Published in final edited form as:

Mol Ecol. 2011 October ; 20(20): 4371–4384. doi:10.1111/j.1365-294X.2011.05270.x.

Isolation driven divergence: speciation in a widespread North American songbird (Aves: Certhiidae)

Joseph D. Manthey^{1,2}, John Klicka³, and Garth M. Spellman^{1,2}

¹Center for the Conservation of Biological Resources, Department of Biology, Black Hills State University, Spearfish, SD 57799, USA

²Integrative Genomics Program, Black Hills State University, Spearfish, SD 57799, USA

³Marjorie Barrick Museum of Natural History, University of Nevada Las Vegas, Las Vegas, NV 89154, USA

Abstract

Lineage, or true “species,” trees may differ from gene trees because of stochastic processes in molecular evolution leading to gene-tree heterogeneity. Problems with inferring species trees due to excessive incomplete lineage sorting may be exacerbated in lineages with rapid diversification or recent divergences necessitating the use of multiple loci and individuals. Many recent multilocus studies that investigate divergence times identify lineage splitting to be more recent than single locus studies, forcing the revision of biogeographic scenarios driving divergence. Here we use 21 nuclear loci from regional populations to reevaluate hypotheses identified in an mtDNA phylogeographic study of the Brown Creeper (*Certhia americana*), as well as identify processes driving divergence. Nuclear phylogeographic analyses identified hierarchical genetic structure, supporting a basal split at roughly 32°N latitude, splitting northern and southern populations, with mixed patterns of genealogical concordance and discordance between datasets within the major lineages. Coalescent-based analyses identify isolation, with little to no gene flow, as the primary driver of divergence between lineages. Recent isolation appears to have caused genetic bottlenecks in populations in the Sierra Madre Oriental and coastal mountain ranges of California, which may be targets for conservation concerns.

Keywords

Certhia; multilocus; species tree; historical demography; coalescent analysis; speciation; population bottleneck

Introduction

In the past decade, it has become routine for phylogenetic and phylogeographic studies to use multiple genetic markers as compared to traditional single-marker studies. The need for multilocus datasets stems from the haphazard nature of sequence evolution within and between lineages. Lineage, or true “species,” trees may differ from gene trees because of stochastic processes in molecular evolution leading to gene-tree heterogeneity [due to: deep

Corresponding author: Manthey, Joseph D., 1200 University Street Unit #9003, Black Hills State University, Spearfish, SD 57799-9003, United States, jdmanthey@gmail.com, Fax: 605-642-6237.

Data accessibility DNA sequences used in this study are deposited in GenBank under accession numbers JN368485-JN370948. Sequences shorter than 200 base pairs are provided in the supplementary info because GenBank currently does not accept sequences of this length. Per locus sequence submissions listed in the Supplementary Submission Data file, and summarized in Supplementary Table 1. Phylogenetic tree Newick files are included at the end of the supplementary information.

coalescence, gene duplication and horizontal gene transfer (Edwards 2009; Slowinski & Page 1999)]. Problems with inferring species trees due to excessive incomplete lineage sorting may be exacerbated in lineages with rapid diversification or recent divergences, necessitating the use of multiple loci and individuals (of each species or population) in current studies (Heled & Drummond 2010; Maddison & Knowles 2006). It has been long recognized that models for lineage reconstruction (i.e. species trees) need to account for stochastic mutation and gene coalescence (Belfiore et al. 2008; Maddison 1997; Takahata 1989). Recently, methods for reconstructing species histories that account for the independent, stochastic evolutionary histories of individual genes have become readily accessible (i.e. BEAST and BEST (Edwards et al. 2007; Heled & Drummond 2010; Liu 2008)).

Many recent, empirical studies have inferred species trees with multiple loci, identifying more robust conclusions than from single markers, in spite of incomplete lineage sorting due to recent divergences and prolific gene-tree heterogeneity (for some recent examples, see: Leache 2009; Li et al. 2010; Pasachnik et al. 2010; Themudo et al. 2009). In those multilocus studies that investigate divergence times, lineage splitting is in general found to be more recent, possibly providing more realistic estimates of lineage divergence (Jennings & Edwards 2005; Lee & Edwards 2008; McCormack et al. 2010; Rovito 2010; VW Walstrom, Pers. Comm.); and forcing the revision of biogeographic scenarios for lineage splitting.

With population sampling and many genetic markers, demographic processes that lead to patterns of lineage splitting and divergence may be inferred. Using simulation studies, Carling and Brumfield (2007) showed that increasing the number of loci sampled improved the accuracy of population demographic parameter estimates, even more so than increasing the amount of total base pairs used. Additionally, when many markers are used, pseudo-genomic patterns may be inferred (i.e. patterns of genetic diversity and deviations from neutrality) that add to the inference of demographic processes leading to patterns within and among lineages.

The Brown Creeper (Aves: Certhiidae, *Certhia americana*) is widespread throughout North America in mature, old-growth coniferous and mixed coniferous-deciduous forests. Currently considered one biological species (AOU 1983), the Brown Creeper has significant phylogeographic structure in mitochondrial DNA (mtDNA; Manthey et al. 2011) with six well-supported geographically-structured clades. The basal divergence, a north-south split, coincides with the largest differences in morphology (overall body size and coloration; Webster 1986), while morphological differences between phylogenetic groups (phylogroups) within the basal lineages are largely clinal (Webster 1986). Parapatry among mtDNA clades is prevalent in western North America (Manthey et al. 2011), where possible contact zones warrant investigation using additional genetic markers to differentiate between introgression through hybridization and sympatry between two distinct non-interbreeding phylogroups. Additionally, the use of multiple markers will help identify processes driving patterns of divergence in the Brown Creeper, as well as investigate if migration (or lack thereof) is prevalent enough to prevent the effects of genetic drift between isolated populations.

To investigate phylogeographic structure, lineage divergence and population demographics, we sequenced 21 nuclear loci (introns and anonymous loci) from regional Brown Creeper lineages as inferred from mtDNA. The goals of this study are: (1) Investigate nuclear population structure within and among regional Brown Creeper populations. (2) Infer a species tree from multiple methods using many nuclear loci. (3) Investigate population diversity, demographic patterns, and divergence to infer processes driving contemporary

patterns in Brown Creeper populations. (4) Compare nuclear genetic patterns and relationships with those of mtDNA.

Methods

Sampling and Laboratory Procedures

Tissue samples of 70 Brown Creeper individuals were obtained from ten populations (using individuals from the mtDNA study; Fig. 1; Supp. Table 2) representing the structured clades recovered in mtDNA analyses (Manthey et al. 2011). These include: Chiapas (A; 8 individuals; Central America mtDNA clade), Jalisco (B; 7 individuals; Mexico mtDNA clade), Morelos (C; 8 individuals; Mexico mtDNA clade), Nuevo Leon (D; 8 individuals; Mexico mtDNA clade), New Brunswick (E in Fig. 1; 7 individuals; eastern mtDNA clade), Mono County California (F; 8 individuals; Pacific mtDNA clade), La Plata County Colorado (G; 8 individuals; Rocky Mountain mtDNA clade), Riverside County California (H; 8 individuals; Pacific mtDNA clade), Monterrey County California (I; 6 individuals; Pacific mtDNA clade) and Coconino County Arizona (J; 2 individuals; Rocky Mountain mtDNA clade). Hereafter, population labels A – J in parentheses refer to population labels in Fig. 1. Three samples of the Eurasian Treecreeper (*C. familiaris*) were used as outgroup taxa for phylogenetic analyses.

Total genomic DNA was extracted from all tissue samples using a QIAGEN DNeasy tissue extraction kit following the manufacturer's protocol. The sequences of 21 loci (anonymous loci and introns) were obtained using PCR amplification with designed primers (anonymous loci; development in Supp. Methods) or previously characterized primers (introns; Backstrom et al. 2008). PCR amplification of all sequences was carried out in 12.5 μ L reactions and included an initial denaturation period of 10 min at 95°C, with 40 subsequent cycles of 94°C for 30 s, T_A for 45 s, and 72°C for 1 min; annealing temperatures varied for each locus from 48 - 61°C (Supp. Table 1). PCR products were purified using Exosap-IT (USB Corporation) and sequenced using 10 μ L ABI BigDye (Applied Biosystems) sequencing reactions. Sequencing reactions were purified using a magnetic bead clean-up (Agencourt Biosciences) or standard ethanol precipitation clean-up followed by sequencing on an ABI 3130 Genetic Analyzer.

Sequences were aligned and edited in Sequencher 4.8 (GeneCodes). Indelligent (Dmitriev and Rakitov 2008a,b) was used to help decipher insertions or deletions (indels) in the sequences. Following editing and alignment, sequences were phased in DnaSP (Librado and Rozas 2009) with output threshold of 0.7 using algorithms provided by PHASE (Stephens et al. 2001, Stephens & Donnelly 2003). Sequences were checked for recombination using seven recombination detection tests utilized in RDP3 (Martin et al. 2010). If recombination was detected, the most informative, non-recombinant segment was trimmed and used further in analysis.

Investigation of population structure

ML phylogenies were constructed for each locus individually. With the outgroup sequences removed, TOPALi v2.5 (Milne et al. 2009) was used to select the model of sequence evolution that best fit the sequence data for each locus using the Bayesian information criterion (BIC). Maximum likelihood analyses were performed using a PhyML analysis (Guindon & Gascuel 2003) in TOPALi v2.5 (Milne et al. 2009) with 100 bootstrap replicates.

A ML tree for each locus was used as input to calculate the genealogical sorting index (GSI; Cummings et al. 2008; www.genealogicalsortingindex.org). The GSI is a normalized statistic, and is thus able to use genealogical sorting statistics from multiple loci to infer an

ensemble statistic from all the analyzed loci. The statistic quantifies the exclusive ancestry of individuals in labeled groups from rooted gene trees. The method divides the minimum number of nodes (coalescences) to produce monophyly for a given defined group divided by the observed amount of nodes to include all of the individuals defined in the group. We defined each population sample as an individual group for the GSI analysis. Pairwise population F_{ST} values for each locus were calculated using DnaSP (Librado & Rozas 2009), and the average F_{ST} values for each population were calculated to compare to the GSI ensemble statistics.

The program STRUCTURE (Falush et al. 2003; Pritchard et al. 2000) was used to identify the number of populations without a priori input using single nucleotide polymorphism (SNP) data from 20 autosomal loci (Ca8 was not included because it was invariable). The admixture model with correlated allele frequencies between populations was used with a constant lambda. Lambda was inferred by estimating the ln likelihood of $k = 1$ and allowing lambda to converge. Values of $k = 1 - 10$ (up to total number of sampled populations) were tested with ten replicates of each run. Each run had a burn-in period of 150,000 steps followed by 1 million iterations. ΔK was calculated using the ad hoc method of Evanno et al. (2005) to identify the true k . The ΔK of Evanno et al. (2005) cannot be calculated at $K = 1$; however the log-likelihood of the data indicated that $K = 1$ could be excluded as a possibility. DISTRUCT (Rosenberg 2004) was used to create publication-quality figures from the STRUCTURE output. Because the STRUCTURE analysis found $k = 2$ (discussed in results), we ran additional runs of STRUCTURE, using the same methods as above, with split datasets for each of the initial $k = 2$ groups.

Species tree estimation

Species tree estimation was performed in a Bayesian inference (BI) framework as well as using a simple concatenation method. These analyses were performed using the following programs: Bayesian evolutionary analysis by sampling trees (*BEAST v1.6.1; Drummond et al. 2006; Drummond & Rambaut 2007; Heled & Drummond 2010), Bayesian estimation of species trees (BEST; Liu 2008), and simple concatenation in TOPALi v2.5 (Milne et al. 2009).

*BEAST infers rooted phylogenies for each of the gene trees and uses these to estimate the species tree that is most probable given the multi-individual, multi-locus sequence data (Heled & Drummond, 2010). In *BEAST, a gene tree was input for each locus with taxa groups set as the seven STRUCTURE groups (see results). We used a UPGMA starting tree for each locus with a relaxed uncorrelated lognormal molecular rate. Twenty six independent analyses were run for 100 million generations each and sampled every 10,000 generations. Logcombiner v1.6.1 was used to combine log and species tree files removing the first 50% of each independent run (runs converged after 50 million generations). TRACER v1.5 (Rambaut & Drummond 2007) was used to assess the parameter distributions output from the combined *BEAST files. *BEAST output trees were summarized with TreeAnnotator v1.6.1 (Rambaut & Drummond 2008) and visualized in FigTree v1.2.2 (Rambaut 2009) to record the species-tree cladogram and support for each branch.

BEST estimates a posterior distribution of trees for each gene using standard Bayesian phylogenetic methods and then estimates the species tree based on these separate gene-tree estimates (Edwards et al. 2007). Because the full dataset could not converge, we used five datasets of three randomly-assigned alleles for each STRUCTURE group in the BEST analysis. The BEST species tree was estimated using fifteen runs (three of each random data set) with one chain each. Each run was sampled every 10,000 generations for 20 million generations using the default parameters with uniform priors and inclusion of a single *C. familiaris* sample as an outgroup. Initially, these analyses failed to converge after the 20

million generations; therefore, we used values of 1 and 1 for the mean and standard deviation of the inverse gamma distribution prior on the effective population size (θ) as described by Wiens et al. (2010). This generates a broader prior distribution on the marginal effective population sizes and appeared to help achieve stationarity quickly. The first 50% of trees were excluded as burn-in as suggested in the BEST manual.

A simple concatenation of all loci was performed to compare to the BI species tree methods. In TOPALi v2.5 (Milne et al. 2009), we ran maximum likelihood analyses (TVM + I + G) using a PhyML method (Guindon & Gascuel 2003) with 100 bootstrap replicates.

Population Analyses

Population analyses were performed for all loci, all sampling localities (except for Coconino, AZ which had a small sample size, $N=2$), as well as all combinations of loci and localities. Genetic diversity per locus was measured by the number of variable sites, number of haplotypes (H), haplotype diversity (H_D), and nucleotide diversity (π) using ARLEQUIN v3.1 (Excoffier et al., 2005). Genetic diversity by population for each locus was measured by haplotype diversity (H_D).

ARLEQUIN was also used to calculate mismatch distributions (Slatkin & Hudson 1991) by population for each locus to test a model of population growth. The significance of each mismatch distribution was assessed using the sum of squared deviations statistic and 1000 replicates of the test. For each population, Tajima's D (Tajima 1989) and Fu's F_S (Fu 1997) neutrality tests were calculated using DnaSP v5 (Librado & Rozas 2009). Mantel Correlation Coefficients (Mantel 1967) and their significance were calculated to identify whether sequence data fit a pattern of isolation by distance. This test calculates the correlation between matrices of genetic (average pairwise Slatkin's Distance [$F_{ST}/(1 - F_{ST})$] across all loci) and geographic (km) distance.

Investigation of population demographics and divergence

The Isolation with Migration (IMa) software (Hey & Nielsen 2007) was used to generate posterior probability distributions for parameters relevant to population demographics and divergence between populations. These comparisons were carried out for each pairwise-comparison of geographically-neighboring populations using the combined mtDNA (ND2; Manthey et al. 2011) and nDNA from all loci. A mutation rate of 2.6×10^{-8} (range: 2.0×10^{-8} to 3.2×10^{-8} ; rate and range from Manthey et al. 2011; Smith & Klicka 2010) mutations/site/year was applied to the mitochondrial locus, with the nDNA allowed to scale to this rate. Following trial runs to identify the appropriate scaled model parameter priors, we ran the program for a burn-in period of 500,000 steps followed by 150 – 200 million iterations. Adequate convergence and mixing were achieved as indicated by effective sample size (> 120 for each parameter) and trend line plots for all comparisons reported here. For all population comparisons that converged, we used log-likelihood ratio tests of nested models in IMa to determine if simpler models could explain the data as well as the full model. For example, a model of historical demography with no or asymmetrical gene flow may be a better fit to the data than a model with bidirectional gene flow, thereby reducing the complexity of the demographic model. Nested models were ranked using Akaike's information criterion (AIC; Carstens et al. 2009).

Results

In total, 5063 base pairs from 21 loci were sequenced for all 70 Brown Creeper samples. For each locus, with the exception of Ca34, at least one outgroup sample was sequenced successfully. In the complete dataset, there are 243 variable sites (171 parsimony

informative) and 9 indels; per locus variability is reported in Fig. 2A and Table 1. Five loci (Ca14, Ca21, Ca34, Ca57, and Ca65) contained SNPs diagnostic between northern (pops. E, F, G, H, I, J) and southern (pops. A, B, C, D) populations. Additionally, one locus (Ca57) contained three consecutive polymorphisms diagnostic of the Nuevo Leon population (D).

Population Structure

ML phylogenies (Newick files available in supp. info.) did not show patterns of monophyly, but contained geographically structured patterns of allele frequencies along branches in many of the trees. The ensemble GSI, which directly measures structure in allele frequencies of gene trees, found significant structure in each population ($p < 0.0009$; Fig. 3); Monterrey, CA, Nuevo Leon, MEX, and Chiapas, MEX (pops. I, D, A, respectively) had the highest ensemble GSI values among all populations. All populations had high average pairwise F_{ST} values ($F_{ST} > 0.2$; Fig. 2B; Table 2); ensemble GSI scores were highly correlated with average pairwise F_{ST} values ($R^2 = 0.7753$, $p = 0.001$).

The number of populations (K) estimated from the initial STRUCTURE analysis was $k = 2$ (Fig. 1; Supp. Fig. 1), dividing northern and southern samples. Further investigation of hierarchical structure found $k = 3$ or 4 in the northern group and $k = 2$ or 3 in the southern group (Fig. 1; Supp. Fig. 1). This hierarchical pattern results from the occurrence of fixed differences between the two basal groups while within basal lineages little or no SNPs are fixed and genetic structure comes from correlated allele frequencies within populations.

In the highest level of structure ($k = 2$), all but one individual (from New Brunswick (E); $Q = 0.94$) had estimated admixture coefficients (Q) greater than 0.99, assigning each individual to either the northern or the southern group. In the second level of hierarchical structure, the southern group had a $k = 2$ or 3; this separated Nuevo Leon (D) from the rest of southern populations with $k = 2$ and additionally separated Chiapas (A) with $k = 3$. For both $k = 3$ and $k = 4$, the New Brunswick population (E) was sorted from western populations. In the west, for both $k = 3$ and $k = 4$, the majority of individuals from inland populations (pops. F, G, J) were divided from populations of the coastal and transverse ranges (pops. I and H, respectively). Additionally, in $K = 4$, the transverse range population is differentiated from the coastal range population.

Estimation of Species Trees

*BEAST analyses resulted in runs that converged on the same consensus tree topologies for well-supported branches ($P > 0.50$). Effective sample size estimates of all species tree parameters were greater than 150, with the exception of two of the gene tree posterior probabilities (loci with few haplotypes). The consensus tree (cladogram inset of Fig. 1; only branches with $P > 0.50$ are shown) supports a basal split between northern and southern populations; additionally, there is weak support for a split between eastern and western populations in the north and a split of the Nuevo Leon population from other southern populations.

BEST analyses failed to converge on the same tree topology each run, likely reaching local optima in the MCMC space; however, when all fifteen runs were combined together, the BEST analysis agreed with branches supported by greater than 0.50 posterior probability in the BEAST analysis (Fig. 1). Maximum likelihood analysis of the concatenated dataset supported the basal split found in the BI analyses (Fig. 1), as well as 100% bootstrap support for nested monophyly, but not a sister-group relationship as in BI analyses of the Nuevo Leon population (D) within the southern group (Newick file available in supp. info.). The concatenated tree did not show support for any other population groups in the dataset; these

results suggest that the concatenation only supported populations and/or groups that contained diagnostic markers in one or more loci.

Population analyses

Summaries of haplotype diversity, nucleotide diversity, neutrality indices and mismatch distributions are presented in Fig. 2C, 2D and all population by locus statistics reported in Supp. Tables 3, 4, 5, & 6. Populations from Chiapas (A), Nuevo Leon (D) and Monterrey (I) exhibited the lowest amounts of genetic diversity, contained the most fixed haplotypes, and exhibited little support for population expansion (Fig. 2C, 2D). One population [Mono, CA (F)] exhibited more deviations from the signature of an expanding population than all other populations; however, this pattern was identified because many of the deviations were due to double peaks from individuals sorting to different genetic groups (i.e. mixed STRUCTURE results in Fig. 1). Interestingly, populations with the most significant deviations in mismatch distributions (pops. A, D, F, I) had no significant deviations from neutrality for any of the loci (Fu's F_S). In general, populations with high population structure (from STRUCTURE and GSI analyses) exhibited the least genetic diversity and most deviations from the signature of an expanding population while populations with less population structure (i.e. pops. B, C, F, G) showed the highest amounts of genetic diversity.

Mantel tests (Supp. Fig. 2) identified a correlation between genetic and geographic distance when all population pairwise comparisons were included in the test ($R = 0.408$, $p = 0.005$), suggesting an overall pattern of isolation by distance between all Brown Creeper populations. When the two basal groups (north and south) were separated into their own tests, northern populations indicated a pattern of isolation by distance ($R = 0.781$, $p = 0.001$) while southern populations showed no significant correlation ($R = 0.328$, $p = 0.526$). Because western populations show genetic differentiation from the eastern population (Fig. 1), we ran a Mantel test on a subset of the northern lineage, including only western populations; this test did not identify an association between genetic and geographic distance ($R = 0.103$, $p = 0.778$).

Population demographics and divergence

IMa results are summarized in Fig. 3 and Supp. Tables 7 & 8. All IMa results reported are highest point estimates and corresponding 95% highest probability density (HPD). One pairwise comparison in the south (comparison of pops B and C) and all pairwise comparisons between western populations (pops F, G, H, I, J) failed to converge in IMa runs (estimates of ancestral N_E and divergence times). These populations likely represent a small number of the actual number of populations exchanging genes with each other in these areas; which violates the assumption that other unsampled populations are not exchanging genes with the populations compared in the IMa model (Hey & Nielsen 2007). A recent phylogeographic study of mtDNA variation within *C. americana* shows extensive overlap between western clades around and in the Great Basin suggesting several western populations are likely exchanging genes. Although, recent simulation studies have found that IMa is fairly robust to moderate violations (genetic structure and gene flow from an unsampled population) of the IMa model (Strasburg & Rieseberg 2010); the potential large violations (considerable retained ancestral variation or ongoing gene flow) between the populations listed above likely preclude model estimation in these pairwise comparisons.

The basal divergence between southern and northern populations was consistent among three comparisons (inset B of Fig. 3); the greatest overlap in divergence curves was ~1.50 million years ago (mya; 0.85 – 3.01 mya HPD from combined basal comparisons). Effective migration rates between northern and southern populations were low in all comparisons (0.01 – 2.12 range inclusive of all HPDs). In basal pairwise comparisons, all populations

show an increase in effective population size compared to a common ancestor, although the magnitude of increase was greater in southern populations (Supp. Table 7). Analyses of nested models between basal groups identified models that could explain the data as well as the full model. The full model was best between populations B and J (Supp. Table 8), but a model with unidirectional northward migration was best in other comparisons (B and H, B and G; Supp. Table 8)

Within the north, divergence between the east and west (pops. G and E) was ~0.63 mya (0.37 – 1.23 mya HPD); Supp. Table 7). Effective migration between east and west is approximately the same, with rates of 0.84 and 0.90 (0.01 – 5.59 and 0.16 – 4.28 HPDs) eastward and westward, respectively. Both east and west show effective population expansions compared to the ancestral population (Supp. Table 7), although the east has a greater magnitude of increase.

In the south, Morelos had a more recent common ancestor with Nuevo Leon than Jalisco (pops. C, D, B, respectively; inset A in Fig. 3). Divergence between Chiapas and Morelos (pops. A and C in Fig. 3) occurred ~1.30 mya (0.21 – 5.06 mya HPD; Supp. Table 7), although there is a high variance in the probability distribution. There was little effective migration between Mexican populations and Nuevo Leon (Fig. 3); however, there is unidirectional effective migration from Chiapas to Morelos (pops. A and C; Supp. Tables 7 & 8). Nuevo Leon shows a population contraction from its ancestral population size with Morelos.

Discussion

Phylogeographic structure and comparisons with mtDNA

Nuclear sequence data identified hierarchical genetic structure, with a basal split dividing northern and southern Brown Creeper populations. This split is identified strongly in STRUCTURE analysis and is supported in all species tree estimation methods. The basal split between northern and southern populations is concordant with mtDNA phylogeographic patterns (Manthey et al. 2011) and the largest differences in morphology among Brown Creeper subspecies (overall body size and darkness of coloration; Webster 1986).

In the northern lineage, significant haplotype sorting was identified in all populations (ensemble GSI; Fig. 2B); however, STRUCTURE analysis identified three or four clusters, with the strongest support splitting eastern and western populations, corroborated by weak support in the coalescent species tree methods (Fig. 1). In the west, populations of the coastal and Transverse ranges (pops. I and H, respectively) are differentiated from all others, although there is some mixing of population assignment within localities. Within the west, species tree reconstructions failed to provide support for the phylogenetic relationships among populations, and most populations [all but Monterrey (I)] show low average pairwise F_{ST} values in the west (< 0.23). Patterns of genetic structure in the north may also be an artifact of the sampling scheme, in actuality representing isolation by distance between eastern and western populations, although this pattern is not apparent between western populations (Supp. Fig. 2).

In the southern lineage, as in the northern, significant haplotype sorting was identified in all populations (ensemble GSI; Fig. 2B). In this lineage, the strongest support for population divergence is the separation of Nuevo Leon (D) from other Mexican populations (strong support in STRUCTURE, weak support in species tree analyses; Fig. 1), while Chiapas (A) is only supported for population structure when the southern STRUCTURE analysis $k = 3$.

Phylogenetic and population genetic analyses of the multilocus dataset show varying levels of concordance with phylogeographic patterns revealed by the analysis of mtDNA alone (Manthey et al. 2011). At the basal level, the mtDNA gene tree (Manthey et al. 2011), STRUCTURE results, and multilocus nuclear species tree support a north-south split between populations of Brown Creeper. Resolution and broad-scale concordance of basal phylogenetic relationships across molecular markers has been observed in several other studies (Douglas et al. 2007; Maniatsi et al. 2009; McCormack et al. 2010; Thomas and Hedin 2008; Walker et al. 2006; Weisrock et al. 2010). The southern lineage similarly has concordance of genetic structure between marker types with the exception of relatively weak Bayesian posterior probability support for some relationships in the nDNA species tree analysis.

In the northern lineage, we recovered varied patterns of genealogical concordance and discordance between mtDNA (Manthey et al. 2011) and the nDNA. STRUCTURE and species tree analyses identified strong genetic structure and weak Bayesian posterior probability support, respectively, for eastern populations to be separated from western populations. In the mtDNA gene tree, the eastern populations are sister to Rocky Mountain populations with these together sister to the pacific populations; the nDNA, however, support a grouping of all western populations sister to the eastern population, with some support for population differentiation in STRUCTURE analyses and no support in species tree analyses. In other multilocus studies, similar patterns of discordance between marker types have been identified. Lee and Edwards (2008) identified a varied number of genetic groups between the mtDNA gene tree (3 clades) and multilocus STRUCTURE analyses (2 clusters) in the Red-backed Fairy Wren (*Malurus melanocephalus*). Walstrom and Spellman (Pers. Comm., unpublished data) found the same structured groups in a multilocus STRUCTURE analyses of the White-breasted Nuthatch (*Sitta carolinensis*) as were identified in the mtDNA gene tree (4 clades; Spellman & Klicka 2007), yet uncovered different sister relationships between groups in the species tree analysis. Both this study and the above cited work, emphasize the need for multilocus studies using both mtDNA and nDNA due to varied patterns of genealogical concordance, and consequently lineage histories, due to coalescent stochasticity.

Discordance among gene trees may arise due to horizontal gene transfer, gene duplication, gene flow and deep coalescence (Maddison 1997). Deep coalescence is the likely cause of gene tree heterogeneity in this study because incongruent gene trees are more common in species with a greater ancestral effective population size (see: Supp. Table 7) and shorter internodal time spans during speciation/lineage splitting (Edwards 2009; Maddison 1997). Additionally, in western North America, gene tree heterogeneity may be due to previous or ongoing male-mediated gene flow, as the mtDNA is generally geographically structured, while the nDNA is not as structured. Even when incomplete lineage sorting is present, greater sampling efforts (both individuals and genetic loci sampled) will increase the accuracy of obtained phylogenetic relationships (Maddison & Knowles, 2006). Because our sampling was high in western North America (32 individuals and 21 genetic markers), phylogenetic relationships may: 1) be complex (recent divergence, high ancestral population size, ongoing migration, etc.) and require a greater sampling effort; or 2) be those suggested by the mtDNA (Manthey et al. 2011) or those shown in Fig. 1.

Population demographics and history between structured populations

The estimation of population demography allows the inference of the evolutionary processes driving lineage diversification. The use of many genetic markers increases the accuracy and decreases the variance associated with estimating Θ (Carling & Brumfield 2007), and consequently migration rates and divergence times, and allows for comparisons to patterns identified in the mtDNA study (Manthey et al. 2011).

All non-basal-split comparisons (i.e. non-North-South comparisons) have large ancestral population sizes (between ~161,000 and 428,000 N_E values; Supp. Table 7) with low levels of migration between populations [$2Nm < 2$; Fig. 3; genetic drift will result in substantial local differentiation if $Nm < 1$ (or $2Nm < 2$) but not if $Nm > 1$ (Slatkin 1987)]. Basal comparisons have mixed values of ancestral N_E (values between 1,270 and 106,000; Supp. Table 7) and very low rates of effective migration ($2Nm < 0.381$; Fig. 3, Supp. Table 7). Additionally, low values of N_E between basal comparisons is likely due to nearly complete lineage sorting, forcing ancestral population sizes to be smaller due to a lack of shared alleles between populations. Large ancestral population sizes (especially among non-basal comparisons) suggest a large amount of ancestral variation should still exist in contemporary populations and evidence for migration events would be preserved and observable. The lack of migration indicates genetic drift, as a consequence of isolation, is driving population divergence as was indicated by mtDNA (Manthey et al. 2011).

Divergence between basal populations occurred ~1.50 mya, with the HPD largely overlapping with Pleistocene glacial cycles. This estimate is more recent and has a much better confidence distribution than mtDNA [~4.63 mya (1.473 – 9.053 HPD); Manthey et al. 2011] results. Other multilocus studies have shown more recent divergence times than single marker studies, with more loci generally showing a greater difference (Jennings & Edwards 2005; Lee & Edwards 2008; McCormack et al. 2010; Rovito 2010; VW Walstrom, Pers. Comm.). These more recent divergence times likely provide more realistic estimates of lineage divergence because of the characteristics of species trees and gene trees. Because gene trees are embedded in species trees, they will overestimate actual lineage divergence (Edwards & Beerli 2000). Additionally, because the use of more loci increases the accuracy in estimating Θ (Carling & Brumfield 2007), and consequently divergence times, multilocus studies will generally show increased accuracy in estimates over single-gene studies.

The geographic break between northern and southern populations includes both the Sonoran and Chihuahuan deserts. During the late Pliocene and Pleistocene (from basal divergence estimates), parts of these deserts were replaced by coniferous woodland that retreated south to refugia along the Gulf of California near the mouth of the Colorado River and further into Mexico (Czaplewski 1999; Graham 1999; VanDevender et al. 1990); the deserts would repeatedly retreat and re-expand during glacial maxima and interglacial periods. The divergence times between basal populations of the Brown Creeper suggests that during the movements of deserts, lineages were separated enough for divergence. Alternatively, if there was not a historically widespread ancestor, dispersal across the desert barrier would have facilitated divergence between lineages. In this case, dispersal would have needed to occur prior to accumulation of genetic structure in the ancestral population.

Between allopatric lineages, niche conservatism maintained allopatry as populations tracked suitable habitat on either side of the Sonoran and Chihuahuan deserts during Pleistocene cycles; ecological niche modeling of these lineages identified similar patterns of niche overlap between contemporary and past (Last Glacial Maximum) periods (Manthey et al. unpublished data), suggesting that gene flow has been low between these lineages for thousands of years, reinforcing the results of gene flow from genetic data (IMa results; Fig. 3). Post-divergence, both the northern and southern lineages exhibit greater effective population sizes than the ancestral population size (Supp. Table 7); however, effective population sizes of the southern lineage are greater than in the northern lineage. This pattern may be due to greater stability of habitat in the southern lineage's Mexican geographic distribution as compared to that of the northern lineage's during periods of climate change in the late Pliocene and Pleistocene (Manthey et al. unpublished data) (excluding Pacific coast ranges and Sierra Nevada; Graham 1999; Loehle 2007; VanDevender et al. 1990). Following the Last Glacial Maximum, there is evidence that the northern lineages

experienced population and range expansions (Manthey et al. 2011; Manthey et al. unpublished data). However, the time period since the expansion has been inadequate to accumulate higher genetic diversity than the southern lineage, thereby showing smaller effective population sizes in the northern lineage.

Patterns from population bottlenecks in two isolated populations

Divergence estimates suggest that Nuevo Leon (D) diverged from other Mexican populations from the south (Morelos, pop. C). The divergence time estimate of Nuevo Leon is similar to mtDNA estimates, taking place ~ 0.17 mya. This suggests that the Nuevo Leon population was connected, and therefore experienced gene flow, with other Mexican populations prior to the last interglacial (~0.12 – 0.14 mya). Our data suggest that the Nuevo Leon population has undergone a severe and rapid population bottleneck following divergence from the other Mexican populations. The support comes from three separate analyses of the data: (1) Coalescent analyses in IMA (Supp. Table 7) suggest a population contraction from the ancestral population size with Morelos. (2) Generally low genetic diversity compared to other populations [both mtDNA and nDNA (Fig. 2C)] and a high number of fixed alleles within the population in nuclear markers (Fig. 2D). (3) Range contraction into interglacial refugia (Manthey et al. unpublished data). Low genetic diversity and range contraction, similar to the Nuevo Leon population of the Brown Creeper, have been exhibited in montane pine species in the Sierra Madre Oriental (González-Astorga et al. 2005; Ledig et al. 2000). Additionally, another bird species (Pygmy Nuthatch, *Sitta pygmaea*) which is highly associated with pine species, exhibits strong population structure in nDNA in populations from Nuevo Leon, possibly due to past population bottlenecks (Spellman unpublished data).

In western North America, coastal populations in California [Monterrey (I)] exhibit lower genetic diversity than all other populations; Monterrey exhibits a fixed haplotype in mtDNA (Manthey et al. 2011), the lowest nDNA genetic diversity among sampled populations (Fig. 2C) and the highest number of fixed nDNA haplotypes (Fig. 2D). Although Monterrey has the most fixed haplotypes of any sampled population, none are private to Monterrey. The coastal range (and associated forests) in California is fragmented along the Pacific coast and isolated from other mountain ranges eastward. These populations were likely better connected to other populations during glacial maxima of the Pleistocene, as forests became less fragmented (VanDevender et al. 1990). Additionally, ecological niche modeling of the last glacial maximum supports connectivity of coastal populations (Manthey et al. unpublished data), where the potential distributional range of western North American Brown Creeper populations shifted to coastal areas (California, Oregon and Washington) and became less fragmented than contemporary forests. These patterns suggest that the lack of genetic diversity in Monterrey (I) has resulted from either: (A) a recent population bottleneck associated with fragmentation of pine forests along the Pacific coast since the last glacial maximum or (B) Monterrey acting as a sink population (no private haplotypes).

Conclusions

Phylogeographic analyses identified a hierarchical genetic structure in the Brown Creeper. Phylogeographic structure supported a basal split between northern and southern populations, concordant with mtDNA structure (Manthey et al. 2011), while within lineages there are mixed patterns of genealogical concordance and discordance between datasets. A lack of gene flow between structured populations ($2Nm < 2$) indicates isolation between populations, where genetic drift is the main cause of divergence. Although we find no evidence for selection in the genetic data, it does not preclude the possibility that selection may be maintaining allopatry. Extreme isolation is exhibited in populations of the Sierra Madre Oriental (Nuevo Leon) and Coastal range of California (Monterrey), where

population bottlenecks have resulted in a lack of genetic diversity (in both populations) and population size contraction (Nuevo Leon). However, strong genetic structure is recovered only in Nuevo Leon, which experienced a much older population bottleneck than the Monterrey population.

Populations already experiencing bottlenecks in interglacial refugia represent populations of concern for conservation studies. Anthropogenic-driven climate change will further imperil populations already experiencing a lack of genetic variation. Isolated populations along coastal California will be less impacted due to the ocean's moderating effects on climate; however, subtropical populations will likely be in greater jeopardy. Sky island populations of the Sierra Madre Oriental would be driven up mountains in climates with rising temperatures, causing further imperilment, especially due to lack of current genetic diversity in these populations.

Given the strong genetic structure and lack of effective migration, recognizing the basal lineages as species is warranted. Additionally, these groups are divergent in morphology (overall body size and darkness of coloration) and ecological biome inhabited (temperate vs. subtropical). A further examination combining genetics, morphometrics and song is warranted to verify this proposed species split.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank Jenna Eddy, Laura Kramer, Jess Moser and Jordan Sheets for assistance with molecular work. We would also like to thank the state, provincial and federal agencies and wildlife officers for their cooperation and help obtaining permits to conduct this research. Fieldwork in Mexico would not have been possible without the help of Patricia Escalante-Pliego (Instituto de Biología, Universidad Nacional Autónoma de México) or Ricardo and Irene Canales (Universidad Autónoma de Nuevo Leon). We also wish to thank Sievert Rowher and Sharon Birks at the University of Washington, Burke Museum of Natural History for providing *C. familiaris* tissue samples. This project was funded in part by NSF DEB 0814841 (Spellman) and 0815705 (Klicka). The project described was also supported in part by Grant Number 2 P20 RR016479 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH). Its contents are solely the responsibility of the authors and do not necessarily represent the official views of NCRR or NIH.

References

- AOU. Check-list of North American Birds. American Ornithologists' Union; Washington, D.C: 1983.
- Backstrom N, Fagerberg S, Ellegren H. Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Molecular Ecology*. 2008; 17:964–980. [PubMed: 17927703]
- Belfiore NM, Liu L, Moritz C. Multilocus phylogenetics of a rapid radiation in the genus *Thomomys* (Rodentia : Geomyidae). *Systematic Biology*. 2008; 57:294–310. [PubMed: 18432550]
- Carling MD, Brumfield RT. Gene Sampling Strategies for Multi-Locus Population Estimates of Genetic Diversity (theta). *PLoS One*. 2007; 2
- Carstens BC, Stoute HN, Reid NM. An information-theoretical approach to phylogeography. *Molecular Ecology*. 2009; 18:4270–4282. [PubMed: 19765225]
- Cummings MP, Neel MC, Shaw KL. A genealogical approach to quantifying lineage divergence. *Evolution*. 2008; 62:2411–2422. [PubMed: 18564377]
- Czaplewski, NJ. Recent (Holocene). In: Mares, MA., editor. *Encyclopedia of Deserts*. University of Oklahoma Press; Norman, OK: 1999. p. 469–470.
- Dmitriev DA, Rakitov RA. Decoding of Superimposed Traces Produced by Direct Sequencing of Heterozygous Indels. *Plos Computational Biology*. 2008a; 4

- Dmitriev, DA.; Rakitov, RA. Indelligent v.1.2. 2008b. Available from <http://ctap.inhs.uiuc.edu/dmitriev/indel.asp>
- Douglas ME, Douglas MR, Schuett GW, Porras LW, Thomason BL. Genealogical concordance between mitochondrial and nuclear DNAs supports species recognition of the Panamint Rattlesnake (*Crotalus mitchellii stephensi*). *Copeia*. 2007;920–932.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A. Relaxed phylogenetics and dating with confidence. *Plos Biology*. 2006; 4:699–710.
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology*. 2007; 7
- Edwards SV. Is a new and general theory of molecular systematics emerging? *Evolution*. 2009; 63:1–19. [PubMed: 19146594]
- Edwards SV, Beerli P. Perspective: Gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution*. 2000; 54:1839–1854. [PubMed: 11209764]
- Edwards SV, Liu L, Pearl DK. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences of the United States of America*. 2007; 104:5936–5941. [PubMed: 17392434]
- Evanno G, Regnaut S, Goudet J. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*. 2005; 14:2611–2620. [PubMed: 15969739]
- Excoffier L, Laval G, Schneider S. Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*. 2005; 1:47–50.
- Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*. 2003; 164:1567–1587. [PubMed: 12930761]
- Fu YX. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*. 1997; 147:915–925. [PubMed: 9335623]
- González-Astorga J, Vovides AP, Cruz-Angón A, Octavio-Aguilar P, Iglesias C. Allozyme variation in the three extant populations of the narrowly endemic cycad *Dioon angustifolium* Miq. (Zamiaceae) from North-eastern Mexico. *Annals of Botany*. 2005; 95:999–1007. [PubMed: 15760914]
- Graham, A. Late Cretaceous and Cenozoic History of North American Vegetation. Oxford University Press; New York, NY: 1999.
- Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology*. 2003; 52:696–704. [PubMed: 14530136]
- Heled J, Drummond AJ. Bayesian Inference of Species Trees from Multilocus Data. *Molecular Biology and Evolution*. 2010; 27:570–580. [PubMed: 19906793]
- Hey J, Nielsen R. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. *Proceedings of the National Academy of Sciences USA*. 2007; 104:2785–2790.
- Jennings WB, Edwards SV. Speciation history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution*. 2005; 59:2033–2047. [PubMed: 16261740]
- Leache AD. Species Tree Discordance Traces to Phylogeographic Clade Boundaries in North American Fence Lizards (*Sceloporus*). *Systematic Biology*. 2009; 58:547–559. [PubMed: 20525608]
- Ledig FT, Bermejo-Velazquez B, Hodgskiss PD, et al. The mating system and genic diversity in Martinez spruce, an extremely rare endemic of Mexico's Sierra Madre Oriental: an example of facultative selfing and survival in interglacial refugia. *Canadian Journal of Forest Research*. 2000; 30:1156–1164.
- Lee JY, Edwards SV. Divergence across Australia's Carpentarian Barrier: statistical phylogeography of the Red-backed Fairy Wren (*Malurus melanocephalus*). *Evolution*. 2008; 62:3117–3134. [PubMed: 19087188]

- Li CH, Orti G, Zhao JL. The phylogenetic placement of sinipercid fishes (“Perciformes”) revealed by 11 nuclear loci. *Molecular Phylogenetics and Evolution*. 2010; 56:1096–1104. [PubMed: 20546903]
- Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25:1451–1452. [PubMed: 19346325]
- Liu L. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics*. 2008; 24:2542–2543. [PubMed: 18799483]
- Loehle C. Predicting Pleistocene climate from vegetation in North America. *Climate of the Past*. 2007; 3:109–118.
- Maddison WP. Gene trees in species trees. *Systematic Biology*. 1997; 46:523–536.
- Maddison WP, Knowles LL. Inferring phylogeny despite incomplete lineage sorting. *Systematic Biology*. 2006; 55:21–30. [PubMed: 16507521]
- Maniatsi S, Kappas I, Baxevanis AD, Farmaki T, Abatzopoulos TJ. Sharp Phylogeographic Breaks and Patterns of Genealogical Concordance in the Brine Shrimp *Artemia franciscana*. *International Journal of Molecular Sciences*. 2009; 10:5455–5470. [PubMed: 20054480]
- Mantel N. Detection of disease clustering and a generalized regression approach. *Cancer Research*. 1967; 27:209–&. [PubMed: 6018555]
- Manthey JD, Klicka J, Spellman GM. Cryptic diversity in a widespread North American songbird: phylogeography of the Brown Creeper (*Certhia americana*). *Molecular Phylogenetics and Evolution*. 2011; 58:102–112.
- Martin DP, Lemey P, Lott M, et al. RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics*. 2010; 26:2462–2463. [PubMed: 20798170]
- McCormack JE, Heled J, Delaney KS, Peterson AT, Knowles LL. Calibrating divergence times on species trees: implications for speciation history of *Aphelocoma* jays. *Evolution*. 2011; 65:184–202. [PubMed: 20681982]
- Milne I, Lindner D, Bayer M, et al. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics*. 2009; 25:126–127. [PubMed: 18984599]
- Pasachnik SA, Echternacht AC, Fitzpatrick BM. Gene trees, species and species trees in the *Ctenosaura palearis* clade. *Conservation Genetics*. 2010; 11:1767–1781.
- Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics*. 2000; 155:945–959. [PubMed: 10835412]
- Rambaut, A. FigTree v1.2.2. 2009. Available from <http://tree.bio.ed.ac.uk/software/figtree>
- Rambaut, A.; Drummond, AJ. Tracer v1.4. 2007. Available from <http://beast.bio.ed.ac.uk/Tracer>
- Rambaut, A.; Drummond, AJ. TreeAnnotator v1.4.8. 2008. Available from <http://beast.bio.ed.ac.uk/TreeAnnotator>
- Rosenberg NA. DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes*. 2004; 4:137–138.
- Rovito SM. Lineage divergence and speciation in the Web-toed Salamanders (Plethodontidae: Hydromantes) of the Sierra Nevada, California. *Molecular Ecology*. 2010; 19:4554–4571. [PubMed: 20854412]
- Slatkin M. Gene flow and the geographic structure of natural populations. *Science*. 1987; 15:787–792. [PubMed: 3576198]
- Slatkin M, Hudson RR. Pairwise comparisons of mitochondrial-DNA sequences in stable and exponentially growing populations. *Genetics*. 1991; 129:555–562. [PubMed: 1743491]
- Slowinski JB, Page RDM. How should species phylogenies be inferred from sequence data? *Systematic Biology*. 1999; 48:814–825. [PubMed: 12066300]
- Smith BT, Klicka J. The profound influence of the Late Pliocene Panamanian uplift on the exchange, diversification, and distribution of New World birds. *Ecography*. 2010; 33:333–342.
- Spellman GM, Klicka J. Phylogeography of the white-breasted nuthatch (*Sitta carolinensis*): diversification in North American pine and oak woodlands. *Molecular Ecology*. 2007; 16:1729–1740. [PubMed: 17402986]

- Strasburg JL, Rieseberg LH. How Robust Are “Isolation with Migration” Analyses to Violations of the IM Model? A Simulation Study. *Molecular Biology and Evolution*. 2010; 27:297–310. [PubMed: 19793831]
- Tajima F. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123:585–595. [PubMed: 2513255]
- Takahata N. Gene genealogy in 3 related populations – consistency probability between gene and population trees. *Genetics*. 1989; 122:957–966. [PubMed: 2759432]
- Themudo GE, Wielstra B, Arntzen JW. Multiple nuclear and mitochondrial genes resolve the branching order of a rapid radiation of crested newts (*Triturus*, Salamandridae). *Molecular Phylogenetics and Evolution*. 2009; 52:321–328. [PubMed: 19348957]
- Thomas SM, Hedin M. Multigenic phylogeographic divergence in the paleoendemic southern Appalachian opilionid *Fumontana deprehensor* Shear (Opiliones, Laniatores, Triaenonychidae). *Molecular Phylogenetics and Evolution*. 2008; 46:645–658. [PubMed: 18053750]
- VanDevender, TR.; Martin, PS.; Betancourt, JL. *Martin Packrat Middens: The Last 40,000 Years of Biotic Change*. University of Arizona Press; Tuscon, AZ: 1990.
- Walker ML, Chirhart SE, Moore AF, Honeycutt RL, Greenbaum IF. Genealogical concordance and the specific status of *Peromyscus sejugis*. *Journal of Heredity*. 2006; 97:340–345. [PubMed: 16785280]
- Webster, JD. *Certhia americana* Bonaparte. In: Phillips, AR., editor. *The known birds of North and Middle America*. A. R. Phillips; Denver, CO: 1986. p. 195-212.
- Weisrock DW, Rasoloarison RM, Fiorentino I, et al. Delimiting Species without Nuclear Monophyly in Madagascar’s Mouse Lemurs. *PLoS One*. 2010; 5
- Wiens JJ, Kuczynski CA, Arif S, Reeder TW. Phylogenetic relationships of phrynosomatid lizards based on nuclear and mitochondrial data, and a revised phylogeny for *Sceloporus*. *Molecular Phylogenetics and Evolution*. 2010; 54:150–161. [PubMed: 19751839]

Biography

JD Manthey is an EEB graduate student at the University of Kansas, interested in evolutionary and ecological processes driving biodiversity, primarily in birds. This paper was part of his thesis research on the evolutionary history of *Certhia americana* (M.S. in Integrative Genomics Program at Black Hills State University). J Klicka is curator of Ornithology at the Barrick Museum interested in the evolution, biogeography, molecular systematics, and natural history of New World nine-primaried oscine songbirds at all taxonomic levels. GM Spellman is an assistant professor at Black Hills State University interested in speciation and evolution of North American birds.

Figure 1.

Sampling localities and STRUCTURE hierarchical results. Samples are colored to match the highest level of structure (North = blue, South = red). Secondary structure population assignment is shown on far right. Southern $k = 2$ or 3 and northern $k = 3$ or 4. Inset shows cladogram of species tree estimate; labels on branches are *BEAST posterior probabilities (top), BEST posterior probabilities (bottom left) and concatenation bootstrap support (bottom right).

Figure 2.

All population labels correspond to labels in Fig. 1. (A) Nucleotide (dots) and haplotype (bars) diversity for each locus. Solid and dotted lines represent the means of haplotype and nucleotide diversity, respectively. Means are calculated for introns (top 5) and anonymous loci (bottom 16) separately. (B) Average F_{ST} across all loci (bars) and ensemble GSI (dots) for ten populations of Brown Creeper. All ensemble GSI statistics have a P-value < 0.0009. The ensemble GSI is correlated with average F_{ST} ($R = 0.8805$, $p = 0.001$). (C) Average haplotype (bars) and nucleotide (open circles) diversity for each population. Population by locus diversity reported in Supp. Tables 3 & 4. (D) Number of loci with fixed alleles (open circles), percent of loci deviating from neutrality (gray bars; Fu's F_S) and percent of loci deviating from the pattern of an expanding population (black bars; from mismatch distributions) for each population. Neutrality and mismatch distribution percentages were calculated from the total number of loci per population with > 1 allele (number of loci with fixed alleles subtracted from 21 total loci). Population by locus statistics are reported in Supp. Tables 5 & 6.

Figure 3.

IMA results for Brown Creeper populations. Map shows effective migration rates (effective rate at which genes come into population per generation ($2Nm$)) between populations. For example, (A) shows a high migration rate into (C), but the rate is low in the opposite direction. Inset A shows divergence times with Nuevo Leon population (D) and other adjacent Mexican populations (B & C). Inset B shows divergence times between basal groups; divergence times between Jalisco (B) and Riverside (H), Coconino (J), and LaPlata (G) are shown.

Table 1

Per locus summary statistics.^a

Locus	Model	C#	Lgth	%GC	VS	PIS	Indels	H	D
13093	TrN ^{IG}	3	322	35.2	24	19	1	51	-0.15446
00895	HKY ^{IG}	3	218	72.6	11	8	2	20	-0.69652
02108	JC	15	194	39.4	13	8	0	11	-0.26038
22528	JC	2	265	41.3	14	11	1	14	-1.01546
04550	JC	26	89	51.2	3	3	1	4	0.33486
Ca4	JC	1A	266	48.0	10	6	0	12	-1.49887
Ca8	JC	UNK	172	54.1	0	0	0	1	-
Ca9	TrNef ^d	20	274	46.5	16	11	0	28	-0.30559
Ca10	HKY ^G	4A	258	31.3	10	10	1	13	0.18037
Ca13	K80 ^l	6	266	48.7	16	11	1	25	-0.22103
Ca14	K80 ^G	3	264	42.4	20	13	1	21	-0.99507
Ca15	K80 ^G	2	227	43.2	13	10	0	18	-0.87816
Ca21	F81	1	216	30.5	9	6	1	10	-0.85282
Ca29	K80	1A	189	40.4	13	7	0	12	-1.42853
Ca34	F81	8	222	37.0	6	3	0	7	0.03868
Ca47	K81 ^G	24	286	44.1	19	13	0	24	-1.24952
Ca50	JC	12	282	53.8	5	4	0	6	-1.17835
Ca51	JC	19	290	44.8	8	3	0	9	-1.95170 [*]
Ca57	JC	1	256	40.5	11	10	0	8	1.07712
Ca58	F81	5	246	32.9	14	10	0	15	-1.27215
Ca65	JC	3	261	45.7	8	5	0	8	-0.45799
Summary: ^b			5063	43.2	243	171	9	317	

^aModel, model used in ML analysis; C#, chromosome BLAST result; Lgth, length of sequence; %GC, percent bases GC; VS, variable sites; PIS, parsimony informative sites; Indels, insertions and deletions; H, number of haplotypes; D, Tajima's D

^{*} (= 0.05 statistical significance).

^bSummary is average of %GC and sum of all other statistics.

Table 2

Average (across all loci) F_{ST} pairwise comparisons across all populations.

	A	B	C	D	E	F	G	H	I
Chiapas, MX (A)									
Jalisco, MX (B)	0.15								
Morelos, MX (C)	0.16	0.05							
Nuevo Leon, MX (D)	0.41	0.28	0.24						
New Brunswick, CA (E)	0.49	0.37	0.35	0.50					
Mono Co. CA, USA (F)	0.45	0.32	0.33	0.50	0.22				
La Plata Co. CO, USA (G)	0.48	0.35	0.35	0.51	0.18	0.06			
Riverside Co. CA, USA (H)	0.54	0.40	0.40	0.57	0.26	0.12	0.12		
Monterrey Co. CA, USA (I)	0.60	0.46	0.48	0.66	0.37	0.17	0.18	0.19	
Coconino Co. AZ, USA (J)	0.51	0.36	0.39	0.52	0.23	0.10	0.07	0.11	0.17