Accounting for molecular stochasticity in systematic revisions: species limits and phylogeny of *Paroaria*

Liliana M. Dávalos[1] and Ana Luz Porzecanski[2,3,4]

1. Department of Ecology and Evolution, SUNY Stony Brook, 650 Life Sciences Building, Stony Brook, NY 11794-5245, USA; email: ldavalos@life.bio.sunysb.edu

2. Department of Ornithology, American Museum of Natural History, New York, NY 10024, USA.

3. Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Avenue, New York, NY, 10027, USA.

4. Current address: Center for Biodiversity and Conservation, American Museum of Natural History, Central Park West at 79th Street, New York, NY 10024, USA; email: alporze@amnh.org

Corresponding author: Liliana Dávalos, Department of Ecology and Evolution, SUNY Stony Brook, 650 Life Sciences Building, Stony Brook, NY 11794-5245, USA; email: ldavalos@life.bio.sunysb.edu; tel: + 1 631 632 1554; fax: +1 631 632 7626.

Both authors contributed equally to this paper.

Abstract

Different frameworks have been proposed for using molecular data in systematic revisions, but there is ongoing debate on their applicability, merits and shortcomings. In this paper we examine the fit between morphological and molecular data in the systematic revision of *Paroaria,* a group of conspicuous songbirds endemic to South America. We delimited species based on examination of > 600 specimens, and developed distance-gap, and distance- and character-based coalescent simulations to test species limits with molecular data. The morphological and molecular data collected were then analyzed using parsimony, maximum likelihood, and Bayesian phylogenetics. The simulations were better at evaluating the new species limits than using genetic distances. Species diversity within *Paroaria* had been underestimated by 60%, and the revised genus comprises eight species. Phylogenetic analyses consistently recovered a congruent topology for the most recently derived species in the genus, but the most basal divergences were not resolved with these data. The systematic and phylogenetic hypotheses developed here are relevant to both setting conservation priorities and understanding the biogeography of South America.

Keywords: DNA barcoding, codon models, parametric bootstrap, two-prior on branch lengths model, tree comparisons.

INTRODUCTION

Delimiting species is a first, crucial step in natural history, systematic and ecological studies. Research in fields as varied as phylogenetics, historical and ecological biogeography, population and community ecology, and conservation depends on understanding the traits, range, life history and requirements of accurately identified populations. The use of molecular data to delimit

species has expanded with the falling cost of DNA sequencing and rapid growth of databases, and pressure has built up to shift molecular data from an ancillary to a central role in systematics through DNA barcoding (Tautz et al., 2003). DNA barcoding has been widely criticized as too reductive to effectively aid species delimitation where it is most needed; in populations currently undergoing speciation and/or for poorly known taxa for which there are few or no experts (DeSalle et al., 2005; Ebach and Holdrege, 2005; Elias et al., 2007; Moritz and Cicero, 2004). Most systematic revisions, however, seek to reconcile data from various sources. Formal approaches to test the validity of species delimitation when different sources of data conflict (Knowles and Carstens, 2007), or concur (Rosenberg, 2007) have been recently developed, but their application remains relatively uncommon.

In this paper we examine the fit between morphological and molecular data in the systematic revision of *Paroaria,* a group of conspicuous songbirds widely distributed in South America (Ridgely and Tudor, 1989). The genus has been traditionally thought to comprise five species, easily distinguishable based on features of the beak and head plumage, and fairly common in their natural habitats. Two of the species, *P. coronata* and *P. capitata*, have been introduced into Hawaii but are captured and traded for their bright colors and song in their native range, meriting placement in CITES Appendix II (UNEP-WCMC, 2009). The phylogenetic position of the genus has been controversial, and *Paroaria* has been assigned to various tribes within the family Fringillidae (Hellmayr, 1938; Paynter and Storer, 1970; Tordoff, 1954). The genus was not part of a comprehensive phylogenetic study until the 2000's, when two mitochondrial studies showed that *Paroaria* is part of a large tanager radiation in the tribe Thraupini (Burns and Naoki, 2004; Yuri and Mindell, 2002).

Species limits within the genus have been based on external morphological variation interpreted under a polytypic biological species concept (Mayr, 1942). In this study, we collected morphological data based on the examination of > 600 specimens, and obtained sequences from multiple individuals in all the species, and all but one of the currently recognized subspecies. We used these data to identify both morphological and molecular unique diagnostic characters, e.g., Cracraft (1992), Ribas et al. (2005), Reddy (2008). We then developed distance and character-based tests of species differentiation using parametric bootstraps and coalescent simulations. Finally, both morphological and molecular data were analyzed using parsimony, maximum likelihood and Bayesian phylogenetic algorithms.

The approaches implemented here provide a framework for evaluating the fit between morphological and molecular species limits, even with small sampling sizes. Based on our results, we revise the systematics and taxonomy of *Paroaria*, and point to avenues for future research relevant to South American biogeography and conservation.

MATERIALS AND METHODS

Morphological Data Collection

To establish species limits using morphological data, and obtain characters for phylogenetic analyses, all available skin specimens of *Paroaria* were examined from four collections (see Table 1): American Museum of Natural History (AMNH), Field Museum of Natural History (FMNH), Academy of Natural Sciences (ANSP), Museo Argentino de Ciencias Naturales Bernardino Rivadavia (MACN). In addition, data were also obtained from specimens located in the following collections: Natural History Museum (NHM, formerly the British Museum), Muséum National d'Histoire Naturelle in Paris (MNHN), Museum of Comparative Zoology at Harvard University (MCZ), and Museu Nacional de Rio de Janeiro (MNRJ). When it was not

possible to examine type material or other specimens, data were obtained from the literature or museum staff.

The diagnosability of the described species and subspecies was assessed using characters previously described by Sharpe (1888), Sclater (1890), Cory & Hellmayr (1925), Hellmayr (1938), Bond & Meyer de Schauensee (1939), Sick (1950, 1993), Short (1975), Ridgely & Tudor (1989). To find new diagnostic characters, specimens were also examined following a list of standard anatomical features of external morphology (Proctor and Lynch, 1993). Specimens were arranged geographically and examined for variation in all characters. Each character was scored as either fixed or not fixed in a population or groups of populations, while accounting for individual and seasonal variation, variation due to specimen preparation, and obvious abnormalities. Continuously varying variables, such as culmen length, were assessed from the literature to determine whether non-overlapping, discrete differences that might be interpreted as diagnosable characters could be found. In the cases where such differences were reported, specimens were measured to test the published values. Names for colors and plumage patterns follow those used by previous researchers, always using the simplest color description. When a standard was needed for comparisons, color names and numbers follow Smithe (1981). Morphological characters were coded and analyzed as unordered. If homology for a certain character could not be established for the outgroups, it was coded as missing or inapplicable.

Molecular Data Collection

Molecular characters can be valuable for evaluating species limits in birds (Ribas et al., 2005; Ribas and Miyaki, 2004; Zink and Blackwell-Rago, 2000), and should be particularly helpful when investigating groups of confusing morphology. By examining both morphological data, which is encoded across multiple loci in the nuclear genome, and the cytochrome *b* gene, which

is encoded in the mitochondrial genome and evolves independently, the chances of finding evidence of divergent evolutionary histories can be increased. In addition, because of their mode of inheritance and smaller effective population size, mitochondrial DNA sequences are expected to accumulate fixed characters before nuclear sequences do (Moore, 1995).

Tissue samples were obtained from field collections in Bolivia and Uruguay, museum tissue collections, and from toe pads of museum study skins. Complete cytochrome *b* sequences (1143 bps) were obtained for all described *Paroaria* taxa with the exception of the forms *P. capitata fuscipes* and *P. baeri xinguensis* for which DNA could not be successfully extracted from study skins. Within each described taxon, sequences were obtained from as many individuals as possible, given tissue sample availability and the age of the study skin (i.e., feasibility for DNA extraction). The final datasets contained a total of 34 *Paroaria* individuals, representing seven of the nine described taxa. Voucher information can be found in Table 2, and the collection locality of each DNA sample is mapped in Figure 1.

Extraction and amplification protocols for fresh tissue samples followed Lee et al. (1997) and Cracraft et al. (1998), with specific primer design for several taxa, as necessary (Table 3). In the case of museum skins, samples were collected following Mundy et al. (1997) and DNA was extracted using a Qiagen DNeasy tissue Extraction Kit (Qiagen, Inc.). All PCR reactions involving DNA from skins were performed using aerosol resistant tips and taxon-specific primers for pieces of ~250 base pairs (bp) in length and with a minimum of 50 bp overlap. Sequences were obtained in an Applied Biosystems 377 Automated Sequencer following protocols described in Lee et al. (1997), or cleaned automatically by BioRobot 9600 (Qiagen), then twice by alcohol precipitation (70% ethanol with and without $MgCl_2$ to prevent salt residue) and read using an Applied Biosystems 3100 Capillary Automated Sequencer. All portions of the

sequence for each taxon were read in both directions and up to four times in overlap zones. Sequences were edited and assembled with Sequencher 4.1 software (Gene Codes, Ann Arbor, Michigan). Complete match at >50 bp overlap areas and reading frame maps were used to detect and prevent cases of cross-contamination or the amplification of nuclear products. New sequences collected as part of this study were deposited in GenBank under accession numbers **FJ715650**-**FJ715683**. Complete cytochrome *b* sequences of *Cissopis leveriana* (**AY383096**), *Schistochlamys melanopsis* (**AY383102**) and *Neothraupis fasciata* (**AY383100**) were used to root the *Paroaria* tree (Burns and Naoki, 2004; Yuri and Mindell, 2002).

Species delimitation

Establishing the limits of *Paroaria* species was a central goal of this study. Two sets of data were used to assess species limits among the populations sampled. First, morphological variation was examined to detect characters or character combinations uniquely shared among individuals from a geographic region. If the character(s) in question were fixed in all the individuals sampled, that population was considered diagnosably distinct and treated as a hypothesized phylogenetic species (Cracraft, 1983). The series of museum study skins examined were large relative to molecular sampling sizes (see Table 1).

Second, sequences from at least two individuals from each putative species were examined to test the units hypothesized based on morphology and to identify unique, potentially fixed mutations. Any assessment of character fixation is contingent on population sampling, and for molecular data, on the stochastic nature of evolution. We designed two approaches to account for molecular stochasticity and sampling limitations. The general goal of these analyses was to assess the consistency of morphology-based species delimitation with molecular characteristics in a quantitative framework. Both approaches used the Perl script MrAIC.pl v. 1.4.2 (Nylander,

2004) and the Phyml v. 2.4.4 algorithm (Guindon and Gascuel, 2003) to select the model of sequence evolution that fit the data best. The first approach used a maximum likelihood (ML) parametric bootstrap to generate confidence intervals around observed intra- and inter-specific distances. The second approach examined if the molecular differentiation observed was consistent with gene flow under selected geological scenarios, and therefore estimated the probability that certain groups would not be genetically isolated. We describe them in more detail below.

The parametric bootstrap analysis allowed us to evaluate the species by comparing observed divergences to confidence intervals around the divergences between and within traditionally recognized taxa. This analysis assumed that divergences within and between closely related species could serve as guidelines for delimiting species, and is equivalent to the gap between intraspecific and interspecific distances proposed in DNA barcoding studies (Wiemers and Fiedler, 2007). The parametric bootstrap generates data sets such that the means of the parameters (branch lengths, substitution rates etc.) over all replicates equal those observed in the phylogeny (see below). An individual replicate is therefore compatible with the observed data, but different in its combination of parameters. The quantitative component of this approach captures the stochastic nature of mutation by relying on the parameters of the model of sequence evolution that best fit the observed data.

To select the best model for the parametric bootstrap, we used the second-order Akaike information criterion (AIC*c*) in MrAIC because the ratio of nucleotides to number of parameters in the model of molecular evolution was < 40 (Posada and Buckley, 2004). The general time reversible (Tavaré, 1986) model of sequence evolution with a four-category discrete gamma distribution of rate variation across sites (Yang, 1996) (GTR+G), and the best tree resulting from

Bayesian analyses (see below for alignment and tree searchers) were used to optimize parameters on the sequence data using PAML 4 (Yang, 2007b). The phylogeny and optimal parameters were used to simulate 1,000 nucleotide data sets using PAML's *evolver* module. The same phylogeny and data were also used to fit parameters under a codon model of sequence evolution, and simulate 1,000 data sets under this alternative model using PAML. Although not evaluated in model selection software, such as ModelTest (Posada and Crandall, 1998) or MrAIC, codon models provide a better fit to protein-coding data with fewer parameters (Dávalos and Perkins, 2008; Ren et al., 2005). Codon models account for sequence evolution by fitting only two parameters to the data, in addition to base frequency parameters; a transition/transversion rate ratio and the rate of non-synonymous to synonymous substitution (Muse and Gaut, 1994).

Each data set simulated under the nucleotide (1,000 alignments) and codon models (1,000 alignments) was then analyzed to obtain the highest ML distance within species, and the lowest ML distance between species using PAUP* version 4.0b10 (Swofford, 2002). Optimal parameters for the ML nucleotide evolution model were applied to obtain distances corrected for multiple substitutions along braches. The significance of distances within and between newly proposed taxa was obtained by comparing observed values with the frequency distributions of simulated distances for traditionally recognized taxa.

The second set of molecular analyses used coalescent simulations to test whether observed sequence divergences were consistent with models of events that would have allowed restricted gene flow between subpopulations, preventing the fixation of diagnostic differences. The simulations allowed us to test the probability that the observed molecular differences were not due to separate evolutionary histories, but to stochastic processes such as sampling and mutation instead.. We applied this approach to two sets of observations: 1) genetic distances

between *P. gularis gularis* (hereafter referred to as *gularis*, see Results) samples from western Amazonia and from the remainder of the range, and 2) mutations fixed between *Paroaria capitata* and *P. gularis cervicalis*. These last two populations were found to be sister species (see Results). These analyses required modeling the evolution of two widespread ancestral populations, each later splitting into two semi-isolated subpopulations. The two proposed splits corresponded to events that would have given rise to: 1) relatively high distances between *gularis* from Ecuador and Peru vs. Brazil, and 2) diagnostic differences between *capitata/cervicalis*.

To perform coalescent simulations, we first used MrAIC applying the Akaike information criterion (AIC) to identify the best model of nucleotide evolution, and PAML to obtain optimal parameters under the selected model. The resulting model and parameters were then applied to estimate the value of θ for each of the two populations (*gularis*, and *capitata/cervicalis*) using the coalescence-based ML sampler *migrate* v. 2.3 (Beerli, 2004; Beerli and Felsenstein, 1999), with an estimated sequencing error of 0.1%. These analyses sampled genealogies using the ML method from Markov chain Monte Carlo simulations. To ensure consistency of results, we ran each analysis 4 times, starting from different random seeds. After a burn-in period of 10,000 steps, each analysis ran 10 short chain searches sampling 10,000 steps and recording values every 20 steps, followed by 3 long chain searches sampling 100,000 steps with similar recording frequency. The length of chains was enough to converge on likelihood and θ estimates across multiple independent analyses (Beerli and Felsenstein, 2001). The parameter θ comprises the product of the effective population size (N_e) and the mutation rate, multiplied by a factor that depends on the ploidy and mode of inheritance of the locus (Fu and Li, 1993). The estimated θ

parameters and their two extremes (95% confidence interval) were then used to generate two sets of 1,000 genealogies using the coalescent sampler *ms* (Hudson, 2002).

Modeling the expected pattern of fixed mutations or distances under restricted gene flow required specifying four parameters besides θ: 1) mutation rate, 2) generation time, 3) age of the split, and 4) gene flow following the split. We obtained an estimate of mutation rate from lineage- and locus-specific analyses of substitution rates across the extant bird phylogeny (Brown et al., 2008). The rate of substitution for mitochondrial third positions in *Hemispingus frontalis* (mean = 0.0286 substitutions/site/million year, s.d. = 0.0081) was used to approximate the mutation rate of cytochrome *b*. This lineage was selected because it is a member of the same family as *Paroaria*: Thraupidae. Generation time was estimated at 2 years, based on the age of greatest breeding success in captive populations. The dates of proposed splits were selected based on geological events relevant to the populations in question. The last glacial maximum, around 21,000 years ago, is thought to have isolated humid Amazonian forests abutting the Andes from central Amazonia in Brazil (van der Hammen and Hooghiemstra, 2000). This would have affected gene flow between subpopulations of *P. gularis*. According to paleoecological data, the last time habitats of *cervicalis* in the Beni region and the northernmost populations of *capitata* in the Pantanal were continuous before the present was the last interglacial, about 130,000 years ago (Assine and Soares, 2004). This period was followed by a glaciation that effectively isolated populations in temperate latitudes (Boulet and Gibbs, 2006; Lessa et al., 2003). Because the forest habitat preferred by *gularis* would have been split into two roughly equal halves, this simulation assumed equal-size subpopulations with low symmetrical gene flow. In contrast, the geographic range of *cervicalis* is much smaller than that of *capitata*, so this

simulation was constrained to maintain limited asymmetrical gene flow, such that most genetic exchange went from the larger to the smaller subpopulation, until the present (Fig. 2).

The coalescent sampler *ms* implements an infinite-sites model that does not account for multiple hits at a given site, so the simulated genealogies were used as input for the molecular sequence simulator Seq-Gen v. 1.3.2 (Rambaut and Grassly, 1997). The best-fit nucleotide models and parameter values, with a sequence length of 1143 bases, were used to simulate sequences on the genealogies. PAUP* was then used to calculate the highest ML distance between proposed subpopulations of *P. gularis*, and to count the number of mutations fixed in *cervicalis* with respect to *capitata*. These simulations established the statistical significance of: 1) the distance observed between *gularis* individuals, and 2) fixed mutations observed between the proposed species *cervicalis* and *capitata*.

A decision to elevate a previously described subspecies or new population to species status, or to synonymize it was made case-by-case based on all available data and analyses.

Phylogenetic Analyses

There were no deletions or insertions in this protein-coding gene, and alignment was performed by eye. Maximum parsimony (MP) analyses of all sequences and concatenated morphological characters were performed using the heuristic search algorithm in PAUP*, with 1,000 random addition sequences and tree-bisection-and-reconnection (TBR) branch swapping. Analyses of phylogenetic species exemplars were performed using the branch and bound search algorithm. One thousand bootstrap pseudoreplicates (each with 10 random addition sequences and TBR branch-swapping) were used to estimate support in MP analyses, and all characters in MP analyses had equal weights and were unordered.

Based on MrAIC under the AIC*c*, Bayesian analyses applied a general time reversible (Tavaré, 1986) model of sequence evolution with a four-category discrete gamma distribution of rate variation across sites (Yang, 1996), and a proportion of invariant sites (GTR+I+G). Bayesian analyses were conducted using MrBayes v3.1.2 (Ronquist and Huelsenbeck, 2003). When analyzing only molecular data we specified the GTR+I+G model with flat priors on model parameters and random starting trees. Partitioned analyses of concatenated molecular and morphological data specified the same model of sequence evolution as above, with estimated state frequencies and equal rates of change across unordered morphological characters (Lewis, 2001). Trees and parameters were sampled every 1,000 generations from four chains (1 unheated), and ran at least twice for 4,000,000 generations for analyses of species exemplars and for 10,000,000 generations for analyses including all sequences. To ensure convergence of the chains, we examined the average standard deviation of split frequencies, which in every case was under 0.01 (range = 0.000550 – 0.005405). In every case the plots of log probability versus generation showed no trend after the first 100 samples (100,000 generations), which were discarded as burn in.

Both empirical and simulation studies have found Bayesian posterior probabilities (BPP) to be high relative to other measures of support (Alfaro and Holder, 2006; Douady et al., 2003; Pickett and Randle, 2005; Suzuki et al., 2002). The prior on the distribution of branch lengths explains these inflated support statistics (Yang and Rannala, 2005), at least in part. For example, using a single prior on the distribution of branch lengths across the tree results in relatively high probability estimates for short internal branches, even when data are simulated on a phylogeny with an internal branch length of zero (Yang, 2007a). This high probability seems to arise from the combination of the large sizes of molecular datasets with the great sensitivity in Bayesian

analyses to the branch length prior when models under comparison are similar in their performance (Yang, 2008). A prior on internal branch lengths that becomes shorter as sequences become longer has been shown to more accurately estimate support in empirical analyses (Yang, 2008). To more accurately estimate clade support with our data, we repeated all Bayesian analyses using the Windows executable version of mb3.12.2E (Yang, 2007a; Yang and Rannala, 2005), a version of MrBayes modified to apply separate priors for internal and external branch lengths. These re-analyses sampled the same length of chains and at the same frequency as initial MrBayes analyses. The mb3.12.2E software used a data size-dependent prior for the exponential distribution of internal branch lengths ($mean_i = 0.1*1143^{-2/3}$ for molecular data, $= 0.1*1156^{-2/3}$ for concatenated data, ≈ 0.0009 for all analyses), and a separate exponential prior for external branch lengths ($mean_e = 0.1$).

Finally, we compared different topologies derived from different analyses as well as *a priori* hypothesis of relationships between taxa. ML scores using the codon model of sequence evolution were calculated for: 1) a tree of exemplars consistent with the best tree resulting from Bayesian analysis of all data, 2) a tree of exemplars consistent with the best tree resulting from Bayesian analysis of all sequences, 3) the MP exemplars tree, and 4) a tree constraining the monophyly of subspecies of *P. gularis*. The implementation of codon models in tree search is many times slower than nucleotide models (e.g., in MrBayes). The hybrid approach of optimizing the codon model only on topologies found using different methods, e.g., Dávalos and Perkins (2008), reduced computational time and allowed the optimization and comparison of different topologies using a more appropriate, less parameter-rich model of sequence evolution. The resulting likelihood scores were compared using the Shimodaira-Hasegawa (SH) (Shimodaira and Hasegawa, 1999) and approximately unbiased (AU) (Shimodaira, 2002) tests in

consel 0.1 (Shimodaira and Hasegawa, 2001). A random tree was included in tree comparisons to ensure the range of differences between log-likelihoods was large, making assessments of significance comparable across analyses (van Rij et al., 2003).

RESULTS

Species limits among *Paroaria*

Sixteen external morphological characters were used for phylogenetic analyses (Appendix A). All of the species and subspecies, except for *Paroaria capitata fuscipes*, were found to be diagnosable based on external morphology (Fig. 3). The results obtained in this study justify revising the taxonomy of the genus *Paroaria*. A revised taxonomy is presented in Appendix B, and includes eight species: 1) *Paroaria coronata*, 2) *P. dominicana*, 3) *P. capitata* (including *P. c. fuscipes*); the three former subspecies of *P. gularis sensu lato*: 4) *P. gularis*, 5) *P. nigrogenis* and 6) *P. cervicalis*; and the two former subspecies of *P. baeri sensu lato*: 7) *P. baeri*, and 8) *P. xinguensis*. We use this new nomenclature hereafter. Discrete-character analyses of the cytochrome *b* gene were congruent with the eight basal taxonomic units delimited based on morphology. Each of the phylogenetic species recognized had two or more mutations potentially diagnostic with respect to congeners (Supplementary Data).

Within species, the highest distances (0.011–0.012 substitutions per site) were observed between the *gularis* individuals from Rondônia, Brazil and the samples from Peru and Ecuador at the westernmost edge of Amazonia (Table 2, Fig. 1). These distances were both significantly lower than simulated interspecific distances ($P < 0.001$), and higher than intraspecific distances ($P < 0.001$) (Fig. 4). More than 200 *gularis* study skins were examined (Table 1), and no diagnostic morphological characters were found to separate individuals from these two regions. Three estimates of θ (mean and the two extremes of the 95% confidence distribution) were used

to construct demographic scenarios of glacial isolation followed by limited post-glacial gene flow (see Methods). The genetic distances observed between these subpopulations were consistent with the modeled scenarios when the mean and high estimates of θ were used to ($P \geq$ 0.232). This scenario was, however, rejected with the low estimate of the θ parameter (P = 0.005), as the maximum distance observed was greater than most of the modeled distribution.

Observed distances between former subspecies now elevated to species based on discrete morphological and molecular characters (i.e., among *gularis*, *nigrogenis* and *cervicalis*) were significantly higher than within-species distances of other *Paroaria* species ($P < 0.001$, Fig. 4). Both the minimum (0.025 substitutions per site) and the mean distance (0.027, std. dev. = 0.002) between *cervicalis* and *gularis* were significantly lower than minimum interspecific distances among already recognized species ($P < 0.001$, Fig. 4).

A total of 23 mutations in *P. cervicalis* were fixed with respect to homologous sites in the sequence of its sister species *capitata*, and 2 mutations were unique and fixed with respect to all other *Paroaria* species (Supplementary Data). Three estimates of θ were used to construct demographic scenarios of continuous gene flow interrupted by the last interglacial and followed by restricted genetic exchange. The observed number of mutations fixed between *cervicalis* and *capitata* was significantly higher than the mutations modeled in every scenario ($P < 0.001$ with low and mean estimates of θ, and $P = 0.008$ with the high estimate of θ).

Phylogeny of *Paroaria*

Bayesian analyses of all individuals recovered a phylogeny where the genus *Paroaria*, each species, and relationships between *capitata*, *cervicalis*, *gularis*, *baeri*, and *xinguensis* were highly supported (Fig. 5). MP analyses of all individuals (not shown) were congruent with these results, but recovered *dominicana* as the most basal branch in the *Paroaria* phylogeny with low

support (55% bootstrap, see also Fig. 6C). The two most basal nodes within *Paroaria* received low support in Bayesian exemplar analyses, and different results were obtained in combined vs. only molecular analyses (Fig. 6). *P. coronata* or *dominicana* were the most basal species in different analyses (Fig. 6). These alternative trees were not significantly different given the molecular data ($P \geq 0.243$ in SH and AU tests), and the MP tree had the highest log-likelihood score (-3164.276) in ML analyses fitting a codon model.

A clade containing *capitata*, *cervicalis*, *gularis*, *baeri*, and *xinguensis* was consistently recovered across all analyses. The three taxa previously subsumed in the biological species *Paroaria gularis* (*cervicalis*, *gularis* and *nigrogenis*) did not form a clade, and constraining their monophyly had significantly lower likelihood given the data (log likelihood of constrained tree = -3266.820, $P \leq 0.001$ in SH and AU tests).

Discussion

Molecular stochasticity and systematic revisions

One of the goals of this study was to delimit *Paroaria* species based on a morphological revision and limited molecular data. The methods, prospects and challenges of species delimitation using molecular data constitute a rapidly growing field of study because this is a first step in phylogenetic analysis and all subsequent applications (Wiens, 2007). One approach, calculating distances between sequences from a single mitochondrial locus or DNA barcoding (Hebert et al., 2004; Kerr et al., 2007), has traditionally relied on the gap observed in the distribution of intra- and inter-specific distances. Grounded on this general principle, the parametric bootstrap of distances across the *Paroaria* phylogeny aimed to establish confidence intervals around intra- and inter-specific divergences of long-accepted species within the genus (Fig. 4).

Because it relies on a purported gap in the distribution of genetic distances (Wiemers and Fiedler, 2007), the barcoding approach is uninformative when distances between proposed species fall within the gap, as was the case for the *capitata*/*cervicalis* species pair and intra-specific variation within *gularis* (Fig. 4). Such intermediate distances could indicate within-population structure (Tavares and Baker, 2008), hybridization (Aliabadian et al., 2009), or recent isolation of the species in question (Johnson et al., 2004). Since these cases are precisely the ones that would benefit the most from molecular tools for species delimitation and identification (Moritz and Cicero, 2004), the general validity of the barcoding approach has been called into question (Rubinoff and Holland, 2005). As with many other birds (Aliabadian et al., 2009), the revision of *Paroaria* did not depend solely on comparisons to a pre-existing or modeled distance distribution. Instead, the molecular data were one source of evidence, complementary but independent from morphological character assessments. Distance-gap barcoding fails so long as the goals are species diagnoses and discovery for taxa with few or no trained taxonomists (Tautz et al., 2003). But when distance measures fail, alternative hypothesis-testing frameworks can take advantage of barcoding data. Together with character-based analyses, these new frameworks could enable barcoding initiatives to continue to contribute to species delimitation (DeSalle et al., 2005).

Another framework for delimiting species using molecular data is to consider reciprocally monophyletic groups as species, provided there are other sources of evidence (e.g., ecology, biogeography, morphology) suggesting these are species (Wiens and Penkrot, 2002), or otherwise units worthy of conservation (Moritz, 1994). This approach has been adopted recently in barcoding studies as a more rigorous way to ascertain species limits (Tavares and Baker, 2008). The probability of observing reciprocal monophyly by chance

when sequences evolve following a null coalescent model is very low, even with small samples (Rosenberg, 2007). For example, this probability for the *capitata*/*cervicalis* pair of sister species is ≤ 0.0069, so the data would significantly support their designation as species. Speciation, however, does not always lead to reciprocal monophyly, e.g., Heckman et al. (2007), Knowles and Carstens (2007). Conversely, in the absence of additional biological evidence not all monophyletic groups warrant species-level recognition, see Brito (2007), Russell et al. (2008), or Zink and Barrowclough (2008).

In this study, we examined external morphology with the goal of uncovering fixed characters that would indicate a long history of genetic isolation, and corroborated the interruption of female-mediated gene flow by identifying fixed mutations in a mitochondrial locus. Despite its use in recent studies, e.g., Tavares and Baker (2008), Ribas et al. (2005), Reddy (2008), this method has lacked the quantitative framework for evaluating the significance of observed substitutions. A single mutation, if fixed, might prove sufficient to propose species-level rank, e.g., Goldstein and Desalle (2003), but the validity of such hypotheses hinges completely on sampling. The coalescent simulations of the *capitata*/*cervicalis* species pair aimed to introduce quantitative testing to approaches based on character fixation. Restricted but continuous gene flow between subpopulations was modeled to exclude the possibility that the substitutions observed could have arisen from within-species population structure. Hence these simulations were more stringent than the test of reciprocal monophyly, which assumes a single panmictic population as the null (Rosenberg, 2007). Fixed characters (substitutions and morphology) provided evidence of the absence of gene flow, and both characters and genetic isolation were considered in evaluating species using these methods. Although the total absence of gene flow is not a requirement of valid

species, e.g, Barrowclough et al. (2005), Vallender et al. (2007), we assume that genetic isolation bolsters the case for delimiting species.

The simulations also tested the statistical power of small samples of genetic data. Both extremes and the mean of θ were analyzed to account more fully for stochasticity in the mutation process and sampling. The null models were rejected in every case, suggesting that matrilineal isolation between these two species is complete and older than assumed by existing systematic classification. Our results provide a starting point for further studies seeking to estimate the timing of divergence and mechanisms of speciation in these sister species, and data from multiple unlinked loci will be necessary to accomplish this goal, see Edwards and Beerli (2000).

In contrast with the *capitata* and *cervicalis* populations, no diagnostic external or molecular characters were found within *gularis*, and the few sequences sampled did not form reciprocally monophyletic groups (Fig. 5, probability of reciprocal monophyly by chance is > 0.1). Save for the low estimate of θ, coalescent models of post-glacial isolation were consistent with the high genetic distances observed. If taken as a probable estimate of population size and genetic diversity, the low extreme of θ would attribute observed distances to processes older than the last glacial maximum. Confidence intervals around θ spanned more than one order of magnitude (Fig. 2a) because sampling within *gularis* was minimal. The results suggest the molecular diversity observed is largely consistent with a glacial isolation/restricted post-glacial gene flow model. Sampling across Amazonia for cytochrome *b* (and other loci) will likely reduce the range of inferred molecular diversity, allowing more exacting tests of the role of glaciation in generating population structure in this tropical species.

Unlike the distance barcoding approach, the simulations used here provided a flexible framework for evaluating the significance of differentiation using different metrics (fixed mutations or distance, but other metrics could be used). Coalescent simulations, however, require the formulation of a speciation scenario linked to credible parameter estimates of time of isolation, rates of gene flow, generation time, mutation rate, and/or population size. The scenarios formulated in this study are the result of decades of research on geology and on neotropical avian speciation, and similar null hypotheses might not be available in all instances. Our coalescent simulations assumed there were no changes in population size in the modeled history of speciation as this was the simplest scenario. In many cases, e.g., in species whose population are known to have expanded following the retreat of ice sheets at higher latitudes, this assumption will not hold.

Further, the examination of > 300 specimens was the basis for formulating species limits. In agreement with many recent studies, e.g., Raxworthy et al. (2007), Rissler and Apodaca (2007), Shaffer and Thomson (2007), our analyses illustrate the need for multiple lines of evidence in species delimitation. Since the availability and analytical reach of molecular data are growing rapidly, another way to improve the prospects for species delimitation —even among well studied vertebrates— is to make the collection and analysis of morphological data as efficient as that of molecular data (Wheeler, 2005).

Phylogenetics of *Paroaria*

The revised *Paroaria* species limits enabled phylogenetic analyses encompassing greater diversity than hitherto acknowledged, providing the first estimate of relationships within the genus. The only prior systematic hypothesis, that implied by lumping *gularis*, *nigrogenis* and

cervicalis into a single species (Fig. 1), was rejected by the molecular data and not supported by any analysis (Figs. 5 and 6).

The analyses presented here provided insights into approaches that haven't been adopted widely in phylogenetic applications. Two methods were applied: 1) Bayesian inference with different prior distributions for terminal and internal branches, and 2) tree comparisons fitting a ML codon model to the data. Results from both methods were consistent: alternative resolutions of deep nodes recovered in different Bayesian analyses were not significantly different under the ML model. Traditional Bayesian analyses using a single prior distribution for all branch lengths were congruent with two-prior analyses of concatenated data and sequences from all taxa (Figs. 5 and 6b), but not for concatenated data from exemplars (Fig. 6a). The combination of differential taxon sampling and the underlying goal of the two-prior method explains these results. Greater taxon sampling increases the accuracy of phylogenetic inference by improving the detection of mutations thus "breaking up" long branches (Dávalos and Perkins, 2008; Delsuc et al., 2005; Rodríguez-Ezpeleta et al., 2007). The goal of the two-prior approach is to correct the bias introduced by the high ratio of characters to nodes in molecular phylogenetics (Yang, 2007a; Yang, 2008). All-taxa analyses converged on similar topologies despite different priors on the distribution of internal branches because the detection of character change along internal branches is made easier by the inclusion of many terminals, even when such changes were supported only by morphological characters (Fig. 5). Since only morphological characters were available to resolve relationships between *baeri* and *xinguensis*, and the detection of changes along internal branches becomes less probable with fewer taxa, the short prior on internal branch lengths collapsed this branch (Fig. 6a). Applying the short prior to concatenated data is

appropriate insofar as the ratio of characters to branches remains high and/or expanded taxon sampling makes the detection of changes along internal branches straightforward.

Although more data are needed to conclusively resolve relationships of *Paroaria*, the systematic and phylogenetic hypotheses developed here are relevant to both setting conservation priorities and understanding the biogeography of South America (Porzecanski and Cracraft, 2005). As agriculture and ranching continue to expand in both forested and arid areas of endemism (da Silva and Bates, 2002; Klink and Machado, 2005; Soares-Filho et al., 2006), clarifying the species limits and origin of the biota remains crucial to enabling effective conservation in the region. Our analyses provide a first quantitative molecular framework to attain those goals, made more urgent by the global trade of captive *P. capitata* and *coronota* (UNEP-WCMC, 2009).

ACKNOWLEDGEMENTS

We thank Joel Cracraft, Leo Joseph, Nancy Simmons, and José Tello for comments on early drafts of this manuscript. ALP received generous support from the Frank M. Chapman Memorial Fund, and the International Graduate Student Fellowship Program of the Center for Biodiversity and Conservation at the AMNH. We thank Joseph W. Brown and Josh Rest for making available their estimates of avian mutation rates. We are very grateful to the curators and collection managers at the American Museum of Natural History, Academy of Natural Sciences of Philadelphia, Field Museum, Louisiana State University Museum of Zoology, Museo Argentino de Ciencias Naturales Bernardino Rivadavia, Muséum National d'Historie Naturelle, Museo Nacional de Historia Natural de Montevideo, Museo de Historia Natural Noel Kempff Mercado, Colección Ornitológica Phelps, Univ. of Kansas Museum of Natural History, and Univ. of Washington Burke Museum for kindly providing us with access to their collections, collection

data, tissue and/or skin samples for this study. Without these collections this project would not have been possible. Jacqueline Weicker, Ana Maria Saavedra, and Santiago Claramunt provided invaluable assistance in the field. We thank Katrina Cook from the Natural History Museum (Tring) for making available photographs of specimens under her care, and Thomas Donegan for facilitating the link to NHM. José Frances kindly helped prepare supplementary data tables. Amy L. Russell provided support in *ms* simulations. This material is based upon work supported by the National Science Foundation under Grant No. NAG5-12333 and is also a contribution from the Cullman Research Facility in the Department of Ornithology, American Museum of Natural History, and has received support from the Lewis B. and Dorothy Cullman Program for Molecular Systematics Studies, a joint initiative of The New York Botanical Garden and The American Museum of Natural History. Some bioinformatic analyses were carried out on the freely available Bioportal (www.bioportal.uio.no), and another part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation.

References

Alfaro, M.E., Holder, M.T., 2006. The posterior and the prior in Bayesian phylogenetics. Annu. Rev. Ecol. Evol. Syst. 37, 19-42.

Aliabadian, M., Kaboli, M., Nijman, V., Vences, M., 2009. Molecular Identification of Birds: Performance of Distance-Based DNA Barcoding in Three Genes to Delimit Parapatric Species. PLoS ONE 4, e4119.

Assine, M.L., Soares, P.C., 2004. Quaternary of the Pantanal, west-central Brazil. Quatern. Int. 114, 23-34.

Barrowclough, G.F., Groth, J.G., Mertz, L.A., Gutierrez, R.J., 2005. Genetic structure, introgression, and a narrow hybrid zone between northern and California spotted owls (*Strix occidentalis*). Mol. Ecol. 14, 1109-1120.

Beerli, P., 2004. Migrate: documentation and program, part of LAMARC.

Beerli, P., Felsenstein, J., 1999. Maximum-likelihood estimation of migration rates and effective population numbers in two populations using a coalescent approach. Genetics 152, 763-773.

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in *n* subpopulations by using a coalescent approach. Proc. Natl. Acad. Sci. USA 98, 4563-4568.

Bond, J., Meyer de Schauensee, R., 1939. Descriptions of new birds from Bolivia. Part I: Oscines. Not. Nat. (Phila.) 12, 1-5.

Boulet, M., Gibbs, H.L., 2006. Lineage origin and expansion of a Neotropical migrant songbird after recent glaciation events. Mol. Ecol. 15, 2505-2525.

Braun, M.J., Finch, D.W., Robbins, M.B., Schmidt, B.K., 2000. A Field Checklist of the Birds of Guyana. Smithsonian Institution, Washington , D.C.

Brito, P.H., 2007. Contrasting patterns of mitochondrial and microsatellite genetic structure among Western European populations of tawny owls (*Strix aluco*). Mol. Ecol. 16, 3423-3437.

Brown, J.W., Rest, J.S., Garcia-Moreno, J., Sorenson, M.D., Mindell, D.P., 2008. Strong mitochondrial DNA support for a Cretaceous origin of modern avian lineages. BMC Biol. 6, 6.

Burns, K.J., Naoki, K., 2004. Molecular phylogenetics and biogeography of Neotropical tanagers in the genus *Tangara*. Mol. Phylogenet. Evol. 32, 838-854.

Cory, C.B., Hellmayr, C.E., 1925. Catalogue of the Birds of the Americas and the Adjacent Islands. Zoological Series of the Field Museum of Natural History XIII, 1-390.

Cracraft, J., 1983. Species concepts and speciation analysis. Curr. Ornithol. 1, 158-187.

Cracraft, J., 1992. The species of the Birds-of-Paradise (Paradisaeidae): applying the phylogenetic species concept to a complex pattern of diversification. Cladistics 8, 1-43.

Cracraft, J., Feinstein, J., Vaughn, J., Helm-Bychowski, K., 1998. Sorting out tigers (*Panthera tigris*): mitochondrial sequences, nuclear inserts, systematics, and conservation genetics. Anim. Conserv. 1, 139-150.

da Silva, J.M.C., Bates, J.M., 2002. Biogeographic patterns and conservation in the South American Cerrado: A tropical savanna hotspot. Bioscience 52, 225-233.

Dávalos, L.M., Perkins, S.L., 2008. Saturation and base composition bias explain phylogenomic conflict in *Plasmodium*. Genomics 91, 433-442.

Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. Nat. Rev. Genet. 6, 361-375.

DeSalle, R., Egan, M.G., Siddall, M., 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Phil. Trans. R. Soc. B 360, 1905-1916.

Desjardins, P., Morais, R., 1990. Sequence and gene organization of the chicken mitochondrial genome: A novel gene order in higher vertebrates. J. Mol. Biol. 212, 599-634.

Douady, C.J., Delsuc, F., Boucher, Y., Doolittle, W.F., Douzery, E.J.P., 2003. Comparison of Bayesian and Maximum Likelihood Bootstrap Measures of Phylogenetic Reliability. Mol. Biol. Evol. 20, 248-254.

Ebach, M., Holdrege, C., 2005. DNA barcoding is no substitute for taxonomy. Nature 434, 697.

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

Elias, M., Hill, R.I., Willmott, K.R., Dasmahapatra, K.K., Brower, A.V.Z., Mallet, J., Jiggins, C.D., 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. B 274, 2881-2889.

Fu, Y.X., Li, W.H., 1993. Maximum Likelihood Estimation of Population Parameters. Genetics 134, 1261-1270.

Goldstein, P.Z., Desalle, R., 2003. Calibrating phylogenetic species formation in a threatened insect using DNA from historical specimens. Mol. Ecol. 12, 1993-1998.

Guindon, S., Gascuel, O., 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52, 696-704.

Haffer, J., 1985. Avian zoogeography in the Neotropical lowlands. Ornithological Monographs 36, 113-146.

Hasegawa, M., Kishino, H., Yano, T., 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22, 160-174.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M., 2004. Identification of Birds through DNA Barcodes. PLoS. Biol. 2, e312.

Heckman, K.L., Mariani, C.L., Rasoloarison, R., Yoder, A.D., 2007. Multiple nuclear loci reveal patterns of incomplete lineage sorting and complex species history within western mouse lemurs (*Microcebus*). Mol. Phylogenet. Evol. 43, 353-367.

Hellmayr, C.E., 1938. Catalogue of the Birds of the Americas and the Adjacent Islands. Zoological Series of the Field Museum of Natural History XIII, 58-67.

Hilty, S.L., 2002. Birds of Venezuela. Princeton University Press, Princeton.

Hilty, S.L., Brown, W.L., 1986. A Guide to the Birds of Colombia. Princeton University Press, Princeton.

Hudson, R.R., 2002. Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18, 337-338.

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

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

Kimura, M., 1981. Estimation of evolutionary distances between homologous nucleotide sequences. Proc. Natl. Acad. Sci. USA 78, 454-458.

Klink, C.A., Machado, R.B., 2005. Conservation of the Brazilian Cerrado. Conserv. Biol. 19, 707-713.

Knowles, L.L., Carstens, B.C., 2007. Delimiting Species without Monophyletic Gene Trees. Syst. Biol. 56, 887 - 895.

Lee, K., Feinstein, J., Cracraft, J., 1997. The phylogeny of ratite birds: resolving conflicts between molecular and morphological data sets. In: Mindell, D. (Ed.), Avian molecular evolution and systematics. Academic Press, San Diego, pp. 173–211.

Lessa, E.P., Cook, J.A., Patton, J.L., 2003. Genetic footprints of demographic expansion in North America, but not Amazonia, during the Late Quaternary. Proc. Natl. Acad. Sci. USA 100, 10331-10334.

Lewis, P.O., 2001. A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. Syst. Biol. 50, 913-925.

Mayr, E., 1942. Systematics and the Origin of Species. Columbia University Press, New York.

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

Moritz, C., 1994. Defining 'Evolutionarily Significant Units' for conservation. Trends Ecol. Evol. 9, 373-375.

Moritz, C., Cicero, C., 2004. DNA Barcoding: Promise and Pitfalls. PLoS. Biol. 2, e354.

Mundy, N.I., Unitt, P., Woodruff, D., 1997. Skin from feet of museum specimens as a non-destructive source of DNA for avian genotyping. Auk 114, 126.

Muse, S.V., Gaut, B.S., 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. Mol. Biol. Evol. 11, 715-724.

NatureServe, 2003. Ecological Systems of Latin American and the Caribbean Database, version 1.01. NatureServe, Arlington, VA. .

NatureServe, 2007. InfoNatura: Animals and Ecosystems of Latin America, Version 5.0. NatureServe, Arlington, Virginia.

Nylander, J.A.A., 2004. MrAIC.pl. Evolutionary Biology Centre, Uppsala University, Uppsala, p. Program distributed by the author.

Parker, T.A., III, Stotz, D.F., Fitzpatrick, J.W., 1996. Ecological and Distributional Databases. In: Stotz, D.F., Fitzpatrick, J.W., Parker, I.T.A., Moskovits, D.K. (Eds.), Neotropical Birds: Ecology and Conservation. The University of Chicago Press, Chicago.

Paynter, R.A., Storer, R.W., 1970. Check-list of birds of the world. A continuation of the work of James L. Peters. Museum of Comparative Zoology, Cambridge, Massachusetts.

Pickett, K.M., Randle, C.P., 2005. Strange Bayes indeed: uniform topological priors imply non-uniform clade priors. Mol. Phylogenet. Evol. 34, 203-211.

Porzecanski, A.L., Cracraft, J., 2005. Cladistic analysis of distributions and endemism (CADE): using raw distributions of birds to unravel the biogeography of the South American aridlands. J. Biogeogr. 32, 261-275.

Posada, D., Buckley, T.R., 2004. Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst. Biol. 53, 793-808.

Posada, D., Crandall, K.A., 1998. MODELTEST: testing the model of DNA substitution. Bioinformatics 14, 817-818.

Proctor, N.S., Lynch, P.J., 1993. Manual of Ornithology. Yale University Press, New Haven.

Rambaut, A.E., Grassly, N.C., 1997. SEQ-GEN: An application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13, 235-238.

Raxworthy, C.J., Ingram, C.M., Rabibisoa, N., Pearson, R.G., 2007. Applications of Ecological Niche Modeling for Species Delimitation: A Review and Empirical Evaluation Using Day Geckos (*Phelsuma*) from Madagascar. Syst. Biol. 56, 907 - 923.

Reddy, S., 2008. Systematics and biogeography of the shrike-babblers (*Pteruthius*): Species limits, molecular phylogenetics, and diversification patterns across southern Asia. Mol. Phylogenet. Evol. 47, 54-72.

Ren, F., Tanaka, H., Yang, Z., 2005. An empirical examination of the utility of codon-substitution models in phylogeny reconstruction. Syst. Biol. 54, 808-818.

Ribas, C.C., Gaban-Lima, R., Miyaki, C.Y., Cracraft, J., 2005. Historical biogeography and diversification within the Neotropical parrot genus *Pionopsitta* (Aves: Psittacidae). J. Biogeogr. 32, 1409-1427.

Ribas, C.C., Miyaki, C.Y., 2004. Molecular systematics in *Aratinga* parakeets: species limits and historical biogeography in the '*solstitialis*' group, and the systematic position of *Nandayus nenday*. Mol. Phylogenet. Evol. 30, 663-675.

Ridgely, R.S., Tudor, G., 1989. The Birds of South America: Volume I: the Oscine Passerines. University of Texas Press, Austin.

Rissler, L.J., Apodaca, J.J., 2007. Adding More Ecology into Species Delimitation: Ecological Niche Models and Phylogeography Help Define Cryptic Species in the Black Salamander (*Aneides flavipunctatus*). Syst. Biol. 56, 924 - 942.

Rodríguez-Ezpeleta, N., Brinkmann, H., Roure, B., Lartillot, N., Lang, B.F., Philippe, H., 2007. Detecting and Overcoming Systematic Errors in Genome-Scale Phylogenies. Syst. Biol. 56, 389 - 399.

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

Rosenberg, N.A., 2007. Statistical tests for taxonomic distinctiveness from observations of monophyly. Evolution 61, 317-323.

Rubinoff, D., Holland, B., 2005. Between Two Extremes: Mitochondrial DNA is neither the Panacea nor the Nemesis of Phylogenetic and Taxonomic Inference. Syst. Biol. 54, 952-961.

Russell, A.L., Goodman, S.M., Fiorentino, I., Yoder, A.D., 2008. Population Genetic Analysis of *Myzopoda* (Chiroptera: Myzopodidae) in Madagascar. J. Mammal. 89, 209-221.

Sclater, P.L., 1890. Catalogue of the birds in the British Museum. Vol. XV. The Trustees of the British Museum, London.

Shaffer, H.B., Thomson, R.C., 2007. Delimiting Species in Recent Radiations. Syst. Biol. 56, 896 - 906.

Sharpe, R.B., 1888. Catalogue of the Passeriformes or Perching Birds in the Collection of the British Museum, Vol. XII (Fringilliformes). The Trustees of the British Museum, London.

Shimodaira, H., 2002. An approximately unbiased test of phylogenetic tree selection. Syst. Biol. 51, 492-508.

Shimodaira, H., Hasegawa, M., 1999. Multiple comparisons of log-likelihoods with applications to phylogenetic inference. Mol. Biol. Evol. 16, 1114–1116.

Shimodaira, H., Hasegawa, M., 2001. Consel: for assessing the confidence of phylogenetic tree selection. Bioinformatics 17, 1246-1247.

Short, L.L., 1975. A zoogeographic analysis of the South American Chaco avifauna. Bull. Am. Mus. Nat. Hist. 254, 165-352.

Sick, H., 1950. Uma nova raça de cardeal procedente do Brasil central, "*Paroaria baeri xinguensis*" N. ssp. (Fringillidae, Aves). Rev. Bras. Biol. 10, 465-468.

Sick, H., 1993. Birds in Brazil. A Natural HIstory. Princeton University Press, Princeton.

Smithe, F.B., 1981. Naturalist's color guide. American Museum of Natural History, New York.

Soares-Filho, B.S., Nepstad, D.C., Curran, L.M., Cerqueira, G.C., Garcia, R.A., Ramos, C.A., Voll, E., McDonald, A., Lefebvre, P., Schlesinger, P., 2006. Modelling conservation in the Amazon basin. Nature 440, 520-523.

Suzuki, Y., Glazko, G.V., Nei, M., 2002. Overcredibility of molecular phylogenies obtained by Bayesian phylogenetics. Proc. Natl. Acad. Sci. USA 99, 16138-16143.

Swofford, D.L., 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Sinauer Associates, Sunderland, Massachusetts.

Tautz, D., Arctander, P., Minelli, A., Thomas, R.H., Vogler, A.P., 2003. A plea for DNA taxonomy. Trends Ecol. Evol. 18, 70-74.

Tavaré, S., 1986. Some probabilistic and statistical problems on the analysis of DNA sequences. Lect. Math. Life Sci. 17, 57-86.

Tavares, E., Baker, A., 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evol. Biol. 8, 81.

Tordoff, H.B., 1954. A systematic study of the avian family Fringillidae based on the structure of the skull. Misc. Publ. Mus. Zool. Univ. Mich. 81, 1-41.

UNEP-WCMC, 2009. UNEP-WCMC Species Database: CITES-Listed Species.

Vallender, R., Robertson, R.J., Friesen, V.L., Lovette, I.J., 2007. Complex hybridization dynamics between golden-winged and blue-winged warblers (*Vermivora chrysoptera* and *Vermivora pinus*) revealed by AFLP, microsatellite, intron and mtDNA markers. Mol. Ecol. 16, 2017-2029.

van der Hammen, T., Hooghiemstra, H., 2000. Neogene and Quaternary history of vegetation, climate, and plant diversity in Amazonia. Quaternary Sci. Rev. 19, 725-742.

van Rij, R.P., Worobey, M., Visser, J.A., Schuitemaker, H., 2003. Evolution of R5 and X4 human immunodeficiency virus type 1 gag sequences in vivo: evidence for recombination. Virology 314, 451-459.

Wheeler, Q.D., 2005. Losing the plot: DNA "barcodes" and taxonomy. Cladistics 21, 405-407.

Wiemers, M., Fiedler, K., 2007. Does the DNA barcoding gap exist? - a case study in blue butterflies (Lepidoptera: Lycaenidae). Front. Zool. 4, 8.

Wiens, J.J., 2007. Species Delimitation: New Approaches for Discovering Diversity. Syst. Biol. 56, 875 - 878.

Wiens, J.J., Penkrot, T.A., 2002. Delimiting Species Using DNA and Morphological Variation and Discordant Species Limits in Spiny Lizards (*Sceloporus*). Syst. Biol. 51, 69 - 91.

Yang, Z., 1996. Maximum-likelihood models for combined analyses of multiple sequence data. J. Mol. Evol. 42, 587-596.

Yang, Z., 2007a. Fair-Balance Paradox, Star-tree Paradox, and Bayesian Phylogenetics. Mol. Biol. Evol. 24, 1639-1655.

Yang, Z., 2007b. PAML 4: Phylogenetic Analysis by Maximum Likelihood. Mol. Biol. Evol. 24, 1586-1591.

Yang, Z., 2008. Empirical evaluation of a prior for Bayesian phylogenetic inference. Philos. Trans. R. Soc. Lond. B Biol. Sci.

Yang, Z., Rannala, B., 2005. Branch-length prior influences Bayesian posterior probability of phylogeny. Syst. Biol. 54, 455-470.

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

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

Zink, R.M., Blackwell-Rago, R.C., 2000. Species Limits and Recent Population History in the Curve-Billed Thrasher. Condor 102, 881-886.

Tables

Table 1. Study skins examined for *Paroaria*. Specimen counts include only adult individuals with locality data. *Paroaria* specimens were examined at the American Museum of Natural History (AMNH), the Field Museum (FMNH), the Academy of Natural Sciences (ANSP) and the Museo Argentino de Ciencias Naturales Bernardino Rivadavia (MACN). Dash lines indicate that a certain taxon was not represented in a given collection.

Species	Subspecies	AMNH	FMNH	MACN	ANSP	Total
Paroaria coronata		115	40	80	12	247
Paroaria dominicana		31	14	-	-	45
Paroaria capitata	*capitata*	31	10	21	6	68
	fuscipes	-	-	-	6	6
Paroaria gularis	*cervicalis*	8	6	-	5	19
	gularis	152	51	1	16	220
	nigrogenis	28	3	-	4	35
Paroaria baeri	*baeri*	2	1	-	-	3
	xinguensis	2	-	-	-	2
Total						645

Table 2. *Paroaria* sequenced for this study, including locality and museum voucher number. A DOT number is also provided for AMNH Department of Ornithology tissue samples, and field collector numbers [in brackets] are given for vouchers without assigned numbers in US collections. For samples obtained from museum study skins, the date of specimen collection is also given. Species names correspond to the traditional taxonomy. Abbreviations: AMNH, American Museum of Natural History; ANSP, Academy of Natural Sciences of Philadelphia; FMNH, Field Museum; LSU, Louisiana State University Museum of Zoology; MNHN, Museo Nacional de Historia Natural, Montevideo; MNKM, Museo de Historia Natural Noel Kempff Mercado; Phelps, Colección Ornitológica Phelps; KUNHM, Univ. of Kansas Museum of Natural History; UWBM, Univ. of Washington Burke Museum.

Taxon with sample ID	Voucher numbers	GenBank No.	Locality
Paroaria c. capitata 1	AMNH 833309, DOT 2292	FJ715667	Bolivia, Santa Cruz, Velasco, near localidad El Tuná, 300 m N of Rio Mercedes
Paroaria c. capitata 2	AMNH 833308, DOT 2308	FJ715669	Bolivia, Santa Cruz, Velasco, near localidad El Tuná, 300 m N of Rio Mercedes
Paroaria c. capitata 3	MNKM 2698, DOT 2303	FJ715668	Bolivia, Santa Cruz, Velasco, near localidad El Tuná, 300 m N of Rio Mercedes
Paroaria c. capitata 4	MNKM 2691, DOT 6173	FJ715670	Bolivia, Santa Cruz, Velasco, near localidad El Tuná, 300 m N of Rio Mercedes
Paroaria c. capitata 5	KUNHM [ATP 99175]	FJ715671	Paraguay, depto. Alto Paraguay; 8 km W Río Negro
Paroaria c. capitata 6	KUNHM 90312	FJ715672	Paraguay, Ñeembucú; 26 35.3’S, 58 08.4’W
Paroaria c. capitata 7	FMNH 256183 (1956)	FJ715666	Argentina, Salta
Paroaria coronata 1	AMNH 833307, DOT 2200	FJ715650	Bolivia, Santa Cruz, Cordillera, Izozog, Comunidad Karapari, Estancia San Julián, 1000 m W of Parapetí
Paroaria coronata 2	MNKM 2465, DOT 2235	FJ715651	Bolivia, Santa Cruz, Cordillera, Izozog, Comunidad Karapari, Estancia San Julián,

			1000 m W of Parapetí
Paroaria coronata 3	LSU [HDC 362]	FJ715660	Bolivia, Santa Cruz, Cordillera; Estancia Perforación, ca. 130 Km. E Charagua
Paroaria coronata 4	KUNHM [KSB 247]	FJ715661	Paraguay, Pr. Hayes, Río Verde
Paroaria coronata 5	KUNHM 90128	FJ715662	Paraguay, Ñeembucú; 26 35.3'S, 58 08.4'W
Paroaria coronata 6	AMNH 833846, DOT 2330	FJ715652	Uruguay, Río Negro, Estancia Las Flores, E of Ruta 3 (km 270), 13 km along Ruta 20 to Pueblo Greco
Paroaria coronata 7	AMNH 833847, DOT 2335	FJ715653	Uruguay, Río Negro, Estancia Las Flores, E of Ruta 3 (km 270), 13 km along Ruta 20 to Pueblo Greco
Paroaria coronata 8	AMNH 833850, DOT 2339	FJ715656	Uruguay, Río Negro, Estancia Las Flores, E of Ruta 3 (km 270), 13 km along Ruta 20 to Pueblo Greco
Paroaria coronata 9	AMNH 833851, DOT 2340	FJ715657	Uruguay, Río Negro, Estancia Las Flores, E of Ruta 3 (km 270), 13 km along Ruta 20 to Pueblo Greco
Paroaria coronata 10	AMNH 833848, DOT2337	FJ715654	Uruguay, San José, Ruta 3, km 114, 500 m N of Arroyo Chamizo
Paroaria coronata 11	AMNH 833849, DOT 2338	FJ715655	Uruguay, San José, Ruta 3, km 114, 500 m N of Arroyo Chamizo
Paroaria coronata 12	AMNH 833852, DOT 2343	FJ715658	Uruguay, Florida, ruta 77, km 4, between pueblos Independencia and 25 de Agosto
Paroaria coronata 13	MNHN 5621	FJ715659	Uruguay, Durazno.
Paroaria coronata 14	UWBM 70248	FJ715663	Argentina, Corrientes, Corrientes
Paroaria g. cervicalis 1	AMNH 792268 (1964)	FJ715678	Bolivia, El Beni, Río Itenez
Paroaria g. cervicalis 2	AMNH 792274 (1965)	FJ715677	Bolivia, El Beni, Arroyo Mercedes
Paroaria g. gularis 3	FMNH 390056	FJ715674	Brazil, Rondônia, Cachoeira Nazaré, W bank Rio Jiparana

Paroaria g. gularis 4	FMNH 390055	FJ715673	Brazil, Rondônia, Cachoeira Nazaré, W bank Rio Jiparana
Paroaria g. gularis 5	LSU 156811	FJ715676	Peru, Ucayali, West bank Río Shesha, 65 Km. ENE Pucallpa
Paroaria g. gularis 6	ANSP 183503	FJ715675	Ecuador, Napo, Zancudo Cocha
Paroaria g. nigrogenis 7	FMNH 297813 (1975)	FJ715680	Colombia, Meta, Carimagua
Paroaria g. nigrogenis 8	Phelps 80841	FJ715681	Venezuela, Bolívar, Río Caroni, Isla El Hornero
Paroaria g. nigrogenis 9	Phelps 80842	FJ715679	Venezuela, Bolívar, Río Caroni, Isla El Hornero
Paroaria dominicana 1	FMNH 392738	FJ715665	Brazil, Sergipe, Caninde do São Francisco, Curituba, Fazenda Serrote
Paroaria dominicana 2	FMNH 392735	FJ715664	Brazil, Sergipe, Caninde do São Francisco, Curituba, Fazenda Serrote
Paroaria b. baeri 1	AMNH 520206 (1906)	FJ715682	Brazil, Goiás, Rio Araguaya
Paroaria b. baeri 2	AMNH 520207 (1906)	FJ715683	Brazil, Goiás, Rio Araguaya

Table 3. Primers used to amplify tissue and skin samples in this study. Non-specific primers were used in conjunction with specific primers. Primers were developed in the Monell Molecular Laboratory in the course of this study, except for: * Lee et al. (1997), ** Cracraft et al. (1998). L and H refer to light and heavy strands, respectively. Numbers refer to base position on the *Gallus gallus* mitochondrial genome (Desjardins and Morais, 1990). Numbers refer to base position in the cytochrome *b* gene, from 1 to 1143.

Primer name	Sequence
Non-specific	
L14851	5-CCTACTTAGGATCATTCGCCCT-3
L14857	5-GGGTCTTTCGCCCTATCAAT-3
L14827	5-CCACACTCCACACAGGCCTAATTAA-3*
L15236	5-TTCCTATACAAAGAAACCTGAAA-3
L15507	5-CCAGACCTCCTAGGAGACCCAGA-3**
H15149	5-TGCAGCCCCTCAGAATGATATTTGTCCTCA-3**
H15712	5-TTCCTATACAAAGAAACCTGAAA-3*
H15710	5-ATAGCGTAGGCGAATAGGAAGTATC-3 **
H16058	5-TTGGTTCACAAGACCAATGTT-3
Paroaria	
L14818	5-CCTACACACCCCTACACTCCGG-3
L160	5-CACAGGTCTCCTACTAGCCA-3
L275	5-GCAAACGGAGCCTCATTCTTCTT-3
L370	5-GAGTCATCCTCCTCCTAACCC-3
L533	5- GACAACCCCACCCTCACTCG-3

L571	5-TCCTCCCATTTGTCATCG-3
L763	5-AACCTCCT(T/C)GGAGACCCAG-3
L764	5-CCTAACCTCCTCGGAGACCCAGA-3
L821	5-CATATCAAACCAGAATGATA-3
L959	5-TACGCTCCATAACATTCCGCCC-3
H211N	5-AACCGAATTGTACGTCTCGG-3
H422	5-ATGACGGTAGC(G/C)CCTCAGA-3
H 517	5-AAGAATCGAGTTAGGGTTGGATT-3
H620	5-GGGACACCTGTTGGGTTGT-3
H814	GCAAATAGGAA(A/G)TATCATTC-3
H825	5-GGATAGCGTATGCGAATAGG-3
H1013	5-ACAGGTTGGCTGCCAACTCAGG-3

Figure legends

Figure 1. Sampled localities, areas of endemism, and distribution of *Paroaria*. All points illustrated were sampled in molecular analyses, except *P. baeri xinguensis*. Areas of endemism were based on NatureServe's (2003) classification of ecological systems of Latin America and the Caribbean. Three moist forest ecoregions; the Guianan Uplands and Highlands, the Guianan Eastern Lowlands and Amazonia, were merged into a single greater Amazonia area. Two areas, the Beni (white crosshatch) and the Pantanal (thatched), were added based on previously identified ornithological areas of endemism (Porzecanski and Cracraft, 2005). Outlines of species and subspecies distributions were obtained from the NatureServe distributional databases (NatureServe, 2007; Parker et al., 1996). The outline for *Paroaria gularis* was made continuous along Guyana, Surinam, French Guiana and northeastern Brazil.

Figure 2. Schematic of coalescent simulations modeling two histories of population subdivision. A. Simulated population split between western Amazonia and the remainder of the range of *P. gularis*. The selected model of sequence evolution was the Kimura 3-parameter (Kimura, 1981), which accounts for variable base frequencies, variable transversion frequencies and equal transition frequencies (rC-T = rA-G = 1.29102, rA-T = rC-G = 0.00001, rG-T = rA-C = 0.27822). The θ parameter for the population as a whole was estimated at 0.0093 individuals x mutations per site per generation (95% confidence interval = 0.0030–0.0406). Gene flow between subpopulations was set at 1 migrant per generation in each direction. B. Simulated subdivision between *Paroaria cervicalis* (small subpopulation) and *P. capitata* (large subpopulation). The selected model of sequence evolution was the Hasegawa-Kishino-Yano (Hasegawa et al., 1985), which accounts for variable base

frequencies and a transition/transversion rate ration (κ = 26.34707). The θ parameter for the population as a whole was estimated at 0.0061 individuals x mutations per site per generation (95% confidence interval = 0.0022–0.0154). After the split, the subpopulations were modeled to have maintained low and asymmetric gene flow between them, such that the smaller population received 5 migrants each generation and the larger population receives 1 migrant each generation.

Figure 3. Differences in bill color, bill shape, head plumage patterns, and head plumage colors for the species of *Paroaria* recognized in this study. The pictures were based on an illustration by W. Hart, in Sharpe (1888).

Figure 4. Percent histogram of ML distances (*r*A-C 4.68319, *r*A-G 29.6912, *r*A-T 1.27583, *r*C-G 0.00030, *r*C-T 32.2672, alpha = 0.15056) from data simulated using a codon model of sequence evolution (rate of non-synonymous/synonymous substitutions = 0.01766, transition/transversion rate ratio = 12.59209, observed codon frequencies). Pairwise comparisons excluded *P. gularis, P. nigrogenis*, and *P. cervicalis*. Black arrows depict observed distances. Results did not differ when data were simulated using a nucleotide model with GTR+G parameters.

Figure 5. Phylogram resulting from analyses of concatenated molecular and morphological data (mean log likelihood of 4 runs = -3846.83). Initial MrBayes analyses applied a single prior on an exponential distribution of branch lengths (mean = 0.1). This default setting effectively assigns a relatively high probability on short internal branches (Yang and Rannala, 2005). Bayesian posterior probabilities (BPP) were re-estimated using a size-dependent prior on an exponential distribution for internal branch lengths ($mean_i$ = 0.0002 for all data sets),

and a separate prior for external branch lengths ($mean_e = 0.1$) (Yang, 2007a; Yang, 2008). Values above branches are percent BPP of analyses of concatenated data.

Figure 6. Alternative topologies supported by different combinations of data and analysis. Bayesian phylograms were obtained using one prior on the distribution of branch lengths across the tree. Bayesian posterior probabilities (BPP), shown as percent, were obtained using different priors for internal and external branch lengths (see Figure 5). A. Phylogeny of *Paroaria* species exemplars recovered from Bayesian analyses of concatenated molecular and morphological data (mean log likelihood of 4 runs = -3621.66). B. Tree congruent with results from Bayesian analysis of sequences from all individuals (mean log likelihood of 4 runs = -3704.49), BPP resulted from analyses of all sequences. Branch lengths were optimized using PAML to fit exemplar sequences to this topology (log likelihood of codon model = -3168.775). C. Tree resulting from MP analysis of concatenated molecular and morphological data (tree length = 450), support values are percent nonparametric bootstrap (BS).

Appendix A. Morphological characters.

External morphological characteristics used to diagnose species of *Paroaria*, as well as concatenated with molecular data in combined phylogenetic analyses. A diagnosis of each species can be found in Appendix B. The table indicates the state name as well as the code used for that state in the matrix (between straight brackets; n/a = non-applicable). Characters were coded and analyzed as unordered.

	Characters															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Phylogenetic species	Head color	Back color	Dorsal spots	Bill color	Tarsus color	Bib color	Distinct orbital	Primaries	Bill shape	Crural feathers	Chest	Back gloss	Black mask	Chin spot	Crest	Bib
P. coronata	red [0]	gray [0]	few, small [1]	pearly gray [0]	gray [0]	red [0]	no [0]	light edged [1]	thick [1]	gray + white [1]	plain [0]	none [0]	no [0]	no [0]	yes [1]	yes [1]
P. dominicana	dark red [1]	mixed [1]	many, large [2]	pearly gray [0]	gray [0]	red [0]	no [0]	light edged [1]	thick [1]	black + white [0]	plain [0]	none [0]	no [0]	no [0]	no [0]	yes [1]
P. capitata	dark red [1]	black [2]	none [0]	yellow [1]	yellow [1]	black [1]	no [0]	no [0]	thin [0]	black + white [0]	plain [0]	glossy [1]	no [0]	no [0]	no [0]	yes [1]
P. cervicalis	dark red [1]	black [2]	none [0]	black + yellow [2]	black [2]	black [1]	no [0]	no [0]	thin [0]	black + white [0]	plain [0]	glossy [1]	no [0]	no [0]	no [0]	yes [1]
P. gularis	dark red [1]	black [2]	none [0]	black + yellow [2]	black [2]	black [1]	yes, black [1]	no [0]	thin [0]	black + white [0]	plain [0]	glossy [1]	no [0]	no [0]	no [0]	yes [1]
P. nigrogenis	dark red [1]	black [2]	none [0]	black + yellow [2]	black [2]	red [0]	yes, black [1]	no [0]	thin [0]	black + white [0]	plain [0]	glossy [1]	yes [1]	yes, black [1]	no [0]	yes [1]
P. baeri	black + red [2]	black [2]	none [0]	black [3]	black [2]	black + red [2]	no [0]	no [0]	thin [0]	black + white [0]	barred [1]	metallic blue [2]	no [0]	no [0]	no [0]	yes [1]
P. xinguensis	black + red [2]	black [2]	none [0]	black [3]	black [2]	black [1]	no [0]	no [0]	thin [0]	black + white [0]	barred [1]	metallic blue [2]	no [0]	no [0]	no [0]	yes [1]

Cissopis leveriana	black [4]	mixed [1]	none [0]	black [3]	black [2]	n/a	no [0]	no [0]	thick [1]	white [3]	plain [0]	glossy [1]	no [0]	no [0]	no [0]	no [0]
Neothraupis fasciata	grey [3]	gray [0]	none [0]	pearly gray [0]	gray [0]	white [3]	no [0]	light edged [1]	thick [1]	gray + white [1]	plain [0]	none [0]	yes [1]	no [0]	no [0]	yes [1]
Schistochlamys melanopis	black [4]	gray [0]	none [0]	pearly gray [0]	gray [0]	black [1]	no [0]	light edged [1]	thick [1]	gray + white [1]	plain [0]	none [0]	no [0]	no [0]	no [0]	yes [1]

Appendix B. Revised taxonomy of *Paroaria*.

For each species we list the first name recorded for the taxon being revised and any other name being synonymized in this study (Hellmayr, 1938). Names for colors and plumage patterns follow those used by previous researchers, always using the simplest color description. When a standard was needed for comparisons, color names and numbers follow Smithe (1981), in which case the name is capitalized. Numbers of specimens examined for each taxon are provided in Table 1.

GENUS *PAROARIA* BONAPARTE, 1831, GIORN. ARCAD., 52, P. 206.

***Paroaria coronata* (Miller, 1776)**

Synonymy. Loxia coronata Miller, 1776, Var. Subj., Nat. Hist., Part 1, pl. 2.

Diagnosis. Distinguished from all other species in the genus by its large crest and long pointed bib that are Geranium red, and the uniformly gray dorsum, which is only sparsely spotted with white in the upper neck area (see Fig. 3). In addition, this species has 24 unique substitutions in the cytochrome *b* gene (Supplementary table).

Distribution. Eastern Bolivia (Santa Cruz), Brazil (southwestern Mato Grosso and southern Rio Grande do Sul), western Paraguay, Uruguay, and northern Argentina (south to northern Mendoza, central La Pampa, and central Buenos Aires); introduced in Hawaii (Ridgely and Tudor, 1989).

Comments. This taxon was described on the basis of an illustration (Miller 1776), and the specimen(s) on which the illustration was based is not known. The earliest specimens cited for

the taxon are those used by Lafresnaye and D'Orbigny, Syn. Av., 1, in Mag. Zool., 7, cl. 2, p. 82, 1837, with localities of Mojos, Bolivia and Corrientes, Argentina (Hellmayr, 1938).

***Paroaria dominicana* (Linnaeus, 1758)**

Synonymy. Loxia dominicana Linnaeus, 1758, Syst. Nat., 10th ed., 1, p. 172.

Diagnosis. Distinguished from all other species of the genus by its black and gray dorsum boldly spotted white in the upper neck area, producing a mottled appearance. It can be further distinguished from *P. coronata* by the absence of a crest, the darker red (Geranium to Crimson) head and bib, and because the bill has the upper mandible completely dark gray (see Fig. 3). In addition, this species has 33 unique substitutions in the cytochrome *b* gene (Supplementary table).

Distribution. Interior northeastern Brazil (southern Maranhão, Piauí and Ceará to northern Minas Gerais) (Ridgely & Tudor 1989).

Comments. The type series (number of specimens unknown) was deposited in the Uppsala Museum (Hellmayr, 1938).

***Paroaria capitata* (Lafresnaye & D'Orbigny, 1837)**

Synonymy.Tachyphonus capitatus Lafresnaye & D'Orbigny, 1837, Syn. Av., 1, in Mag. Zool., 7, cl. 2, p. 29

Paroaria capitata fuscipes Bond & Meyer de Schauensee, 1939, Not. Naturae, 12, p. 2.

Diagnosis. As with *P. gularis* and *P. cervicalis*, this species has a Crimson red head and black bib, white breast, belly, and nuchal collar, and black dorsum, wings, and tail. It can be distinguished from these two species (and all other species in the genus) by its completely yellow bill (see Fig. 3) and dull yellow tarsi. This species has five unique substitutions in the cytochrome *b* gene (Supplementary table).

Distribution. Southwestern Brazil (in the Pantanal region), eastern Bolivia, central Paraguay, south along the Paraguay-Paraná basin to Argentina (Córdoba, western Buenos Aires, and Entre Ríos), and southern Bolivia (Tarija); introduced in Hawaii (Ridgely and Tudor, 1989).

Comments. The type is from Corrientes, Argentina, and was deposited in the MNHN (Hellmayr, 1938). Hellmayr (1938, p. 65) asserts that in addition to the above morphological characters *P. capitata* has a smaller wing and bill than *P. gularis*, much shorter tarsi, and a "very nearly complete nuchal collar." This was discounted after examining and measuring a few study skins in the AMNH collections, because a number of *P. capitata* specimens had values and collar characteristics matching those of several *P. gularis* specimens.

Paroaria capitata fuscipes was described based on a series of eight specimens from Fortín Campero, Tarija, Bolivia, and is only known from that type series (Bond and Meyer de Schauensee, 1939). According to Bond & Meyer de Schauensee (1939) these specimens had "brownish" as opposed to "brownish yellow" tarsi, and had a smaller bill measuring 10.5 to 11.5 mm compared to the 12 to 15 mm range found in the rest of the species. Tarsus color can vary in live *P. capitata* from brownish yellow and yellowish flesh to rosy flesh (ALP personal observation). Furthermore, the tarsi tend to appear brownish in study skins and usually darken as the specimen ages, making a subsequent evaluation of this character difficult. The series of Tarija specimens was examined, and the tarsi were found to be completely within the range of

colors observed in other *P. capitata* specimens of similar age and condition. Thus, this character could not be considered diagnostic for the Bolivian population. As far as the bill length, two of the eight Tarija specimens, an adult (ANSP 134294) and a juvenile bird (ANSP 134295), had bill lengths greater than or equal to 12 mm, thus overlapping with the rest of the species. Given that this character cannot be considered diagnostic either, this population is not considered here to be distinct.

***Paroaria gularis* (Linnaeus, 1766)**

Synonymy. Tanagra gularis Linnaeus 1766, Syst. Nat., 12th ed., 1, p. 376, based on Brisson, 1760, Orn., 6, Suppl., p. 67, pl. 4, Fig. 4.

Paroaria gularis gularis Hellmayr, 1907, Nov. Zool., 14, p. 10.

Diagnosis. This species can be distinguished from the other species in the genus by the presence of a narrow region of short black feathers that surround the orbital region, and extend into the lores, but are vestigial or absent in the auricular region. These black feathers do not separate the Crimson feathers of the head and those of the throat (see Fig. 3). This species has two unique substitutions in the cytochrome *b* gene (Supplementary data).

Distribution. The Amazon basin in eastern Ecuador, northwestern Peru, the Guianas, and northern and central Brasil (south to northern Goiás, northern Mato Grosso and Rondônia) (Hellmayr, 1938; Hilty, 2002; Hilty and Brown, 1986; Paynter and Storer, 1970).

Comments. The presumed type of *gularis* from the locality "Amérique," was part of the collection of M. Mauduyt (Hellmayr, 1938). This complete collection was lost due to the inadequate specimen preservation practices of the collector (C. Voisin, MNHN, personal

communication). The first detailed locality description attached to this species is "upper Esequibo River" in Guyana, by Schomburgk, 1847, Reisen Brit. Guiana, 1, p. 310, (specimens not identified), and was later restricted to "Cayenne" by Berlepsch, 1908, Nov. Zool., 15, p.122, based on a specimen collected by Sclater.

Hellmayr (1938, p. 64) stated that birds from Peru and Ecuador tend to be "slightly larger" and have the black orbital space less conspicuous, "thus approaching the southern race *P. g. cervicalis*." The extent of the black orbital space varied within populations examined for this study, but was not correlated with geographic origin (n = 114 from Brazil, 61 from Peru, and 17 from Ecuador). Further differentiation within this widespread Amazonian taxon would be compatible with the available molecular data and could be investigated with a larger sample (see Discussion).

***Paroaria cervicalis* Sclater, 1862**

Synonymy. Paroaria cervicalis Sclater, 1862, Cat. Coll. Amer. Bds., p. 108.
Paroaria gularis cervicalis Allen, 1889, Bull. Amer. Muse. N. H., 2, p. 84

Diagnosis. The species is similar to *P. gularis* and *P. capitata* in proportions and coloration, including the Crimson red head, black bib, and black dorsum, but can be distinguished from *P. capitata* by its black and yellow bill and black tarsi, and from *P. gularis* by the absence of any black feathers or black coloration around the orbital region (see Fig. 3). In addition, this species has two unique substitutions in the cytochrome *b* gene (Supplementary data).

Distribution. Eastern Bolivia (along the Madeira, Beni, Mamoré, and Iténez rivers and their tributaries) and adjacent Brazil (Mato Grosso along the Rio Guaporé) (Paynter & Storer

1970). Although no specimens are known from the upper Madeira in extreme western Rondônia, the species may meet *P. gularis* in this region. Historical records indicate that the species reached Mato Grosso city (Vila Bela) in the upper Guaporé (Pelzeln 1868).

Comments. The holotype of *P. gularis cervicalis* from Bolivia was deposited in the collection of the NHM (Sharpe, 1888). The illustration in Sharpe (1888, p. 814, pl. 16) was most likely based on this specimen and depicts a yellowish, intead of black, upper mandible. Hellmayr (1938, p. 64) examined the type and reported that it had a partly brownish yellow upper mandible and that the tarsi were horn brown instead of black. For this reason, he considered the type of *P. gularis cervicalis* to be an "intergrade" or hybrid between *P. capitata* and *P. gularis*. The photographs of the type are consistent with Hellmayr's observations (K. Cook, pers. comm.). Further study is required to determine whether this specimen is a hybrid or an aberrant individual, and thus clarify the status of the type specimen. The diagnosability and separate species rank for both *Paroaria cervicalis* and *Paroaria capitata* are nevertheless highly supported by specimens from several collections as well as the molecular data obtained and analyzed in this study.

***Paroaria nigrogenis* (Lafresnaye, 1846)**

Synonymy. Nemosia nigro-genis Lafresnaye, 1846, Rev. Zool., 9, p. 273

Paroaria gularis nigrogenis Hellmayr, 1938, Cat. Bds. Amer., 13, p. 61.

Diagnosis. This species resembles *P. gularis* and *P. cervicalis* in proportions, color of the dorsum, head, bill and tarsi, but can be distinguished from all congeners by the much larger black patch around the eye that encompasses the lores, the orbital region, and broadens as it

reaches the auricular region forming a conspicuous black mask that completely separates the red feathers of the head and the throat (see Fig. 3). In addition, it is the only species in the genus with a contrasting black spot at the base of the lower mandible (chin). It is further distinguished from *P. gularis*, *P. cervicalis*, and *P. capitata* by a Crimson red bib instead of a black one (see Fig. 4). The species has 20 unique substitutions in the cytochrome *b* gene (Supplementary data).

Distribution. Venezuela and Colombia (Llanos along the Orinoco river and its tributaries), Brazil (upper Río Negro) (Hellmayr, 1938; Paynter and Storer, 1970). Appears not to meet *P. gularis* in northwestern Guyana due to unsuitable habitat for either species along the Venezuelan and Guyanan coast (Braun et al., 2000; Hilty, 2002), but may meet that species in northern Amazonas and central Colombia (Hilty & Brown 1986; see Fig.1).

Comments. The type of *P. nigrogenis*, from "the mouth of the Orinoco" in Venezuela, is housed in the MCZ (Hellmayr, 1938). In addition to the diagnostic characters mentioned above, specimens of *P. nigrogenis* have a tendency to have the bib and crown feathers much more compact and lanceolate than those of *P. gularis* and *P. cervicalis*, and the crown feathers are also generally longer. While the bill color and pattern is the same as in *P. gularis* and *P. cervicalis*, the base of the lower mandible tends to be lighter than in those two species. This taxon should now be considered an endemic species for the Llanos avian area of endemism (Haffer, 1985).

***Paroaria baeri* Hellmayr, 1907**

Synonymy. Paroaria baeri Hellmayr, 1907, Bull. Brit. Orn. Cl., 19, p. 43, Jan.

Paroaria baeri baeri Sick, Rev. Bras. Biol., 10, p. 465, 1950.

Diagnosis. Easily distinguished from *P. coronata, P. dominicana, P. capitata, P. gularis, P. cervicalis*, and *P. nigrogenis* by its metallic bluish-black head and dorsum, with only the tips of the forehead, crown, neck, and bib feathers colored dark red. The sides of neck are bluish-black as opposed to white, and the chest feathers are broadly banded with black in the basal portion (see Fig. 3). The bill appears black in specimens but the basal portion of the lower mandible is reportedly dark reddish (Hellmayr 1938, p. 65) or bluish gray in live birds (Sick, 1950). It shares the above characters with *P. xinguensis* but can be distinguished from that species because *xinguensis* lacks any red in the throat region (see Fig. 3). The species has 11 unique substitutions in the cytochrome *b* gene (Supplementary data). Since sequences of its sister species, *P. xinguensis,* are lacking in this study, this may be an overestimate of the unique substitutions for this species, and issue that can be investigated with additional information.

Distribution. Brazil (Goiás, along the Rio Araguaia, its tributary the Rio das Mortes, and recorded occasionally from the Rio Xingú and its tributary the Rio Culuene in Mato Grosso) (Sick 1950).

Comments. The type is from the Rio Araguaya, near Leopoldina in Goiás, Brazil, and is housed in the AMNH.

***Paroaria xinguensis* Sick, 1950**

Synonymy. Paroaria baeri xinguensis Sick, Rev. Bras. Biol., 10, p. 465, 1950.

Diagnosis. This species can be distinguished from *P. baeri* by the almost complete absence of red on the neck and bib feathers, which results in a contrasting red malar streak (see Fig. 3). It can be distinguished from the remaining species in the genus by that same character, plus the fact that is has a metallic bluish-black head and dorsum with only the tips of the forehead and crown colored dark red, bluish-black sides of the neck, and chest feathers broadly banded with black in the basal portion. Black barring on the chest also appears more pronounced than in *P. baeri*, but given the small series available for examination the precise nature of the variation in this character cannot be determined.

Distribution. Known from the type locality and Jacaré, 120 km. to its south (Sick 1950).

Comments. The type pf *P. xinguensis* is from Acampamento Iauarun, Alto Xingú, Mato Grosso, and is housed at the MNRJ (Sick 1950). Although sequence data were not available for *P. xinguensis* and the skin series examined was small (n = 2), a larger series of specimens sharing the diagnostic character has been reported in the literature (n = 7; Sick 1950). Sick (1950, p. 466) suggests the taxon may be more widely distributed along the Rio Xingú, but isolated from *P. baeri* in the southern portions of their distributions by the Serra do Roncador, a chain of hills reaching 500 m that form the hydrological divide between the upper parts of the Araguaia and Xingú rivers (Sick 1950, Almeida & Hennes 1969). This taxon is elevated to species level based on the morphological differences from *P. baeri* mentioned above, and its reported allopatric distribution. More research on this species is needed to clarify its distribution, taxonomy, and conservation status.

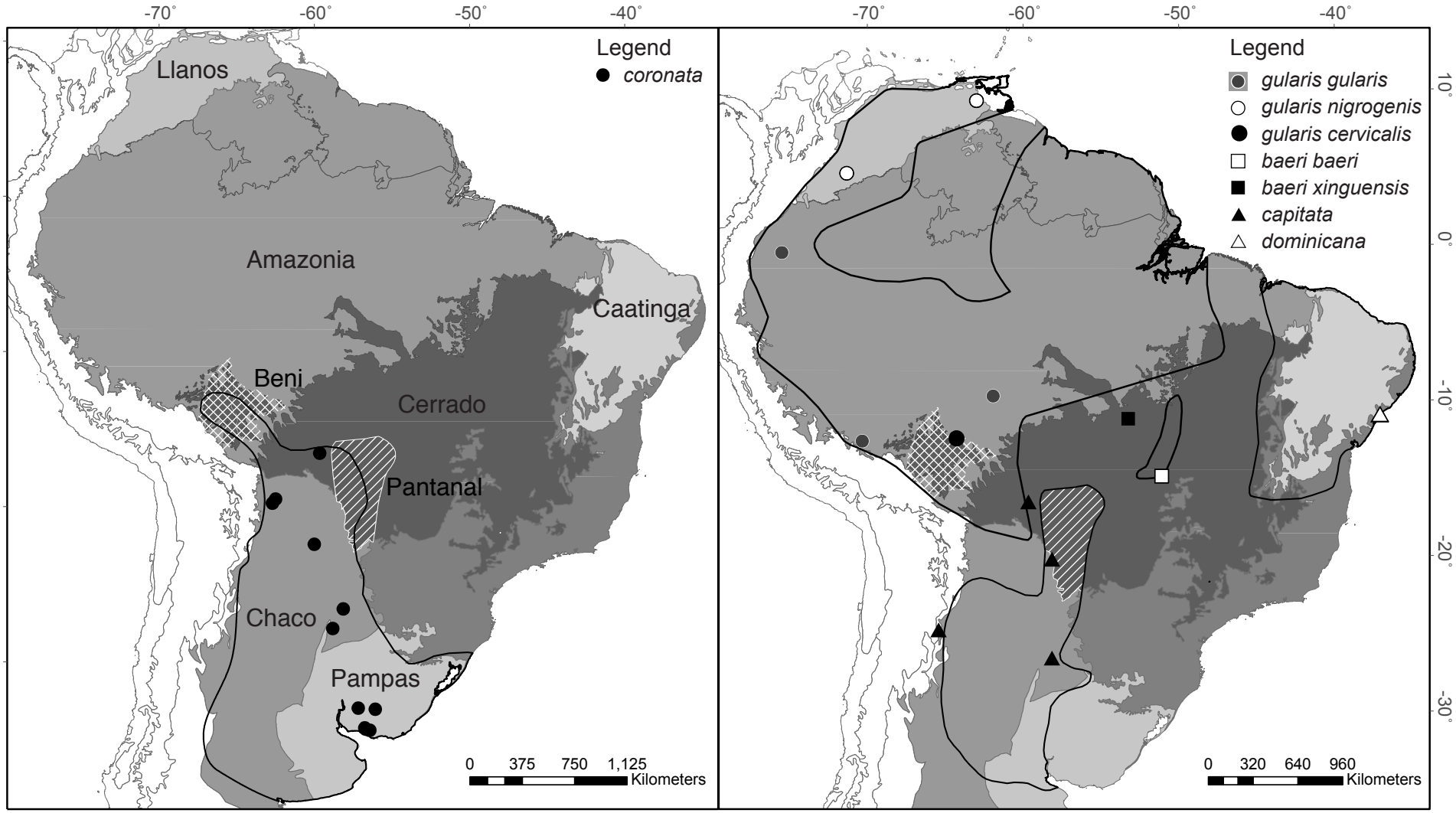
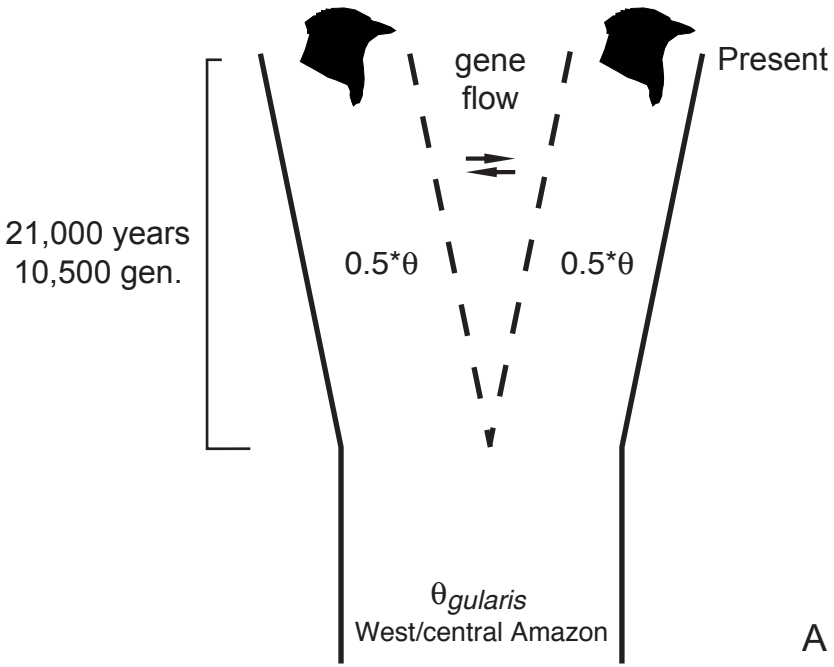


Figure 1



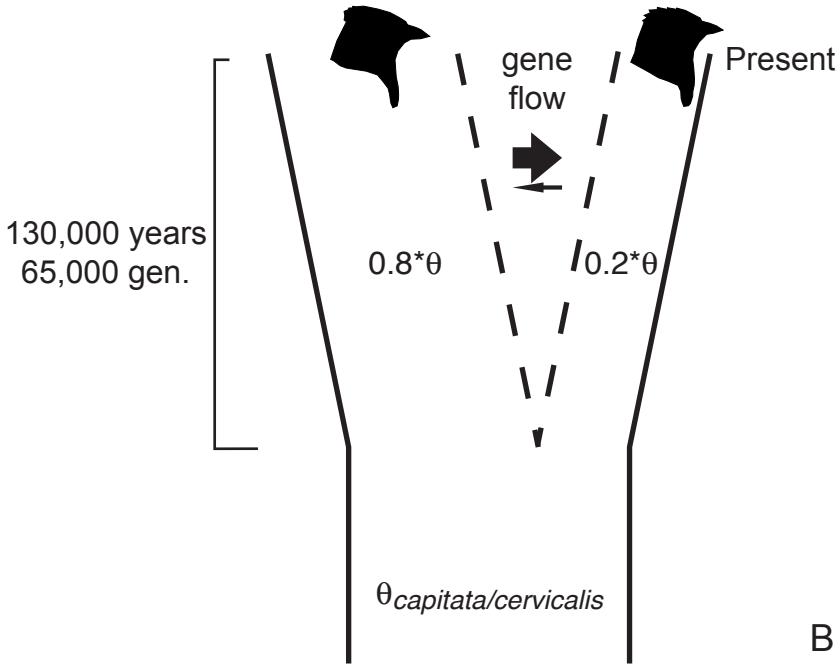


Figure 2

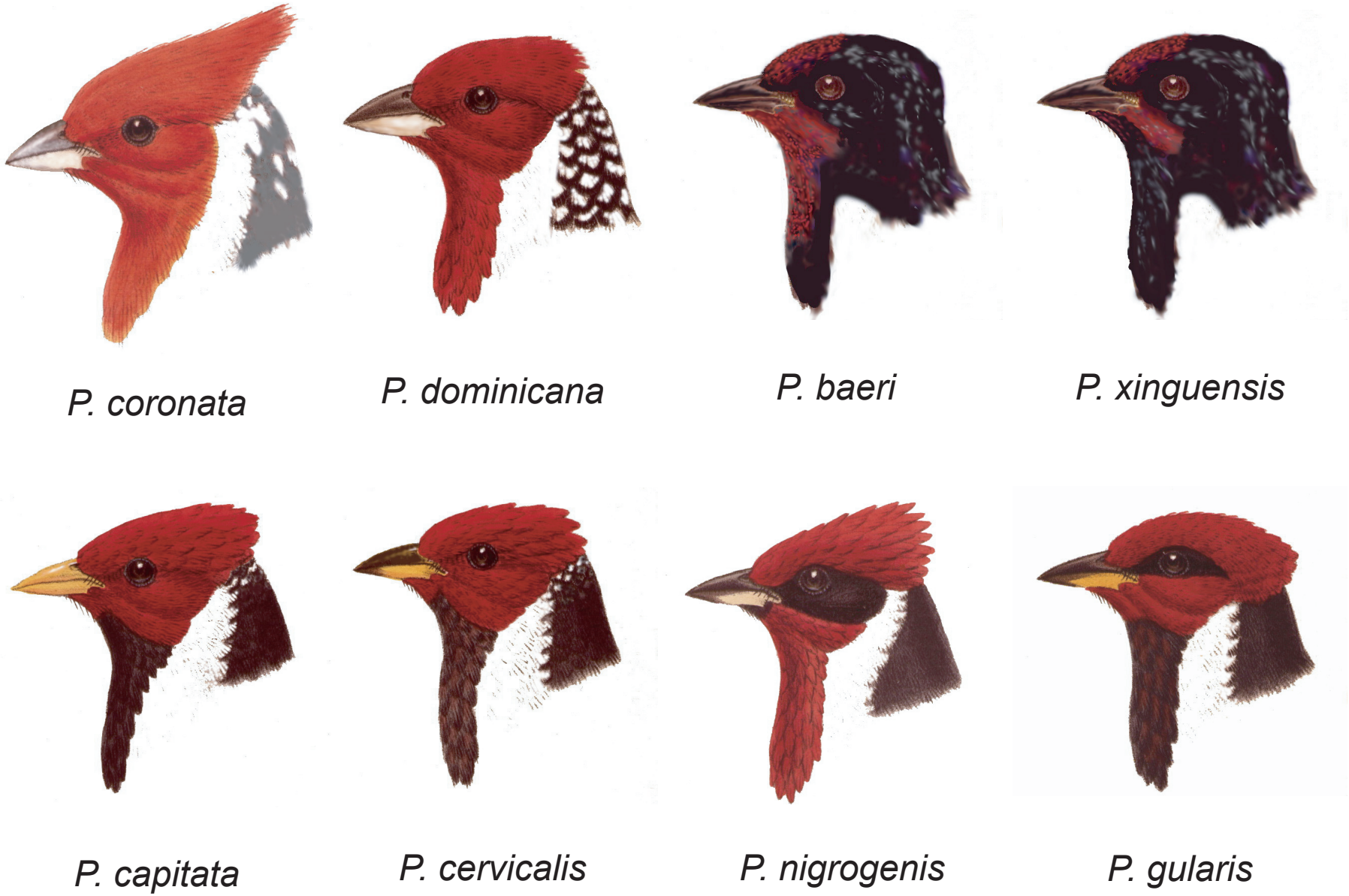


Figure 3

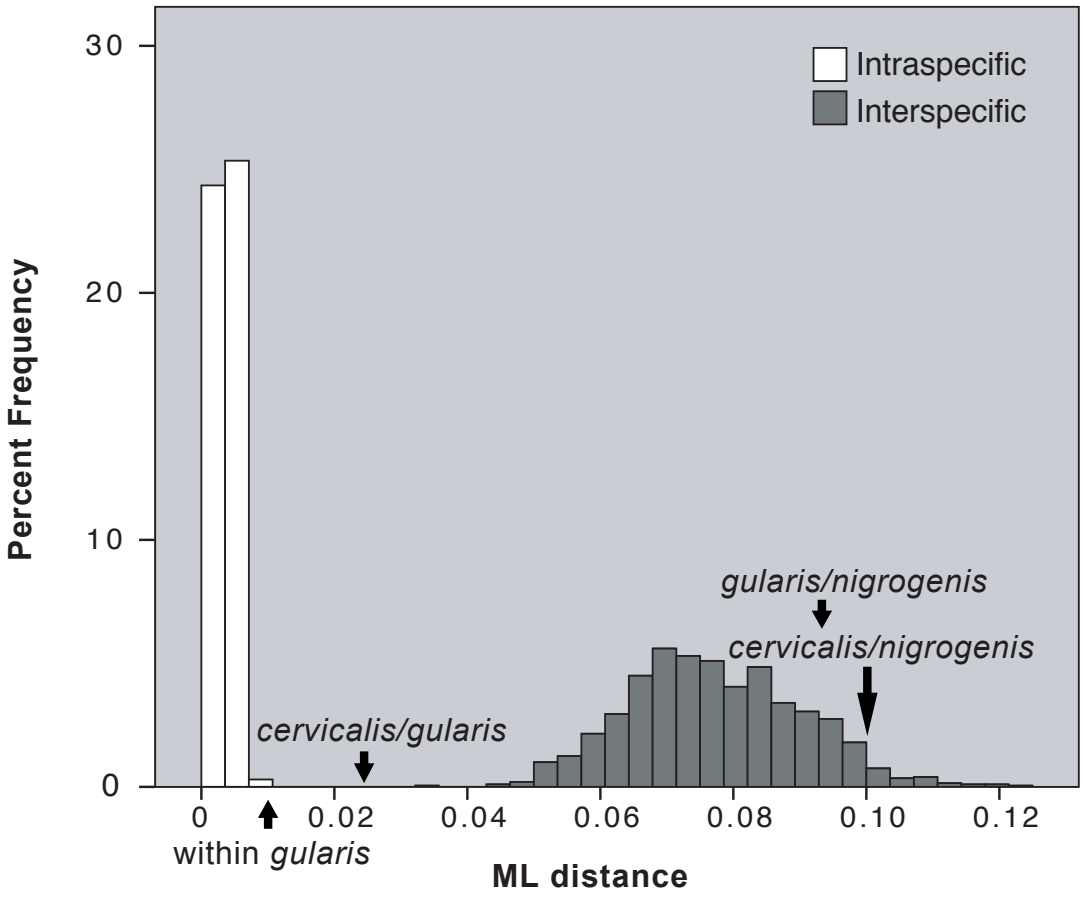


Figure 4

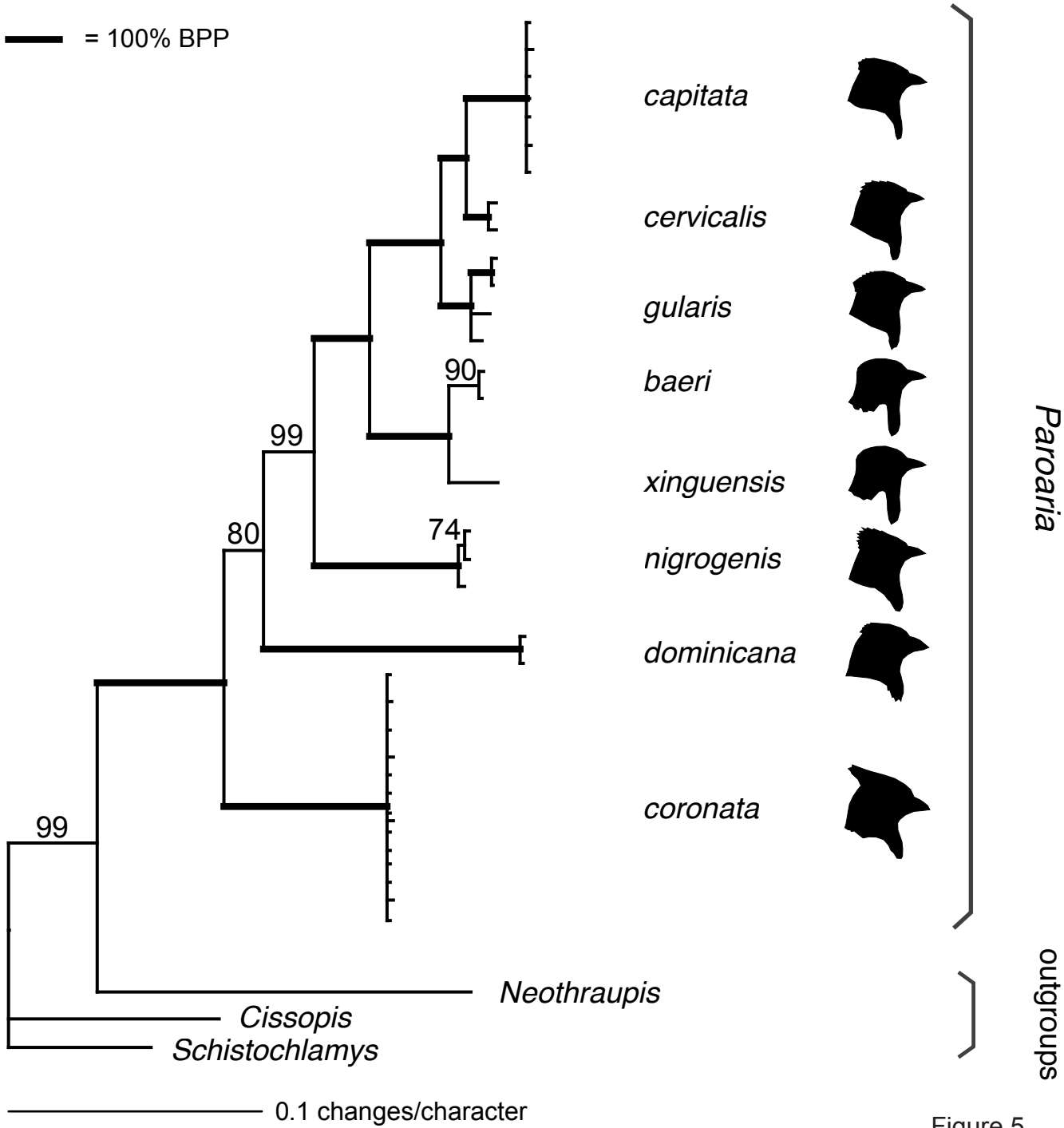


Figure 5

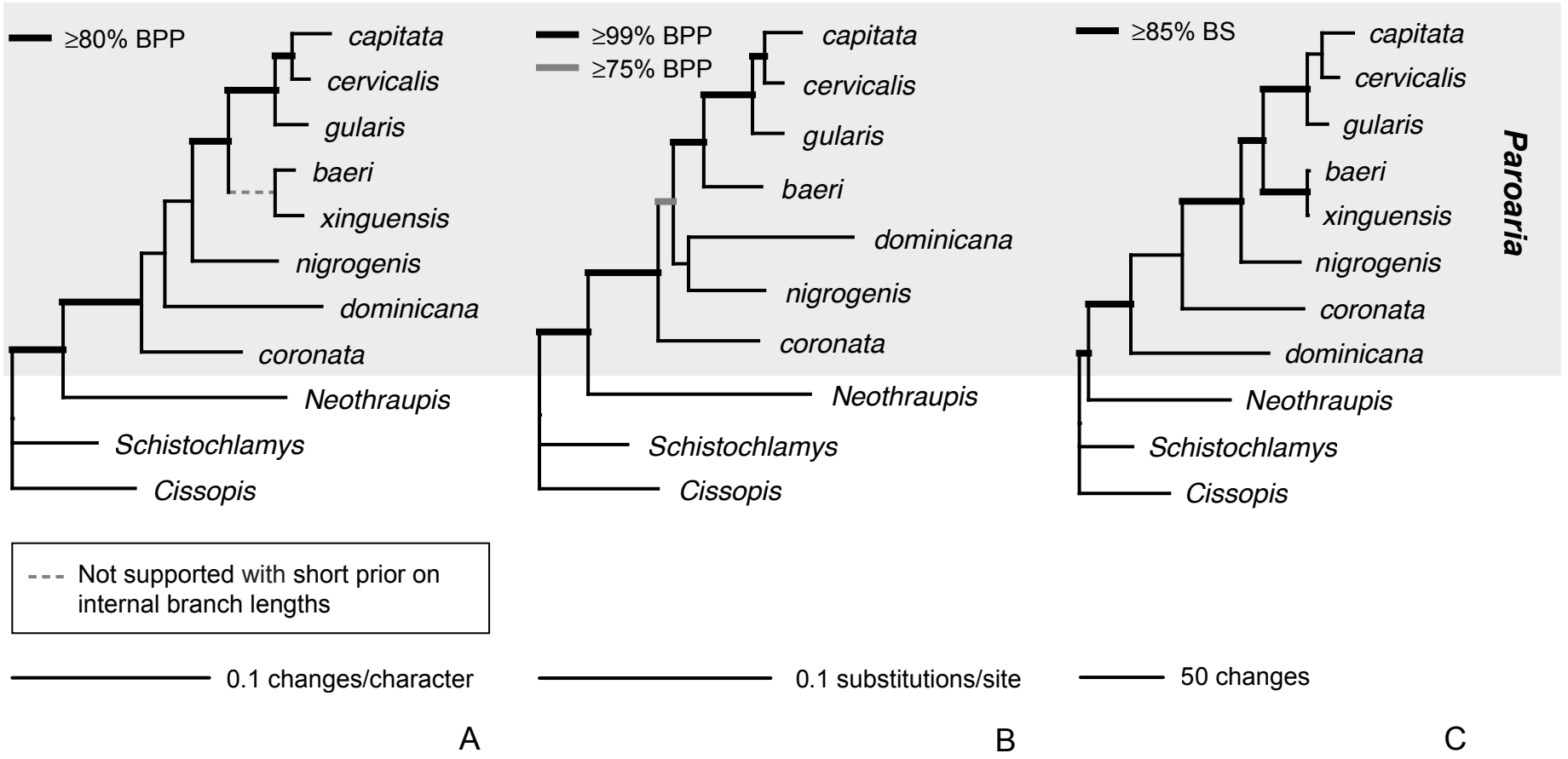


Figure 6