Mame-AL05	553	61 (11.0)	0.017	0.007	−1.279	0.019	0.601	0.022	2.282	1, 1, 1, 1, 4
Mame-AL06	445	39 (8.8)	0.006	0.007	−1.713*	0.005	−0.934	0.006	−0.803	3, 1, 7, 1
Mame-AL07	494	59 (11.9)	0.016	0.014	−0.921	0.015	0.016	0.016	0.350	28, 1, 3, 5, 1, 5, 2, 1
Mame-AL08	467	44 (9.4)	0.009	0.007	−1.203	0.010	−0.726	0.009	0.098	1, 5, 3, 1, 7
Mame-AL09	495	64 (12.9)	0.016	0.017	−0.791	0.013	−0.835	0.016	0.661	7, 1, 4, 12, 1
Mame-AL10	229	4 (1.7)	0.003	0.004	−0.008	0.002	−1.294	0.004	0.341	N/A
Mame-AL11	396	34 (8.6)	0.010	0.014	−0.661	0.007	1.090	0.005	−0.952	1
Mame-AL12	457	59 (12.9)	0.023	0.023	0.582	0.023	0.744	0.024	0.627	2, 4, 1, 1
Mame-AL13	437	50 (11.4)	0.013	0.014	−0.467	0.010	−1.042	0.011	−0.579	1, 19, 1, 7
Mame-AL14	389	23 (5.9)	0.004	0.005	−1.287	0.003	−0.979	0.004	−0.507	N/A
Mame-AL15	433	18 (4.2)	0.007	0.004	−0.668	0.007	0.231	0.001	−0.498	35, 2, 2
Mame-AL16	419	54 (12.9)	0.018	0.018	−0.580	0.016	0.638	0.019	0.381	5, 4, 5, 1, 10, 7, 3, 1
Mame-AL17	441	60 (13.6)	0.028	0.028	0.360	0.025	0.125	0.025	1.061	10, 12, 3, 22, 1, 1
Mame-AL18	423	42 (9.9)	0.011	0.011	−1.231	0.008	−0.040	0.013	0.502	1, 2, 5
Mame-AL19	398	30 (7.5)	0.008	0.007	−0.963	0.009	−0.848	0.007	0.072	1, 16, 9, 1
Mame-AL20	464	66 (14.2)	0.025	0.021	0.599	0.028	0.539	0.022	0.922	3, 1, 4, 1
Mame-AL21	511	48 (9.4)	0.014	0.017	0.114	0.011	−0.959	0.009	−0.309	12, 1, 4, 21
Mame-AL22	500	39 (7.8)	0.012	0.015	−0.395	0.007	−1.112	0.008	0.784	3, 30, 2, 35, 3
Mame-AL23	439	77 (17.5)	0.014	0.015	−1.504*	0.011	−0.973	0.016	−0.99	1, 2, 1, 3, 1, 1
Mame-AL24	530	74 (14.0)	0.027	0.026	0.894	0.023	1.147	0.029	0.438	2, 1
Mame-AL25	460	34 (7.4)	0.010	0.009	−0.977	0.007	−0.644	0.012	0.141	3, 72, 1, 4
Mame-AL26	389	54 (13.9)	0.019	0.020	−0.836	0.016	−0.232	0.018	0.097	2, 21, 1, 1
Mame-AL27	473	90 (19.0)	0.046	0.048	0.133	0.030	1.375	0.040	1.657	1, 1, 1, 1, 2, 2, 16, 3, 6, 2, 10
Mame-AL28	460	36 (7.8)	0.018	0.017	0.760	0.007	−0.952	0.018	1.537	1
Mame-AL29	309	16 (5.2)	0.005	0.002	−1.192	0.007	−0.020	0.005	0.171	3
		Mean(SE)	0.016	0.016		0.014		0.015		
			(0.010)	(0.010)		(0.009)		(0.010)		
Introns										
AB4	203	12 (5.9)	0.006	0.006	0.248	0.004	−0.372	0.002	−0.860	6
α-Globin2	179	14 (7.8)	0.010	0.008	0.114	0.009	−0.426	0.011	0.743	1, 1
GAPDH	343	30 (8.7)	0.008	0.009	−1.205	0.007	−0.479	0.007	0.427	4, 3
GTP	637	54 (8.5)	0.010	0.011	−1.164	0.008	−0.010	0.011	0.260	N/A
RI2	269	23 (8.6)	0.011	0.008	−0.527	0.009	−0.503	0.016	1.216	1, 2, 1, 1
TGFb2	542	47 (8.7)	0.010	0.010	−0.755	0.011	−0.912	0.008	−0.844	9, 43, 1, 12, 9, 5, 1
		Mean(SE)	0.010	0.009		0.008		0.010		
			(0.002)	(0.002)		(0.002)		(0.005)		
mtDNA										
ND2	467	10 (2.1)	0.013	0.007	−0.897	0.009	−1.004	0.008	−1.800*	**N/A**

Figure 2. Genetic diversity (π) across 29 anonymous nuclear markers, six introns, and ND2 region. Dotted vertical bars represent the mean genetic diversity of introns and nuclear anonymous markers.

loci sampled, and the variance usually stabilized after 10–15 loci. This was the case for the variance of all parameters except those for ancestral population size and divergence time, which changed little regardless of the number of loci analyzed (Fig. 7C, D).

Discussion

We used a multilocus approach to study phylogeography of the red-backed fairy wren, a denizen of the widespread savannah forests of northern and eastern Australia. This study not only supports the idea that the Carpentarian Barrier had played a primary role in differentiating the avian fauna of the northern Australia, but also provides an example and a time frame for the Pleistocene "population expansion hypothesis," which suggests that populations distributed across the Carpentarian region were formerly sequestered in refugia around the periphery of northern and eastern Australia. Our study, conducted within a single taxonomic species, found a relatively recent divergence across the Carpentarian barrier (270 kyr), less than half the time estimated for Australian grassfinches (*Poephila*), which are phenotypically clearly distinct species (Jennings and Edwards 2005). Furthermore, the present study demonstrates the applicability of nuclear

anonymous loci markers in phylogeography. Although there have been many studies on speciation patterns in vertebrates of northern Australia (Edwards 1993; Worthington-Wilmer et al. 1999; Cardinal and Christidis 2000; Ford and Blair 2005; Toon et al. 2007), this is the first (aside from Jennings and Edwards 2005) to examine population structure by applying multilocus sequence-based markers, allowing us to estimate demographic parameters with confidence (Brito and Edwards 2008). In addition, the large number of novel SNPs we discovered ($n = 1395$) will be useful for understanding fairy wren behavior and mating systems, particularly using high-throughput SNP genotyping approaches.

CHARACTERISTICS OF ANONYMOUS LOCI

One of the reasons that mitochondrial DNA has thus far been widely used for phylogeographic studies is that it is expected to show higher genetic divergence due to its small effective population size, which is only a quarter that of nuclear DNA when the sex ratio is even. Therefore, phylogeographers expect to get more information per length of sequence from mtDNA than from nuclear DNA. However, considering that genetic diversity is usually measured as the product of mutation rate (μ) multiplied by effective population size (N_e), substantial genetic diversity should
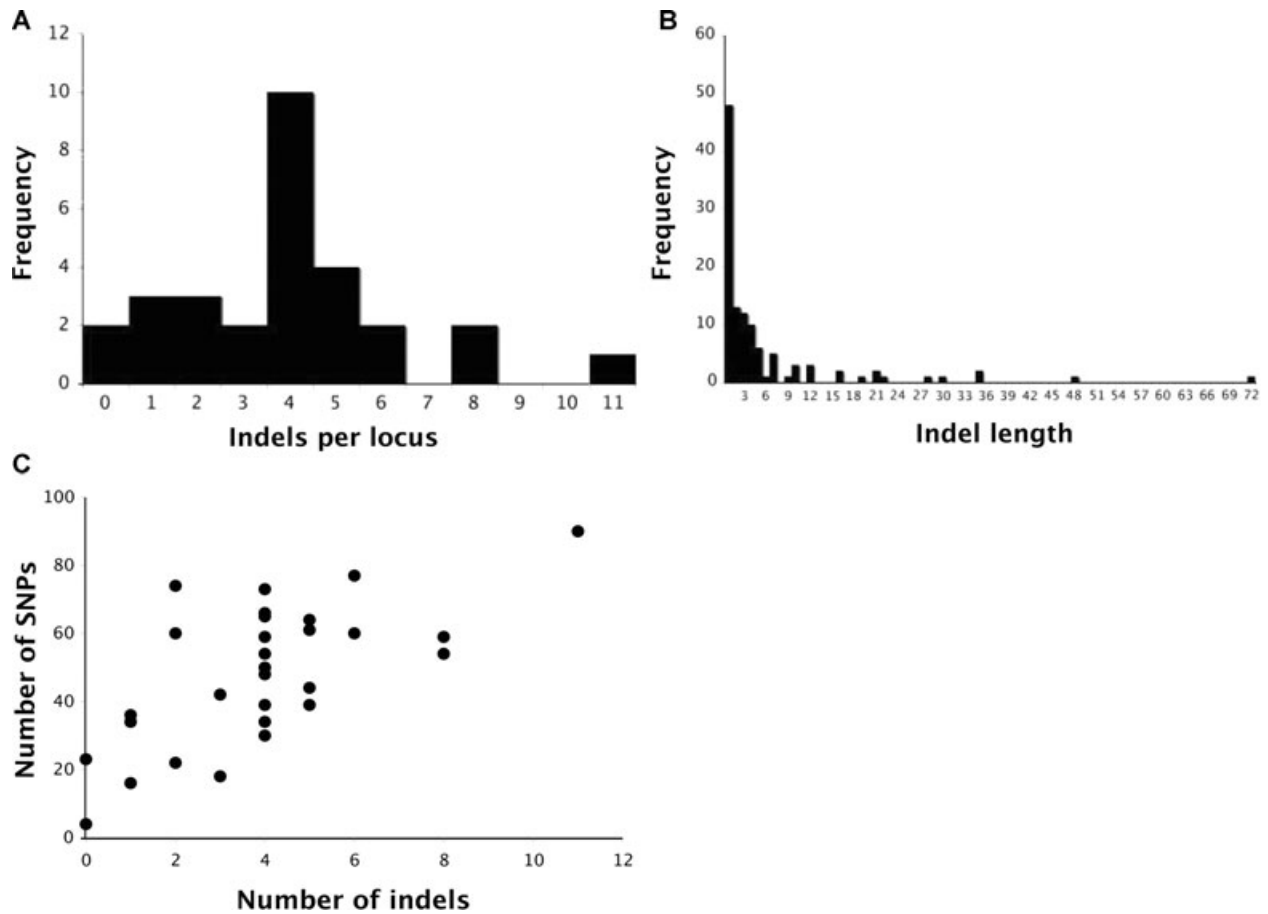
Figure 3. Distribution of indel frequency per locus (A) and indel length (B). Nine sequences that showed any evidence of multiple indels per locus per individual were excluded in our analysis. Panel C represents the relationship between the number of indels per locus and the number of SNPs (Pearson's $r = 0.660$, $N = 29$, $P < 0.01$).

be detectable in nuclear DNA as well if the mutation rate is high enough. This was the case in our study. Even though we cannot directly compare the diversity at anonymous loci to that in the mitochondrial ND2 gene, a coding region, the number of SNPs we detected in anonymous loci (mean $= 48 \pm 21$ per locus, 10.6%) compared favorably to that found in mtDNA (10 SNPs in 467bp, 8%). Our rate of SNP encounter in anonymous loci was less than that found in a study of control region sequences in another passerine bird (96 of 400 sites, or 24%) distributed almost identically to the red-backed wren, the gray-crowned babbler (*Pomatostomus temporalis*; Edwards 1993). However, as judged by mtDNA, the babblers are somewhat more differentiated, both within and between subspecies, than the red-backed wrens appear to be, so this comparison is also not straightforward. Regardless, our results suggest that anonymous loci in birds will exhibit abundant variation for phylogeographic analysis. The level of genetic diversity also appears to be almost six times greater than that found in a trio of grassfinches ($t_{29} = -7.3071$, $P < 0.001$; Jennings and Edwards 2005) and collared flycatcher (Backström et al. 2008), although each study either had limited sampling of individuals (grassfinch)

or included markers representing known genes (flycatcher). This high genetic diversity was also reflected on the pattern of indel frequency in anonymous loci (Johnson 2004). When the number of SNPs (as a proxy of genetic diversity) was plotted across the number of indels, we found a positive correlation (Pearson's $r = 0.66$, $P < 0.01$; Fig. 3C), suggesting that both SNPs and indels may have a common underlying mutational basis, perhaps varying among loci or regionally in the fairy wren genome.

In addition, the genetic diversity detected at our anonymous loci was much higher than that of introns (Fig. 2). This is interesting given that the effective population size is the same across these markers. Because we had ruled out any possible paralogous loci (see Methods), we believe that the high genetic diversity in our study is the result of a high substitution rate in anonymous loci. Introns in vertebrates are known to possess abundant regulatory sequences that presumably are under at least mild stabilizing selection (Collins 1988; Castillo-Davis et al. 2002; Hong et al. 2006). By contrast, anonymous regions of the genome may be genuinely free from stabilizing selection, although of course some noncoding regions of the genome show extensive conservation

Figure 4. Four examples of gene trees constructed with three anonymous loci (Mame-AL*) and the mitochondrial ND2 marker (lower right), each of which exhibits different degrees of population structure (Slatkin's *s*, indicated). All gene trees were rooted with splendid fairy wren *M. splendens* and variegated fairy wren *M. lamberti* sequences. Samples collected from the putative hybrid zone were labeled with "+."
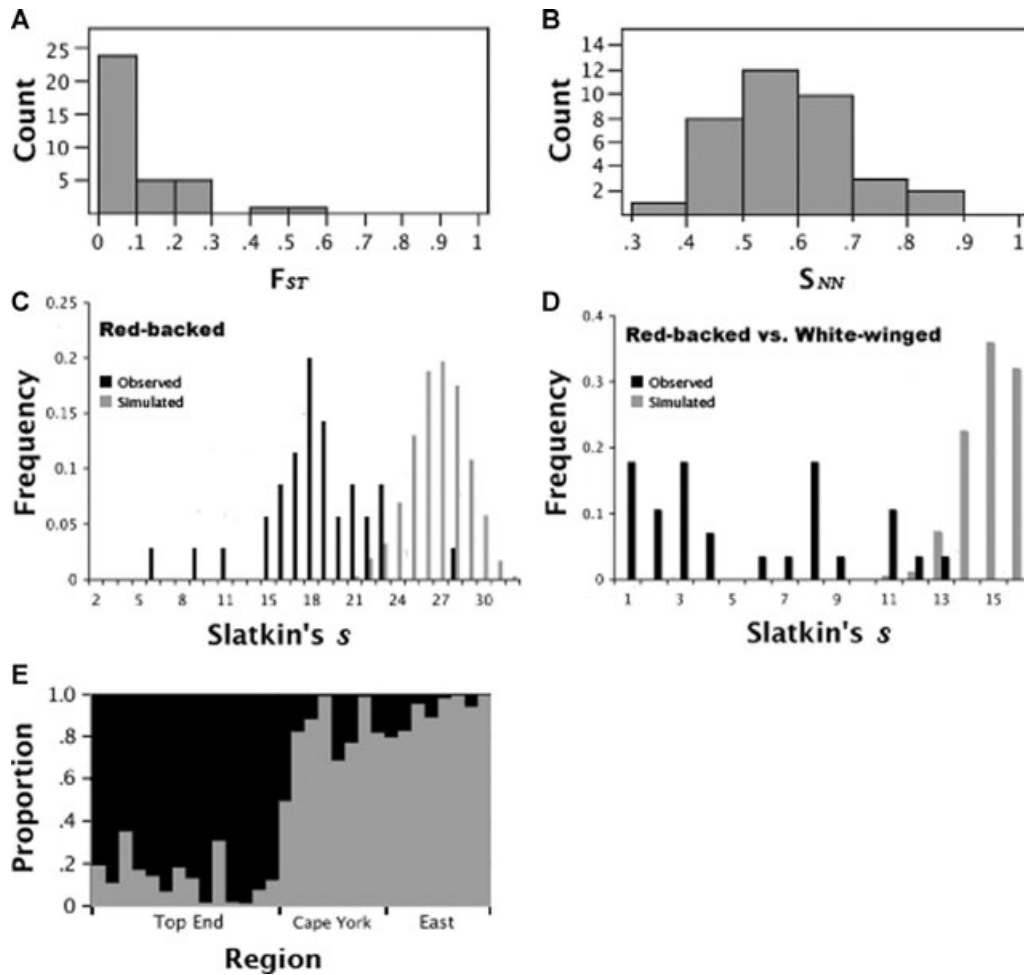
Figure 5. Four different statistics to detect genetic differentiation among populations. Figure A and B depict the distribution of F_{ST} and S_{NN} among loci. See main text for discussion. Figure C and D show the distributions of Slatkin's s within the red-backed fairy wren, and between red-backed wrens and the sister species (white-winged fairy wren), respectively. The distributions of observed Slatkin's s (black bars) were compared to Slatkin's s that we obtained from 1000 simulated gene trees (gray bars). Figure E is the result from a clustering analysis using STRUCTURE, representing probabilistic assignments of individual genotypes to two populations in which we found the likelihood estimate was highest.

over evolutionary time (Katzman et al. 2007). This disparity may arise from the difference in substitution rate between the two classes of markers, suggesting that substitution rates in anonymous loci are higher in this species.

When all genetic markers were analyzed across species, we found that red-backed and white-winged fairy wrens are still sharing their nuclear genes, but not their mtDNA (ND2 gene in our study; Fig. 4). Given that the white-winged fairy wren shows population structure between east and west (Driskell et al. 2002), this incomplete lineage sorting for the nuclear genes may be due to our biased sampling on eastern population of the white-winged fairy wren, which is genetically more closely related to the red-backed fairy wren. However, they are phenotypically very divergent and their distributions do not overlap (Rowley and Russell 1997), so this lack of reciprocal monophyly in nuclear genes is noteworthy.

This pattern will be common in bird studies when most species have arisen rapidly or had ancestors with large effective population sizes (Edwards et al. 2005; Jennings and Edwards 2005). In addition to the fact that mtDNA may in some cases be driven by natural selection (Bazin et al. 2006) and represents only a portion of a species history due to its maternal inheritance, our multiple nuclear gene study emphasizes the necessity of incorporating nuclear genes in understanding evolutionary history of species.

POPULATION STRUCTURE
Given that haplotype diversity was very high, we estimated two commonly used statistics (F_{ST} and S_{NN}) with genetic distance as well as allele frequency. Both statistics suggested that there is significant differentiation among populations, a pattern that is also observed in the analysis using Slatkin's s. The fact that this
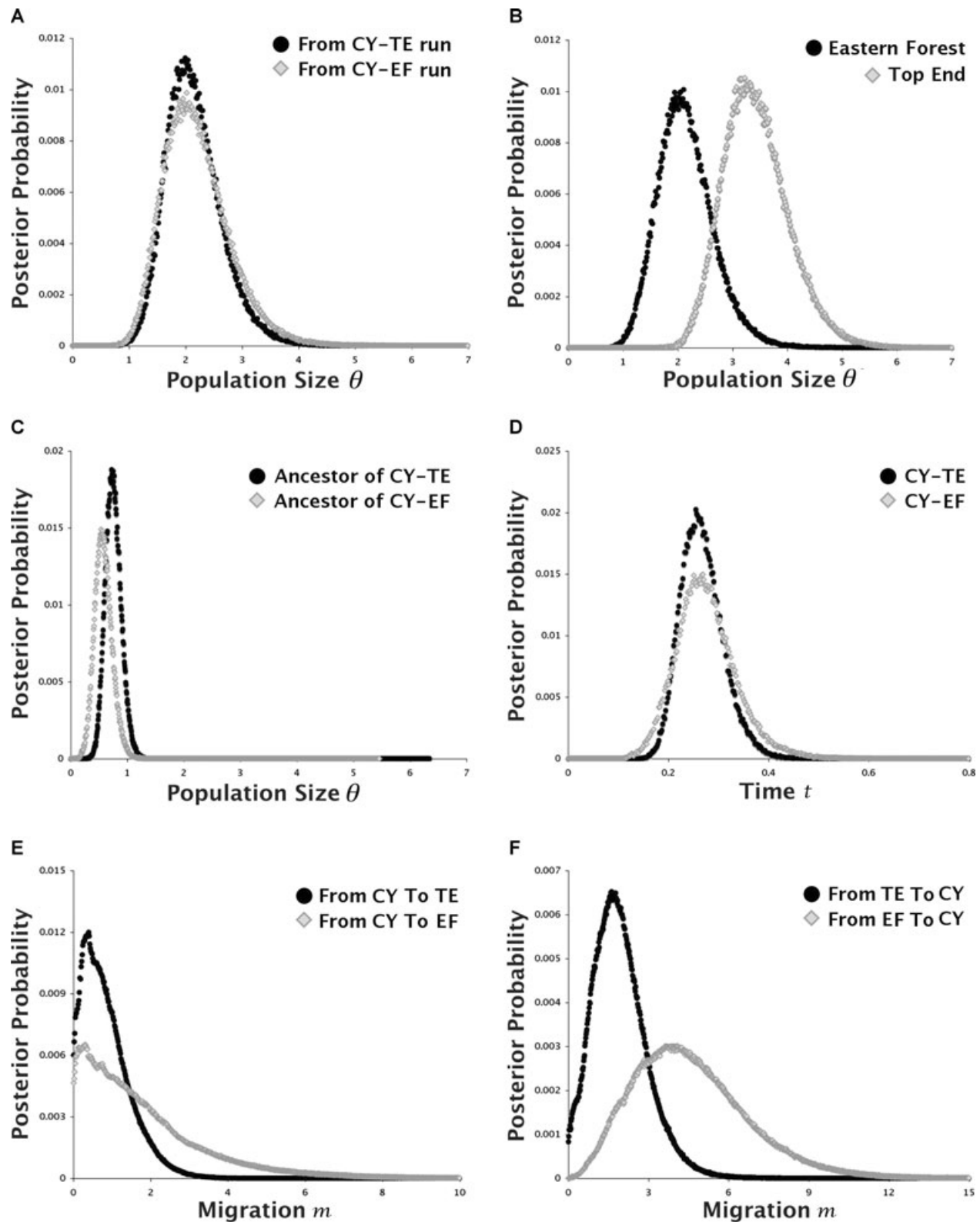
Figure 6. Estimated posterior probability distributions for six parameters using IM. All estimates are scaled by the mutation rate (μ). The distributions in black and gray are the results of CY-TE and CY-EF runs, respectively. Each IM run was conducted with 20 million steps (after two million burn-in period) and four different chains: (A) population size (θ) for the CY population; (B) population size (θ) for the EF and the TE population; (C) population size (θ) for the ancestor of CY/TE and CY/EF; (D) divergence times (t) between the CY and the TE population, and between the CY and the EF population; (E) migration rate (m) from the CY population to the TE and EF populations; (F) migration rate (m) into the CY population from the TE and EF populations.

Table 2. Results of IM analysis of two two-population subsets of the red-backed fairy wren data (TE vs. CY and CY vs. EF). All estimates shown here are scaled by mutation rate (μ): θ_1 and θ_2 represent the population size for the first and the second population (e.g., θ_1 and θ_2 in the first row are the population size of the TE and the CY population, respectively); θ_A is the ancestral population size of two populations included in the run; t indicates the divergence time between two populations; m_1 and m_2 stand for the migration rate at which gene comes into the first population and the second population, respectively. HPD90Hi and HPD90Lo represent the upper and lower bound of the estimated 90% highest posterior density interval, respectively, and HiPts are our parameter estimates.

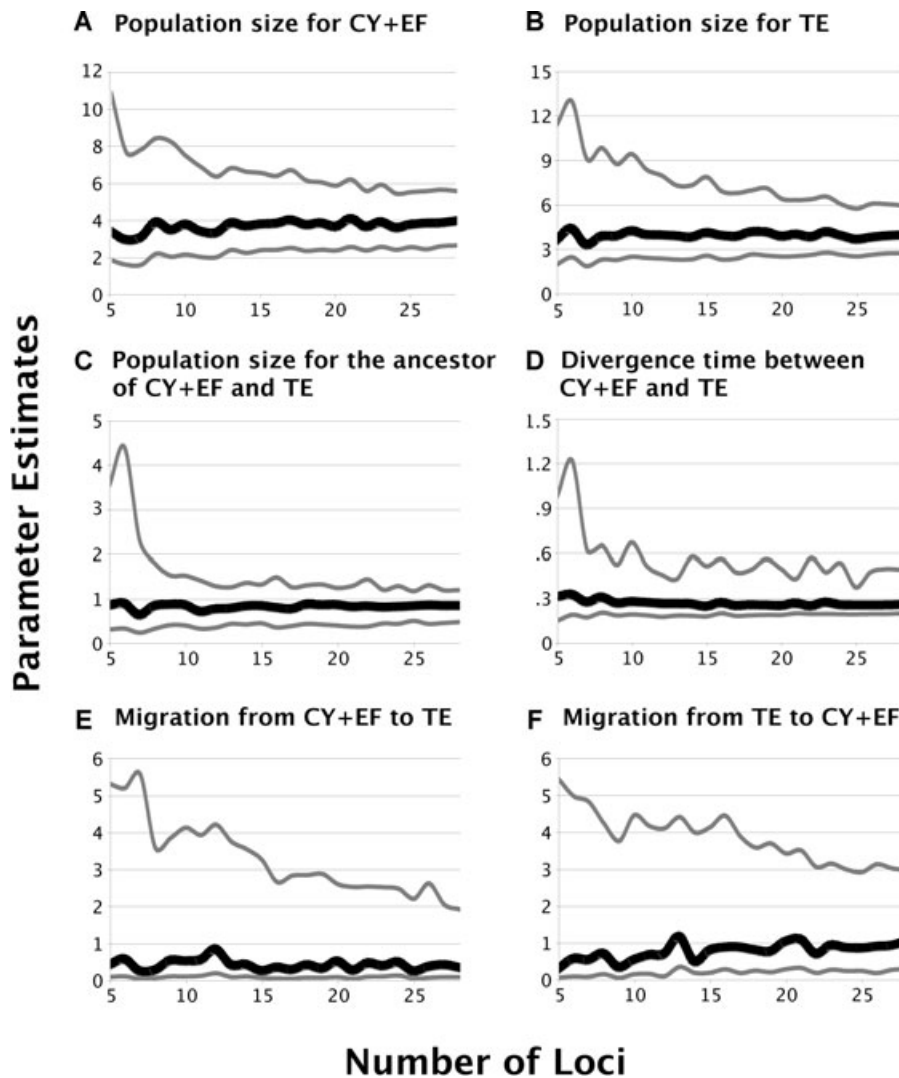| Comparison	θ_1	θ_2	θ_A	t	m_1	m_2
TE vs. CY						
HiPt	**3.223**	**1.995**	**0.727**	**0.255**	**0.398**	**1.628**
HPD90Hi	4.379	2.902	0.981	0.335	1.688	0.218
HPD90Lo	2.452	1.348	0.518	0.195	0.008	3.383
CY vs. EF						
HiPt	**2.021**	**2.108**	**0.539**	**0.269**	**4.088**	**0.308**
HPD90Hi	3.080	2.933	0.828	0.369	7.763	3.893
HPD90Lo	1.256	1.307	0.320	0.175	1.163	0.008

Figure 7. Parameter estimation and variance as a function of the number of independent anonymous loci. In each graph, the distribution of mean of peak posterior estimates (solid line) along with 95% posterior densities interval (between upper and lower gray lines) is depicted. Datapoints are averages of five analyses involving randomly generated subsets of the data (see Materials and Methods). All estimates shown here are scaled by mutation rate (μ).

analysis is dependent on gene trees, which may be not fully resolved, can cause problems if Slatkin's *s* is used to estimate gene flow. However, we have used this statistic primarily as a broad descriptor of the presence of structure, rather as a means to estimate rates of gene flow. These statistics suggest that there is significant population structure, but they do not specify which populations are closer than one another, which is one of the main questions in our study. By grouping the CY population with the EF population, the clustering algorithm implemented in STRUCTURE shows that there is much greater differentiation across the Carpentarian Barrier than across the Burdekin Barrier (barriers D, Fig. 1). Based on morphological data, the subspecies boundary had been set near the Burdekin Barrier (Schodde and Mason 1999). However, our study shows that genetic inference favors the "Carpenatarian Barrier first" hypothesis over the morphological taxonomy. Considering that there is a substantial plumage color difference between the two subspecies (Schodde and Mason 1999), the plumage-based taxonomy might not represent the true demographic history of this species. This discrepancy may result from subjective assessment of the plumage-color, and may represent another case in birds of conflict between molecular markers and traditional taxonomy (Zink 2007). This discrepancy might be resolved with more quantitative and detailed assessment (e.g., spectrophotometry), or resolved by taking other morphological characters into account.

DIVERGENCE TIMES IN THE RED-BACKED FAIRY WREN COMPLEX

The results of the IM program suggest that the divergence time between the CY population and the TE population is approximately 270,000 years ago. Our estimate is more recent than the estimated divergence time (0.61∼0.71 mya) for two grass finches (Jennings and Edwards 2005), also using anonymous markers. This discrepancy could be due to the method we used to fit our sequence data to the IM model. Because the IM program assumes no recombination within loci, we took from each locus the longest block of sequence identified as nonrecombining by the four-gamete test. Even though this method can avoid violating the assumption, we expect such loci to have shorter genealogies, on average, in terms of recovered mutations (Hey and Nielsen 2004; Becquet and Przeworski 2007). Consequently, our divergence time of 0.27 mya may be an underestimate. Another possibility is that on average the anonymous loci we have chosen for this study evolve slower than those in the grass finch study, or that we have misestimated divergence times relative to the grass finch study by using an incorrect generation time calibration. One deficiency of the anonymous locus approach is that frequently no locus will overlap with those used in studies of other taxa, as is the case in our comparison with Jennings and Edwards (2005) study. Thus, one could argue that such studies cannot be compared because they

do not use homologous loci. We have argued elsewhere that such a conclusion is probably too conservative, given that such studies at least focus on a homologous subgenome (Brumfield et al. 2003). In addition, the intron sequences that we used should be amplifiable in other lineages of birds such as the grassfinches, and could be used to estimate relative rates of genome-wide evolution between the two lineages (Brumfield et al. 2003). In general we are comfortable comparing our study directly to the grassfinch study, with the caveats that we have outlined above. Using a standard set of introns across avian phylogeographic studies, such as those provided by Backström et al. (2008) may also be a useful approach for comparing different studies. It is reasonable that our estimate of divergence time across the Carpentarian Barrier is less than that for the grassfinches, primarily because we are studying taxa that are morphologically conspecific and presumably much more recently diverged than the grassfinches, which are conventionally recognized as clearly different species with highly distinct plumage color, song, and other traits (Cracraft 1986; Boles 1988; Schodde and Mason 1999).

We can gain additional perspective on the divergence times in the present study by examining mtDNA divergence. Using the simple pairwise method for estimating net nucleotide diversity (Nei and Li 1979; Edwards and Beerli 2000), we estimated the divergence time for the ND2 gene between the TE and CY + EF regions. Net nucleotide diversity (*d*) was calculated by the equation $d = d_{xy} - 0.5\ (d_x + d_y)$, where d_{xy} is the average divergence between the TE and CY + EF regions (0.018), and d_x and d_y are the average pairwise divergence within the TE and CY + EF regions, 0.006 and 0.011, respectively. The net nucleotide diversity between the TE and the CY + EF regions in ND2 region was 0.9%, corresponding to approximately 450,000 years for the divergence time across the Carpentarian Barrier given a standard mtDNA rate calibration of 2% divergence per million years (Lovette 2004; Weir and Schluter 2008). This divergence time is about twice as old as that from anonymous loci (about 260,000 years when we ran the IM program using only anonymous loci). This discrepancy can be due to error in calibration rate for mtDNA or anonymous loci, or both. Also, given that the mtDNA rate calibration is often used for groups that show reciprocal monophyly, our estimate of divergence time may be overestimated.

Even though our study supports the idea of the Carpentarian Barrier as the primary geographical barrier in the northern part of Australia, the cause of the morphological divergence recognized by taxonomists across the Atherton Plateau is still unclear. The uplift of the plateau is presumed to have occurred prior to the opening of the Coral Sea in the Late Cretaceous (Ollier 1982), but this event is clearly unrelated to divergences in our study. The Atherton plateau may represent a barrier that has experienced mulitple vicariant events in the same place over many different times (Leaché et al. 2007). One explanation can be found

in the long-term vegetation history in this area (Hopkins et al. 1990; Kershaw 1994). Based on charcoal layers found in volcanic craters located on Atherton plateau, these studies claimed that there was a series of extensive fires across this plateau around the late Pleistocene (150,000 $\sim$ 120,000 years ago). Given the evidence that red-backed fairy wrens have a strong tendency to avoid burnt areas (Rowley and Russell 1997), the environmental change caused by these fires could have played a role in structuring populations. An alternative explanation is that the subspecies separation actually occurred at the Burdekin Barrier, not on the Atherton Plateau. Many species exhibit high divergence across the Burdekin Barrier (Joseph et al. 1993; James and Moritz 2001; Dolman and Moritz 2006), suggesting that it is a major barrier for rainforest species in the northeastern part of Australia. Even though the origin of the Burdekin Barrier is still unclear, it could have dampened vagility for this species and caused subspecies divergence (James and Moritz 2001). Given that red-backed fairy wrens mainly inhabit open forest/savannah rather than rainforest (the predominant habitat on the Atherton Tableland), this idea also provides a plausible scenario for the population history of this species. Our estimated rates of gene flow suggest that there has been a significant number of migrants among the three regions, especially into the CY region, and mainly from the EF region. Taking all this information into account, we suggest that the fluctuation of Pleistocene sea levels in the Gulf of Carpentaria and environmental change in the Atherton Plateau or Burdekin Barrier were the primary factors initiating the differentiation of red-backed fairy wrens. Later on, as the plateau was revegetated, birds in the EF region could have spread back north to Cape York, whereas the Carpentarian barrier, whatever its source is, remained a potent barrier to gene flow maintaining differentiation between the CY and the TE populations.

HOW MANY LOCI DO WE NEED?

One way to reduce the variance in coalescent estimates of population parameters is to increase the number of loci (Plunzhnikov and Donnelly 1996). The question then is, how many loci do we need to get proper estimation with sufficiently minimized variance? In our analysis of subsamples of our loci, we failed to obtain stable parameter estimates when we used fewer than five loci, presumably due to lack of information in this small subset for our particular dataset (Fig. 2). Despite the large amount of variation we found, particularly in anonymous loci, presumably the recent divergence times and presence of gene flow in our system require a large number of variable sites that are not quite provided by five loci. However, the variances for most of parameters rapidly decrease from five to 15 loci, and slowly stabilize thereafter. A similar pattern was also observed in the grassfinch study, which used different multilocus statistical methods (e.g., MCMCcoal; Rannala and Yang 2003) to estimate parameters among species

(Jennings and Edwards 2005). Given that some parameters (such as ancestral population size and divergence time) may require more data to estimate than current population sizes do (Yang 1997), it is hard to predict how many loci will be needed for a specific study. The number of loci required likely will depend not only on population history, but also on how frequently recombination occurs within loci. It has been suggested that recombination is a difficult variable to include in genealogy-based, likelihood framework for historical model-fitting (Kuhner et al. 2000; Nielsen 2000). But some recent progress has been made in this area (Becquet and Przeworski 2007). New pyrosequencing approaches will likely make it possible to sample even much larger numbers of loci easily from nonmodel species, and will dramatically increase the potential of statistical phylogeography (Knowles 2004).

LITERATURE CITED
Arbogast, B. S., and G. J. Kenagy. 2001. Comparative phylogeography as an integrative approach to historical biogeography. J. Biogeogr. 28:819–825.

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

———. 2000. *Phylogeography*. Harvard Univ. Press, Cambridge, MA.

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

Avise, J. C., J. Arnold, R. M. Ball, E. Bermingham, T. Lamb, J. E. Neigel, C. A. Reeb, and N. C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annu. Rev. Ecol. Syst. 18:489–522.

Backström, N., S. Fagerberg, and H. Ellegren. 2008. Genomics of wild bird populations: a gene-based set of references makers spread across the avian genome. Mol. Ecol. 17:964–980.

Bazin, E., S. Glémin, and N. Galtier. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science 312:570–572.

Becquet, C., and M. Przeworski. 2007. A new approach to estimate parameters of speciation models with application to apes. Genome Res. 17:1505–1519.

Bensch, S., D. E. Irwin, J. H. Irwin, L. Kvist, and S. Åkesson. 2006. Conflicting patterns of mitochondrial and nuclear DNA diversity in *Phylloscopus* warblers. Mol. Ecol. 15:161–171.

Boles, W. E. 1988. Comments on the subspecies of Australian native and introduced finches. Emu 88:20–24.

Brito, P. H., and S. V. Edwards. 2008. Multilocus phylogeography and phylogenetics using sequence-based markers. Genetica Online First doi: 10.1007/s10709-008-9293-3.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256.

Cardinal, B. R., and L. Christidis. 2000. Mitochondrial DNA and morphology reveal three geographically distinct lineages of the large bentwing bat (*Miniopterus schreibersii*) in Australia. Aust. J. Zool. 48:1–19.

Carstens, B. C., and L. L. Knowles. 2007. Shifting distributions and speciation: species divergence during rapid climate change. Mol. Ecol. 16:619–627.

Carstens, B. C., J. D. Degenhardt, A. L. Stevenson, and J. Sullivan. 2005. Accounting for coalescent stochasticity in testing phylogeographical hypotheses: modelling Pleistocene population structure in the Idaho giant salamander *Dicamptodon aterrimus*. Mol. Ecol. 14:255–265.

Castillo-Davis, C. I., S. L. Mekhedov, D. L. Hartl, E. V. Koonin, and F. A. Kondrashov. 2002. Selection for short introns in highly expressed genes. Nat. Genet. 31:415–418.

Christidis, L., and R. Schodde. 1997. Relationships within the Australo-Papuan fairy-wrens (Aves: Malurinae): an evaluation of the utility of allozyme data. Aust. J. Zool. 45:113–129.

Collins, R. A. 1988. Evidence of natural selection to maintain a functional domain outside of the 'core' in a large subclass of group I introns. Nucleic Acids Res. 16:2705–2715.

Cracraft, J. 1986. Origin of evolution of continental biotas: speciation and historical congruence within the Australian avifauna. Evolution 40:977–996.

———. 2001. Avian evolution, Gondwana biogeography and the Cretaceous-Tertiary mass extinction event. Proc. R. Soc. Lond. B 268:459–469.

Di Rienzo, A., P. Donnelly, C. Toomajian, B. Sisk, A. Hill, M. L. Petzl-Erler, G. K. Haines, and D. H. Barch. 1998. Heterogeneity of microsatellite mutations within and between loci, and implications for human demographic histories. Genetics 148:1269–1284.

Dolman, G., and C. Moritz. 2006. A multilocus perspective on refugial isolation and divergence in rainforest skinks (*Carlia*). Evolution 60:573–582.

Driskell, A. C., S. Pruett-Jones, K. A. Tarvin, and S. Hagevik. 2002. Evolutionary relationships among blue- and black-plumaged populations of the white-winged fairy-wren (*Malurus leucopterus*). Aust. J. Zool. 50:581–595.

Edwards, S. V. 1993. Long-distance gene flow in a cooperative breeder detected in genealogies of mitochondrial DNA sequences. Proc. R. Soc. Lond. B 252:177–85.

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.

Edwards, S. V., S. B. Kingan, J. D. Calkins, C. N. Balakrishnan, W. B. Jennings, W. J. Swanson, and M. Sorenson. 2005. Speciation in birds: genes, geography, and sexual selection. Proc. Natl. Acad. Sci. USA 102:6550–6557.

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

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

Ford, J. 1974. Speciation in Australian birds adapted to arid habitats. Emu 74:161–168.

———. 1977. Subspeciation in the white-throated warbler of Australia and New Guinea. Emu 78:90–92.

———. 1987a. Hybrid zones in Australian birds. Emu 87:158–178.

———. 1987b. Minor isolates and minor geographical barriers in avian speciation in continental Australia. Emu 87:90–102.

Ford, F., and D. Blair. 2005. Neat patterns with a messy history: Savannah refuges in northern Australia. Mammal Study 30:S45–S50.

Hagelberg, E., N. Goldman, P. Lio, S. Whelan, W. Schiefenhövel, J. B. Clegg, and D. K. Bowden. 2000. Evidence for mitochondrial DNA recombination in a human population of island Melanesia. Proc. R. Soc. Lond. B 267:1595–1596.

Hare, M. P. 2001. Prospects for nuclear phylogeography. TRENDS Ecol. Evol. 16:700–706.

Hewitt, G. M. 2001. Speciation, hybrid zones and phylogeography—or seeing genes in space and time. Mol. Ecol. 10:537–549.

Hey, J., and C. A. Machado. 2003. The study of structured populations—new hope for a difficult and divided science. Nat. Rev. 4:535–543.

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.

Hong, X., D. G. Scofield, and M. Lynch. 2006. Intron size, abundance, and distribution within untranslated regions of genes. Mol. Biol. Evol. 23:2392–2404.

Hopkins, M. S., A. W. Graham, R. Hewett, J. Ash, and J. Head. 1990. Evidence of late Pleistocene fires and eucalypt forest from a north Queensland humid tropical rainforest site. Aust. J. Ecol. 15:345–347.

Hudson, R. R. 2000. A new statistic for detecting genetic differentiation. Genetics 155:2011–2014.

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

Hughes, J. M., A. M. Baker, G. De Zylva, and P. B. Mather. 2001. A phylogeographic analysis of southern and eastern populations of the Australian magpie: evidence for selection in maintenance of the distribution of two plumage morphs. Biol. J. Linn. Soc. 74:25–34.

Hughes, J. M., C. L. Lange, P. B. Mather, and A. Robinson. 2002. A comparison of fitness components among different plumage morphs of the Australian magpie *Gymnorhina tibicen*. Emu 102:331–338.

James, C. H., and C. Moritz. 2000. Intraspecific phylogeography in the sedge frog Litoria fallax (Hylidae) indicates pre-Pleistocene vicariance of an open forest species from eastern Australia. Mol. Ecol. 9:349–358.

Jennings, W. B., and S. V. Edwards. 2005. Speciational history of Australian grass finches (*Poephila*) inferred from thirty gene trees. Evolution 59:2033–2047.

Johnson, K. P. 2004. Deletion bias in avian introns over evolutionary timescales. Mol. Biol. Evol. 21:599–602.

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.

Joseph, L., and C. Moritz. 1994. Mitochondrial DNA phylogeography of birds in eastern Australian rainforest: first fragment. Aust. J. Zool. 42:385–403.

Joseph, L., and T. Wilke. 2007. Lack of phylogeographic structure in three widespread Australian birds reinforces emerging challenges in Australian historical biogeography. J. Biogeogr. 34:612–624.

Karl, S. A., and J. C. Avise. 1993. PCR-based assays of Mendelian polymorphisms from anonymous single-copy nuclear DNA: techniques and applications for population genetics. Mol. Biol. Evol. 10:342–361.

Karubian, J. 2002. Costs and benefits of variable breeding plumage in the red-backed fairy-wren. Evolution 56:1673–1682.

Katzman, S., A. D. Kern, G. Bejerano, G. Fewell, L. Fulton, R. K. Wilson, S. R. Salama, and D. Haussler. 2007. Human genome ultraconserved elements are ultraselected. Science 317:915.

Keast, A. 1961. Bird speciation on the Australian continent. Bull. Mus. Compar. Zool. 123:305–495.

Kershaw, A. P. 1994. Pleistocene vegetation of the humid tropics of north-eastern Queensland, Australia. Palaeogeogr. Palaeoclimatol. Palaeoecol. 109:399–412.

Klicka, J., and R. M. Zink. 1999. Pleistocene effects on North America song-bird evolution. Proc. R. Soc. Lond. B 266:695.

Knowles, L. L. 2004. The burgeoning field of statistical phylogeography. J. Evol. Biol. 17:1–10.

Knowles, L. L., and W. P. Maddison. 2002. Statistical phylogeography. Mol. Ecol. 11:2623–2635.

Kuhner, M. K., J. Yamato, and J. Felsenstein. 2000. Maximum likelihood estimation of recombination rates from population data. Genetics 156:1393–1401.

Kuo, C., and J. C. Avise. 2005. Phylogeographic breaks in low-dispersal species: the emergence of concordance across gene trees. Genetica 124:179–186.

Leaché, A. D., S. C. Crews, and M. J. Hickerson. 2007. Two waves of diversification in mammals and reptiles of Baja California revealed by hierarchical Bayesian analysis. Biol. Letts. 3:646–650.

Lovette, I. J. 2004. Mitochondrial dating and mixed-support for the "2% rule" in birds. Auk 121:1–6.

MacDonald, J. D. 1969. Variation in the black-faced woodswallow. Emu 69:168–171.

Maddison, D. R., and W. P. Maddison. 2000. MacClade version 4: analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, MA.

McCracken, K. G., and M. Sorenson. 2005. Is homoplasy or lineage sorting the source of incongruent mtDNA and nuclear gene trees in the stiff-tailed ducks (*Nomonyx-Oxyura*)? Syst. Biol. 54:35–55.

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

Nickerson, D. A., V. O. Tobe, and S. L. Taylor. 1997. PolyPhred: automating the detection and genotyping of single nucleotide substitutions using fluorescence-based resequencing. Nucleic Acids Res. 25:2745–2751.

Nielsen, R. 2000. Estimation of population parameters and recombination rates from single nucleotide polymorphism. Genetics 154:931–942.

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

Nix, H. A., and J. D. Kalma. 1972. Climate as a dominant control in the biogeography of northern Australia and New Guinea. Pp. 61–91 *in* D. Walker, ed. Bridge and barrier: the natural and cultural history of Torres Strait. Australian National Univ., Canberra.

Nylander, J. A. A. 2004. MrModeltest 2.0. Program distributed by the author. Evolutionary Biology Centre, Uppsala Univ.

Ollier, C. D. 1982. The great escarpment of eastern Australia: tectonic and geomorphic significance. J. Geol. Soc. Aust. 29:13–23.

Pakendorf, B., and M. Stoneking. 2005. Mitochondrial DNA and human evolution. Annu. Rev. Genom. Hum. Genet. 6:165–183.

Plunzhnikov, A., and P. Donnelly. 1996. Optimal sequencing strategies for surveying molecular genetic diversity. Genetics 144:1247–1262.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Rannala, B., and Z. Yang. 2003. Bayes estimation of species divergence times and ancestral population sizes using DNA sequences from multiple loci. Genetics 164:1645–1656.

Rowley, I., and E. Russell. 1997. *Fairy-wrens and Grasswrens*. Oxford Univ. Press Inc., New York.

Rozas, J., and J. C. D. B. Sanchez. 2003. DNAsp, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Schodde, R. 1982. Origin, adaptation and evolution of birds in arid Australia. Pp. 191–224 *in* W. R. Barker and P. J. M. Greenslade, eds. Evolution of the flora and fauna of arid Australia. Peacock Publications, Adelaide.

Schodde, R., and I. J. Mason. 1999. The directory of Australian birds. CSIRO Publishing, Canberra.

Slatkin, M., and W. P. Maddison. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123:603–613.

Smart, J. 1977. Late Quaternary sea-level changes, Gulf of Carpentaria, Australia. Geology 5:755–759.

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

Sorenson, M., C. N. Balakrishnan, and R. B. Payne. 2004. Clade-limited colonization in brood parasitic finches (*Vidua* spp.). Syst. Biol. 53:140–153.

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.

Swofford, D. L. 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods) version 4. Sinauer Associates, Sunderland, MA.

Toon, A., J. Hughes, A. Baker, and P. Mather. 2003. Discordance between morphology and genetic structure among three plumage morphs of the Australian magpie. Emu 103:337–343.

Toon, A., P. B. Mather, A. M. Baker, K. L. Durrant, and J. M. Hughes. 2007. Pleistocene refugia in an arid landscape: analysis of a widely distributed Australian passerine. Mol. Ecol. 16:2525–2541.

Tsaousis, A. D., D. P. Martin, E. D. Ladoukakis, D. Posada, and E. Zouros. 2005. Widespread recombination in published animal mtDNA sequences. Mol. Biol. Evol. 22:925–933.

Wakeley, J., R. Nielsen, S. N. Liu-Cordero, and K. Ardlie. 2001. The discovery of single-nucleotide polymorphisms—and inferences about human demographic history. Am. J. Hum. Genet. 69:1332–47.

Waltari, E., and S. V. Edwards. 2002. Evolutionary dynamics of intron size, genome size, and physiological correlates in Archosaurs. Am. Nat. 160:539–552.

Webster, M. S., C. W. Varian, and J. Karubian. 2008. Plumage color and reproductive success in the red-backed fairy-wren: why be a dull breeder? Behav. Ecol. 19:517–524.

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

Wooller, R. D., D. A. Saunders, J. S. Bradley, and C. P. De Rebeira. 1985. Geographical variation in size of an Australian honeyeater (Aves: Meliphagidae): an example of Bergmann's rule. Biol. J. Linn. Soc. 25:355–363.

Worthington-Wilmer, J., L. Hall, E. Barratt, and C. Moritz. 1999. Genetic structure and male-mediated gene flow in the ghost bat (*Macroderma gigas*). Evolution 53:1582–1591.

Yang, Z. 1997. On the esitmation of ancestral population sizes of modern humans. Genet. Res. 69:111–116.

Zhao, Z., L. Jin, Y. Fu, M. Ramsay, T. Jenkins, E. Leskinen, P. Pamilo, M. Trexler, L. Patthy, L. B. Jorde, et al. 2000. Worldwide DNA sequence variation in a 10-kilobase noncoding region on human chromosome 22. Proc. Natl. Acad. Sci. USA 97:11354–11358.

Zink, R. M. 2007. Ecological exchangeability versus neutral molecular markers: the case of the great tit. Anim. Conserv. 10:369–373.

Zink, R. M., and G. F. Barrowclough. 2008. Mitochondrial DNA under siege in avian phylogeography. Mol. Ecol. 17:2107–2121.

Associate Editor: M. Webster

Supporting Information

The following supporting information is available for this article:

Table S1. The detailed information on each sample. Specimen numbers are the numbers that each depository assigned to a specimen, and state represents in which state the sample was collected.
Table S2. Primer sequences (29 anonymous loci, 6 introns and a partial ND2 region) used in this study.

Supporting Information may be found in the online version of this article.
(This link will take you to the article abstract).

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