

THE INFLUENCE OF SAMPLING DESIGN ON SPECIES TREE INFERENCE: A NEW RELATIONSHIP FOR THE NEW WORLD CHICKADEES (AVES: *POECILE*)

Rebecca B. Harris,^{1,2,3,4} Matthew D. Carling,^{1,2,5} and Irby J. Lovette^{1,2}

¹Fuller Evolutionary Biology Program, Cornell Lab of Ornithology, Cornell University, Ithaca, New York 14850

²Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14850

³Department of Biology and Burke Museum, University of Washington, Seattle, Washington

⁴E-mail: rbharris@uw.edu

⁵Department of Zoology and Physiology, Berry Biodiversity Conservation Center, University of Wyoming, Laramie, Wyoming 82071

Received April 11, 2013

Accepted September 19, 2013

Data Archived: Dryad doi: 10.5061/dryad.c413k

In this study, we explore the long-standing issue of how many loci are needed to infer accurate phylogenetic relationships, and whether loci with particular attributes (e.g., parsimony informativeness, variability, gene tree resolution) outperform others. To do so, we use an empirical data set consisting of the seven species of chickadees (Aves: Paridae), an analytically tractable, recently diverged group, and well-studied ecologically but lacking a nuclear phylogeny. We estimate relationships using 40 nuclear loci and mitochondrial DNA using four coalescent-based species tree inference methods (BEST, *BEAST, STEM, STELLS). Collectively, our analyses contrast with previous studies and support a sister relationship between the Black-capped and Carolina Chickadee, two superficially similar species that hybridize along a long zone of contact. Gene flow is a potential source of conflict between nuclear and mitochondrial gene trees, yet we find a significant, albeit low, signal of gene flow. Our results suggest that relatively few loci with high information content may be sufficient for estimating an accurate species tree, but that substantially more loci are necessary for accurate parameter estimation. We provide an empirical reference point for researchers designing sampling protocols with the purpose of inferring phylogenies and population parameters of closely related taxa.

KEY WORDS: Coalescent theory, gene flow, hybridization, phylogenetics.

A practical dilemma in employing species tree approaches is determining which loci to use. In addition to the evolutionary processes of incomplete lineage sorting (ILS) and gene flow, factors that can play a crucial role in the accuracy of species tree estimation include the number of independent loci sampled (Edwards et al. 2007; Edwards 2009), the variability and lengths of those loci (Kuhner et al. 2000; Knowles 2009; Camargo et al. 2012), and perhaps even the information content of the gene trees themselves (e.g., proportion of resolved nodes). A number of recent

studies have explored the relationship between the number of loci sampled and species tree accuracy and found that only a modest number of loci may be necessary (Edwards et al. 2007; Camargo et al. 2012). Coalescent-based species tree estimation methods allow for the estimation of population genetic parameters, such as the population size ($\theta = 2N\mu$) and divergence time (τ , generation length in years), which can be yet another gauge of phylogenetic accuracy. Theoretical and simulation work show that accurate estimates of θ and τ require multiple independent loci,



because the addition of more loci is analogous to gaining independent replicates of the evolutionary processes (Kuhner et al. 2000; Edwards and Beerli 2000). Therefore, using fewer loci for species tree inference might have negative impacts on coalescent-based estimates of population genetic parameters (Felsenstein 2006). This leads to the question of whether there is an optimal range of loci that both minimizes the number of loci needed to estimate the correct species tree, but still provides an accurate estimate of population genetic parameters.

New methods of data acquisition coupled with next-generation sequencing are providing a wealth of data (Hudson 2008; Lerner and Fleischer 2010), but Bayesian species tree estimation is presently incapable of dealing with large numbers of loci. Researchers hence need to decide which subset of loci from the hundreds or thousands of available loci should be included in species tree analyses. Here we investigate the issue of how many loci are needed to infer accurate species tree relationships, and whether loci with particular attributes (e.g., parsimony informativeness, variability, gene tree resolution) outperform others. We use subsampling strategies to evaluate the optimal sampling approach when loci are chosen randomly or based on certain characteristics, including sequence variation, the number of nodes resolved, and total information content. First, we examine how many loci are needed to estimate the reference tree (i.e., probable “true” tree) when chosen under different sampling strategies. Second, we determine how the number and information content of loci influences population genetic parameter estimates (θ and τ). Third, we compare the relative effectiveness and consistency of four species tree inference methods.

STUDY SYSTEM

We address these questions with an empirical data set focused on the New World chickadees (Aves: Paridae) composed of 40 anonymous nuclear sequence loci and mitochondrial DNA (mtDNA) sequences. The seven congeneric species of chickadees (*Poecile atricapillus*, *Poecile rufescens*, *Poecile cinctus*, *Poecile hudsonicus*, *Poecile carolinensis*, *Poecile sclateri*, and *Poecile gambeli*) are an analytically tractable group with which to explore species tree inference owing to their modest species diversity and relatively recent divergence and large population size. The phylogenetic relationships of *Poecile* are of general interest because the clade is used widely in ecological and behavioral research (e.g., Chaplin 1974; Healy and Krebs 1996; Gould et al. 2001; Hill et al. 1980; Olson et al. 2010), making an understanding of its evolutionary history important for comparative analyses.

The most robust previous phylogenetic hypotheses for the chickadees used mtDNA to determine that *Poecile* includes two well-defined groups: (*[hudsonicus, rufescens]*, *cinctus*) and (*atricapillus, gambeli*) (Gill et al. 2005; see also Johansson et al.

2013). However, the relationships among these groups relative to each other and the two remaining species, *P. carolinensis* and *P. sclateri*, were poorly resolved (Gill et al. 2005). One surprising result was that the morphologically similar species *P. atricapillus* and *P. carolinensis* were not sister taxa. Rather, the more morphologically and behaviorally divergent species *P. atricapillus* and *P. gambeli* were sister in the mtDNA tree, a relationship consistent with previous reconstructions based on mtDNA restriction site character matrices (Gill et al. 1993).

In this study, we also test for evidence of gene flow between three *Poecile* species known to hybridize: *P. atricapillus*, *P. gambeli*, and *P. carolinensis*. *Poecile atricapillus* and *P. carolinensis* have long been known to hybridize along an extensive zone of contact in eastern North America (Robbins et al. 1986; Bronson et al. 2001; Reudink et al. 2007). *Poecile atricapillus* and *P. gambeli* occasionally hybridize in the mountains of western North America (Curry 2005). Gene flow among species is a potential source of gene tree discordance (Slatkin and Maddison 1989; Maddison 1997). The difficulty of distinguishing instances of incongruence stemming from ILS or gene flow has impeded the development of phylogenetic methods that can accommodate both the processes simultaneously (but see Kubatko 2009; Yu et al. 2012), yet failing to account for gene flow during species tree estimation effects parameter estimation (Leaché et al. 2013). Here, we test if gene flow influences species tree estimation even when species sampling is specifically designed to decrease the signal of hybridization. Sampling species from distant populations may help mitigate potentially confounding affects of gene flow. However, it is especially important to measure gene flow in this group, as it is a potential source of conflict between the nuclear and mitochondrial gene trees.

Materials and Methods

TAXON SAMPLING

Previous phylogenetic work has shown that North American *Poecile* are monophyletic and the sister clade of the Eurasian *Parus* (*Parus palustris*, *Parus montanus*, and *Parus davidi*; Gill et al. 2005). We sampled two individuals of each seven *Poecile* species and included one individual of the Eurasian species *Pa. palustris* for rooting (Table S1). Species tree reconstruction methods do not account for gene flow, and instead assume that gene tree heterogeneity is due to ILS. Failing to assign individuals to the correct species can result in overestimation of effective population sizes and incorrect species tree topologies (Leaché 2009). Therefore, to decrease the chance of sampling hybrids, individuals from species known to hybridize were sampled far from species contact zones (Fig. S1). Furthermore, we explicitly test for gene flow to ensure that it will not be an issue.

MOLECULAR SEQUENCE GENERATION

Total DNA was extracted from blood or tissue samples using the DNeasy kit (QIAGEN, Valencia, CA). We used DNA from a single *P. atricapillus* individual to construct a small-insert genomic library from which we generated anonymous nuclear sequence markers (for details, see Table S2). We designed and optimized 80 primer pairs from a corresponding set of chickadee-insert sequences of which only 40 pairs amplified across all *Poecile* species. We also included the mtDNA genes *nadh-2* (*ND2*) and *nadh-3* (*ND3*). All of these target sequences were then amplified from the remaining individuals using standard PCR conditions (for details, see Table S2). To investigate the genomic identity of these loci, we conducted a heuristic search for each marker using BLAST (Altschul et al. 1990) against the GenBank nucleotide database.

We visually aligned and edited sequences using Sequencher version 4.6 (GeneCodes, Ann Arbor, MI). Putative heterozygous sites were coded with appropriate ambiguity codes. Indels were present in 18 of our 40 anonymous nuclear loci, all of which were aligned readily by eye. To check for intralocus recombination, we used the difference in sum-of-squares (DSS) method implemented in TOPALi version 2.5 (Milne et al. 2009). A sliding window of 100-bp searched the sequence in increments of 10-bp for possible recombination breakpoints. The statistical significance of each peak was assessed using 500 bootstrapping threshold runs. Peaks above the 95% threshold significance across the length of each locus were considered evidence in support of recombination. For any locus marked by the DSS method, we discarded the shortest fragment adjacent to the hypothesized recombination breakpoint, retaining only the longest adjacent region for further analysis.

Phasing of allelic variation is important because most existing species-tree analysis methods are unable to incorporate information from ambiguous sites. The phase of heterozygous genotypes was resolved using PHASE version 2.1 (Stephens et al. 2001). The algorithm was run 10 times automatically, doubling the number of iterations in the final run. Indels were ignored.

The best-fitting DNA substitution model for each locus was selected using jModelTest version 1.1 (Posada 2008) under the Akaike information criterion (Table S3). To test for the molecular clock, we used PAUP version 4.0b10 (Swofford 2002) to calculate the likelihood of the data with the clock enforced and with an unenforced clock. We rejected the clock when the resulting likelihood ratio test showed a significant difference between these scores.

GENE TREE ANALYSES

Nuclear and mitochondrial gene trees were estimated using the programs RAxML version 7.2.8 (Stamatakis et al. 2005, 2008) and MrBayes 3.1.2 (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). The details of these analyses are de-

scribed in the Supporting Information. The nuclear gene trees were necessary for several species tree methods that use gene trees as starting points, rather than sequence data. By adding two more mtDNA loci and more individuals to the analysis, we hope to resolve the mitochondrial relationships proposed by Gill et al. (2005).

GENE FLOW

We took two approaches for detecting gene flow: one that quantifies gene flow on a fixed tree, whereas the other is a phylogenetic approach that estimates trees with reticulate evolution. To determine the extent to which migrants are being exchanged between species, we used the isolation-by-migration model implemented in IMA2 (Hey 2010b) and quantified gene flow based on our reference tree. Because IMA2 requires allelic data, we compiled data sets from phased haplotypes with 0.90 or greater posterior probability in the species known to hybridize (*P. atricapillus*, *P. carolinensis*, and *P. gambeli*; for loci see Table S3). For all analyses, we used the HKY model of nucleotide substitution. To rescale results into units of time, we used the widely used 2% divergence estimate for avian families to calculate a rate of $1.296e-05$ mutations/locus for the concatenated mtDNA genes (Lovette 2004), as well as calculating the geometric mean of the nuclear loci mutation rates. We conducted 10 independent runs from different starting seeds, each sampling 10,000 genealogies with 50,000 burn-in steps, 40 chains, and a geometric heating scheme ($ha = 0.975$, $hb = 0.95$). These runs were combined to generate 100,000 gene trees for use in the nested models of population divergence which were compared using likelihood ratio tests.

To infer a species network that accounts for both ILS and gene flow, we used the maximum-likelihood (ML) method InferNetwork_ML (Yu et al. 2012), implemented in the software package PhyloNet version 3.4 (Than et al. 2008). We used the RAxML gene trees, rooted on *Pa. palustris*, and compiled data sets from the 40 loci genotypic data set. Because RAxML ignores ambiguous sites, we also analyzed the phased haplotype data set to see if any signatures of gene flow are masked by the genotypic data set. In both instances, we ran PhyloNet 10 times from different starting seeds, searching a maximum of 50 network topologies. The top five optimal ML networks were returned.

SPECIES TREE INFERENCE WITH SUBSAMPLED LOCI

We devised a locus subsampling scheme to answer two main questions: (1) What is the minimum number of loci required to produce the same species tree topology obtained by analyzing the full 40-locus data set? and (2) How does the amount of information content effect the minimum number of loci needed to resolve this same tree? In this study, we define the species tree obtained from the 40 loci analysis as our reference topology (i.e., the probable “true” tree), an assumption supported by the fact that the reference

species topology was found by the majority of the methods when using all available data (Edwards 2009).

Minimum number of loci needed

First, we investigated the effects of marker sampling on species tree inference. By subsampling loci drawn at random, we tested how many loci were needed to estimate the reference species tree consistently. We generated data sets of randomly selected nuclear loci in sets of 5, 10, 15, 20, 25, and 30 loci using a random integer generator (Fig. S2a). We generated four replicates for each subsampling scheme to quantify the influence of locus-specific effects. The 36 clock-like loci were used to generate subsets of loci for the maximum likelihood methods (i.e., STEM, STELLS), as STEM requires clock-like data. The compositions of loci sampled for each data set are described in the Supporting Information (Table S4).

Information content

For each locus, we calculated three metrics: (1) the proportion of variable sites (total variable sites/total length); (2) the proportion of parsimony informative sites (total parsimony informative sites/total length); and (3) the number of nodes resolved in the gene tree (Table S3). The first two metrics are calculated directly from the multiple sequence alignment. To calculate the number of nodes resolved, we used the gene trees estimated by MrBayes. Posterior distribution of gene trees were summarized with the *sumt* command and only nodes with >0.50 posterior probability were counted as resolved. All three metrics were calculated for the ingroup sample only, because the inclusion of the outgroup added substantial variation. Loci were ranked according to each category (from “high” to “low”) and divided into quartiles (Fig. S2b). This binning of data resulted in four tiers of loci differing in information content, with each tier containing 10 loci. For each tier, we filled as many of the 10 slots as we could before encountering a tie. Ties in the rankings were resolved by selecting among the equally ranked loci at random (Table S5). The 36 clock-like loci have nine loci in each tier. Each tier corresponds to a data set later analyzed by species tree methods.

We sought to account for the total variance in information content among the loci by conducting a principal component analysis (PCA) in R using all three data metrics (Fig. S3). Principal component 1 (PC1) accounted for $>70\%$ of the variability, and the loci were ranked according to their PC1 values from “high” to “low” (Table S3). Starting with the three top ranked loci, we analyzed sets of loci incrementally increasing in size until the species tree inference method converged upon the reference topology. We repeated the procedure by assembling data sets starting with the 10 lowest ranked loci and increasing by 2 or 5 (Fig. S2). Our expectation is that more suboptimal loci should be required to

find the reference tree compared to analyses starting with the top-ranked loci.

SPECIES TREE ANALYSES

We used four methods of coalescent-based species tree inference as implemented in the programs BEST version 3.1.2 (Liu and Pearl 2007; Liu 2008), *BEAST version 1.7.5 (Heled and Drummond 2010), STEM version 2.0 (Kubatko et al. 2009), and STELLS version 1.6 (Wu 2012). All these methods assume that discord between the gene trees and the species tree is solely a result of incomplete lineage sorting. In addition, they assume free recombination among loci, no recombination within loci, and no gene flow.

The Bayesian approaches, BEST and *BEAST, start from the original sequence alignment and use the multispecies coalescent model to directly infer the species tree. They also calculate the posterior probability distribution for gene trees, species tree, population sizes, and divergence times. The ML approaches, STEM and STELLS, use point-estimates of gene trees as input rather than starting directly from sequence alignments, which makes them more computationally efficient than the Bayesian methods. These methods also estimate population sizes and divergence times. A fundamental difference between the ML methods is that STELLS requires nothing more than the gene tree topology as input, assuming that the extra information provided by branch lengths contributes to phylogenetic noise and decreases the accuracy of species tree estimation (Huang et al. 2010; Wu 2012). STELLS does not fix population size or specify the number of generations; rather, the user assumes population size and calculates divergence times from the standard coalescent given by the branch lengths. By contrast, STEM uses both gene tree branch lengths and topology as input, and further requires a user-supplied population size estimate, adherence to the molecular clock, rate multipliers for each gene, and species assignments (Kubatko et al. 2009).

BEST

BEST was run for 1 billion generations and sampled every 25,000 generations. Larger θ values allow for a larger effective population sizes and more incomplete lineage sorting, so our θ prior was set to 0.015 (thetapr = (3,0.03); Leaché 2009). The *genemupr* was set to uniform on 0.1–2.5. We ran 10 single-chain runs for the full data set and compared their results (Linnen and Farrell 2008). Our final consensus tree was constructed by combining the final 10% of sampled trees using the “*sumt*” command, generating a majority rule consensus tree across all 10 independent replicate runs. Convergence was assessed with AWTY, as well as using plots of likelihood values and parameter estimates (Wilgenbusch et al. 2004).

***BEAST**

Five independent runs of the full 40 nuclear loci data set were performed for 1 billion generations each, sampling every 50,000 generations. For both analyses, the lognormal on the species tree population size hyper prior was set to 0.015 with an inverse γ distribution and the population size was set to constant over time. A strict clock was applied to all gene trees that adhered to the molecular clock, whereas the remaining genes were given a lognormal uncorrelated relaxed clock. The posterior probability distributions of the five independent runs were combined, summarized, and then visualized in FigTree version 1.3.1 (Rambaut 2009). Convergence was assessed using Tracer version 1.5 (Rambaut and Drummond 2007).

STEM

STEM requires that all trees are rooted and conform to a molecular clock. RAxML trees, rooted using *Pa. palustris*, were tested for adherence to the clock in PAUP version 4.0b10. The likelihood score from the ML tree and the clock-constrained tree were compared, and those trees with a *P*-value of less than 0.05 were discarded for the STEM analyses. Rate multipliers were calculated by standardizing the average pairwise distance to the outgroup by their overall mean (Yang 2002; Kubatko et al. 2011).

θ was set to 0.015 to match the mean of the prior distribution for θ used in BEST. Because different species trees can have equal likelihood scores, we calculated the ML for each of the 15 most likely species trees. All trees with equal ML score were retained in the results. We also calculated the specific likelihood of the mitochondrial topology given the set of nuclear gene trees. All other settings were set to their default values.

STELLS

We used the trees from the RAxML runs rooted with *Pa. palustris* as our input data. Branch lengths were removed from each gene tree. STELLS defines starting tree topologies by the method of minimizing deep coalescences (Maddison 1997), which it then ranks and divides into classes. Here, STELLS began with the five most parsimonious tree classes (-d command). We searched the species tree space for the 15 trees (-N 15 command) that lead to the highest coalescent likelihood of the gene trees. We analyzed the full 40 loci data set, as well as the 36 clock-like trees to enhance comparability with STEM.

BOOTSTRAPPING STEM AND STELLS

We incorporated uncertainty into the ML species tree methods by devising a gene-tree bootstrapping approach: randomly sampling gene trees from the posterior distribution generated by *BEAST according to their probability, and then generated a consensus tree from these runs with bootstrap support indicating the proportion

of sampled trees that supported each node. We selected a single tree for each locus from the Bayesian posterior distribution and repeated this process 100 times. Each subsampled data set was analyzed by STEM and STELLS, as described earlier. To summarize the results from the 100 analyses on one species tree, the ML species trees from each replicate were grouped into a common file and the sumt command in MrBayes was used to generate a 50% majority rule consensus tree from them. The values on nodes represent the number of times (out of 100) that a particular clade was supported by STEM/STELLS. We employed this method on the full data set and in the PCA data sets.

TREE METRICS

To evaluate the difference in topologies given by different methods, we used the squared path difference tree metric to compute the topological distance between two trees (Steel and Penny 1993). We chose this metric over the Robinson–Foulds distance because it generates a broader range of values over which to compare trees. Furthermore, the Robinson–Foulds distance is not robust to small changes, as moving a single terminal branch anywhere in the tree can generate a large Robinson–Foulds distance (Steel and Penny 1993). The squared path distance from the reference tree was calculated using the treedist command in the phangorn package (Schliep 2011) in R.

POPULATION GENETIC PARAMETERS

We investigated the effect of increasing the number of loci on the estimation of τ and θ , and their variance, from data sets that include, (a) all 40 loci, (b) randomly sampled loci, (c) the four tiers of loci ranked by parsimony informativeness, and (d) the top 5 loci as determined by the PC analysis. We expect that as the number of loci increases, the parameter estimates should approach the values estimated in the full-data analysis, and that the standard deviation should decrease (Jennings and Edwards 2005; Carling and Brumfield 2007). Parameter estimates (θ and τ) and corresponding standard deviation of θ for each of the seven ingroup species, and the divergence times (τ) for species pairs, were calculated from the *BEAST posterior distribution of species trees. Only those data sets that matched the reference tree topology were used.

Results

Each marker amplified a region of 300–1000 bp (Table S2). Of the 40 nuclear loci we sequenced, 36 are unannotated in chicken and zebra finch. Two loci (locus 33 and locus 35) have a significant match to hypothetical proteins LOC100229954 (98% coverage, *e*-value of $4e-37$) and LOC100223518 (86%, $5.0e-123$), respectively. Locus 9 has multiple hits, matching with the lowest *e*-value to *Anser anser* bifunctional acyltransferase mRNA

Table 1. Nested model analysis test values and corresponding *P*-values. All 2LLR test statistics followed a mixed χ^2 distribution.

Model description	df	2LLR	<i>P</i> -value
Coalescent migrate zero for <i>P. atricapillus</i>	1	0.086	0.769
Coalescent migrate zero for <i>P. carolinensis</i>	1	0.292	0.589
Coalescent migrate zero for <i>P. gambeli</i>	1	0	1
Migration rate zero between <i>P. atricapillus</i> and <i>P. carolinensis</i>	2	255.7	2.99E–56
Migration rate zero between <i>P. carolinensis</i> and <i>P. gambeli</i>	2	0.292	0.864
Migration rate zero between <i>P. atricapillus</i> and <i>P. gambeli</i>	2	3.039	0.219
Migration rates are zero between <i>P. carolinensis</i> and <i>P. gambeli</i> , and <i>P. carolinensis</i> and (<i>P. atricapillus</i> , <i>P. gambeli</i>)	4	363.8	1.84E–77
Migration rates are all zero	8	363.8	1.02E–73
Migration rates zero between all three populations	6	403.1	6.03E–84

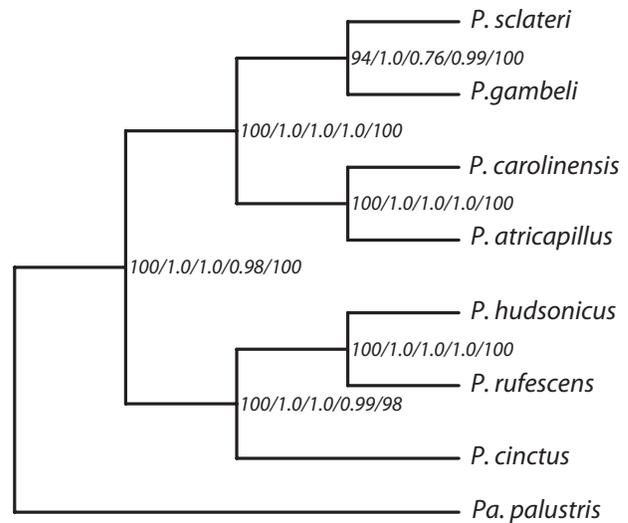
(95%, 6e–59). Locus 1 contains a fragment of the CR1 gene (83%, 1e–13).

Because of our limited intraspecific sampling, we were unable to sample the entire 40 loci nuclear data set with high confidence. The haplotypes of *P. atricapillus*, *P. gambeli*, and *P. carolinensis* were resolved with greater than 0.90 probability in 16 of the 40 nuclear loci. The haplotypes of all 8 species were resolved for 6 of 40 nuclear loci (Table S3).

All nuclear gene trees are shown in Figure S4. The Bayesian estimate of the mitochondrial phylogeny found using MrBayes produced a similar topology (Fig. S5) to that reported previously by Gill et al. (2005). The mitochondrial tree generated in this study was similar to the previously published tree in three respects: (1) the monophyly of the brown-backed chickadees was supported, (2) the sister relationship between *P. atricapillus* and *P. gambeli* was supported, and (3) the relationship of *P. carolinensis* and *P. sclateri* remained unresolved (Gill et al. 2005).

GENE FLOW

Gene flow was not evident into or out of *P. gambeli*, but a signature of unidirectional gene flow from *P. carolinensis* into *P. atricapillus* was recovered at a rate of 0.32 migrants per generation. Likelihood ratio tests comparing nested demographic models show that all models assuming no migration between *P. carolinensis* into *P. atricapillus* are rejected (Table 1).

**Figure 1.** The nuclear species tree topology produced by all methods, except STEM, using the full data set. Support values on each node correspond to posterior probabilities or bootstrap values as given by RAxML, MrBayes, BEST, *BEAST, and STELLS, respectively.

PhyloNet did not detect hybridization in either of the analyses. Both the allele and genotype data sets found a sister relationship between *P. atricapillus* and *P. carolinensis* (Fig. S6).

SPECIES TREE INFERENCE WITH THE FULL DATA SET

Both Bayesian species tree methods, BEST and *BEAST, recovered the reference topology (Fig. 1). The BEST consensus tree was constructed from 8 independent runs, each of which individually recovered that same topology. The consensus tree has high support except in the (*P. gambeli*, *P. sclateri*) node (pp = 0.76). Both runs of the *BEAST analysis found the reference tree with >0.98 support for all nodes. The only parameters that failed to reach convergence were the estimates of effective population sizes at the internal nodes. In both *BEAST runs, the effective sampling size (ESS) of all other parameters was greater than 1200 (Heled and Drummond 2010). We assume that our estimated posterior conditional distribution converged to the true joint distribution, as there was high convergence between the two runs (Kubatko et al. 2011).

We compared the bootstrapping results to those of the regular implementation of STEM and STELLS, in which single ML trees output by RAxML are input. Only STELLS found a single ML tree in all runs, which was consistently the reference tree, and had high support at all nodes in the bootstrap analysis (Fig. 1). STEM found multiple trees tied in ML score. In the regular implementation, STEM returned six ML trees with different topologies, none of which were the reference topology. STEM failed to resolve the relationship of *P. cinctus*, *P. carolinensis*, and *P. atricapillus* (CCA) returning all possible variations of this clade. This

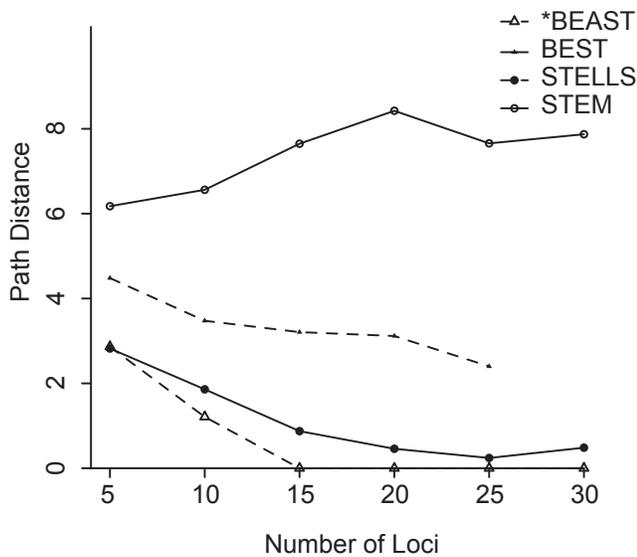


Figure 2. The performance of each species tree method with an increasing number of randomly selected loci. The average path difference from the four runs of each data set size is plotted with its standard deviation shown as error bars. Because of limited computational time, we did not analyze the 30 loci subsampled data sets in BEST.

behavior corresponds to a polytomy in the ML tree. The bootstrapped implementation of STEM, however, resolved this clade with 87% bootstrap support for the topology of CCA (Fig. S7).

SPECIES TREE INFERENCE WITH SUBSAMPLED DATA

Random subsampling of loci

With the exception of STEM, the accuracy of all methods increased as more loci were added to the analysis. Surprisingly, the addition of more loci did not help STEM resolve a single ML tree nor cause it to converge upon the reference topology (Fig. 2). We found that STEM does not merely plateau, but begins to produce inconsistent results when number of loci increases but number of individuals remains constant; it does a worse job of converging upon one tree as the number of loci increase (see also McCormack et al. 2009; Huang et al. 2010). With smaller samples of up to 20 loci, STEM identified a single ML tree across all four runs of each data set. At 30 loci, STEM returned multiple ML trees, none of which was consistent with the reference topology. This behavior may result in part from the relationship of *P. cinctus*, *P. carolinensis*, and *P. atricapillus* possibly representing a polytomy. The only clade consistently recovered across all STEM runs was the sister relationship between *P. gambeli* and *P. sclateri*. STEM was not able to find the reference species tree using the 36 molecular clock-like genes, nor with any of the randomly subsampled data sets.

STELLS found the reference topology consistently with 25 loci (Fig. 2). *BEAST converged on the reference topology with sampling of 15 or more loci. In contrast, BEST failed to consistently find the reference topology with 25 loci (Fig. 2).

Ranking loci based on information content

All methods generally performed better (i.e., found the reference topology) when analyzing loci with more information content. With the exception of STEM, all methods found the reference topology with the top tier loci ranked according to variable sites, parsimony informativeness, or number of nodes resolved by the gene trees (Fig. 3) and recovered the reference topology with many fewer loci when those loci had the highest information content. *BEAST required the fewest loci, followed by BEST, then STELLS, each needing 4, 5, or 6 loci, respectively. When using the least variable loci, these same methods needed substantially more loci (*BEAST 20, STELLS 22). BEST did not find the reference topology even with the 25 bottom-ranked loci.

POPULATION PARAMETER ESTIMATES

We provide an empirical example of how population genetic parameters estimated by *BEAST are influenced by increased sampling of loci and the information content of those loci. Because we cannot know the true value of these parameters, we used the estimate given by the full 40-loci data set as our proxy.

As predicted by simulation studies, the θ point estimates and variance increased as the number of loci decreased (Figs. 4, 5; Jennings and Edwards 2005; Carling and Brumfield 2007). The same pattern is seen as the information content of loci decreased. In general, the θ point estimate from the 10 most parsimony-informative loci performed the same or better than 20 randomly chosen loci (Fig. 5). For all species, except *P. carolinensis*, the top five informative loci gave large point estimates—resembling the five randomly selected loci. The τ point estimates and variance also decreased as the number of loci increased in the analysis, resulting in more recent divergence times (Figs. 6, 7). The estimates of τ increased as information content increased and, by and large, the top five informative loci had the largest τ estimates and error (Fig. 7).

Discussion

Both simulation and empirical studies show that species-tree estimation methods benefit substantially from an increase in sampling effort (McCormack et al. 2009; Heled and Drummond 2010; Camargo et al. 2012). In this study, we used four species tree estimation methods to investigate how these different methods perform under increasing numbers of loci. There are few studies comparing the accuracy of different methods on empirical data sets where genealogies are shaped by unknown historical and

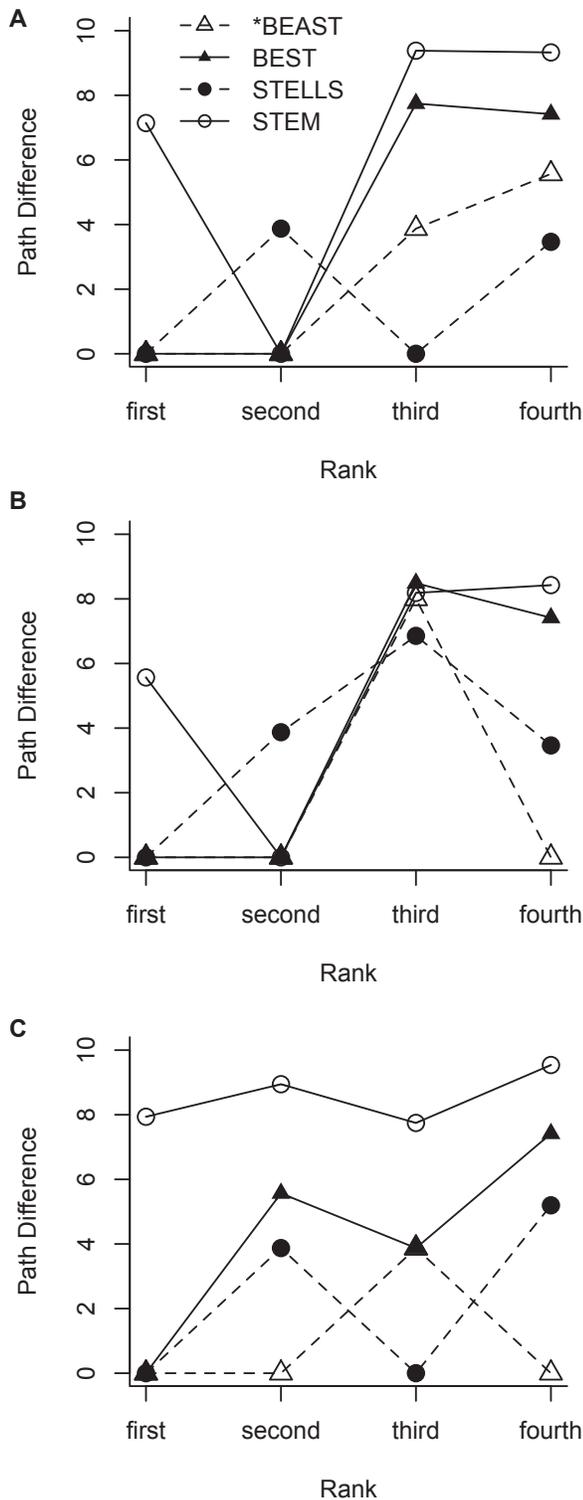


Figure 3. The performance of attributes analyzed by each species tree method. Increasing path distance indicates increasing deviation from the reference topology. Loci were ranked according to (A) number of nodes resolved, (B) percent parsimony informativeness, and (C) percent variable sites, and binned into four data sets based on ranking.

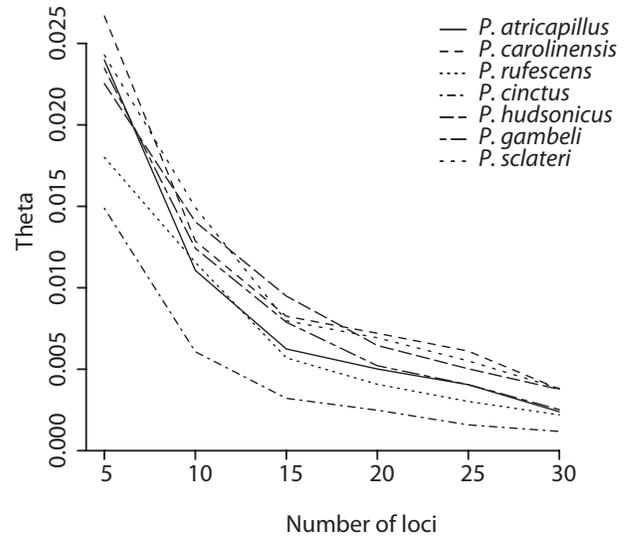


Figure 4. The variance in the population size parameter (θ) estimated by *BEAST for each species as given by the randomly sampled 5, 10, 15, 20, 25, and 30 loci. Variance is given as the average standard deviation of the posterior distribution of trees.

demographic processes (but see, Lee et al. 2012). Our results suggest that increasing the number of loci sampled is not a panacea for some of the issues inherent in these approaches.

SPECIES TREE ESTIMATION

We found that *BEAST required between 10 and 15 loci to resolve the reference species tree, similar to the finding of Camargo et al. (2012) that suggests that only eight loci are needed for accurate species tree reconstruction. However, our BEST results do not support this low number; instead, we were unable to recover our reference species tree even with 20 randomly selected loci. This difference in performance seen across these studies is probably due to the nature of empirical systems, each with a different set of speciation times, number of species, population sizes, etc. Simulation studies show that the highest accuracy in reconstructing species trees occurs when the probability of deep coalescences is minimized, and that as trees shorten and population sizes increase (thereby increasing the likelihood of deep coalescence), the accuracy of species tree inference decreases (Leaché and Rannala 2011). However, the varying performance of different methods we see in our study is likely due to algorithmic differences.

We suggest that considering the number of parsimony informative sites is the most simple and tractable metric to consider when choosing loci. Sets of loci selected on this criterion found reference tree with the top tier loci and performed as well, or better, with the second tier loci (Fig. 3). Selecting sets of loci based on their number of variable sites did not show a decrease in performance with decrease in variability, suggesting that the number of variable sites is not a strong predictor of phylogenetic

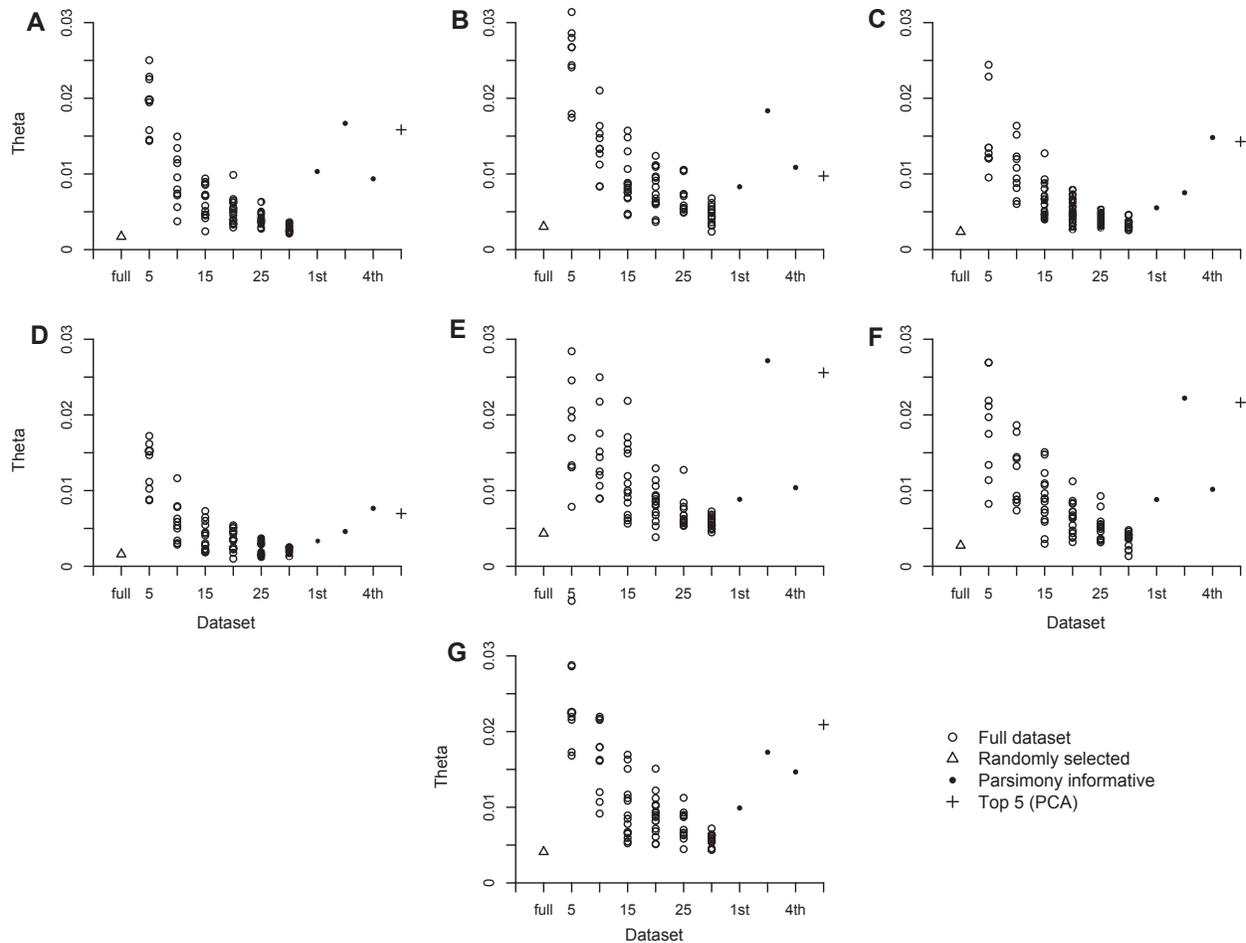


Figure 5. Population size parameter (θ) point estimates for each species as given by *BEAST. Each panel corresponds to (A) *Poecile atricapillus*, (B) *Poecile carolinensis*, (C) *Poecile rufescens*, (D) *Poecile cinctus*, (E) *Poecile hudsonicus*, (F) *Poecile gambeli*, and (G) *Poecile sclateri*. Only those data sets that found the topology of the reference tree are shown. Multiple points reflect estimates from each of the 16 replicates of the randomly selected data sets. *Parus palustris* is not included, as only one individual of this species was included in the study.

performance. Although using loci that result in greater numbers of resolved gene trees performed equally as well as parsimony informative sites, this is a more difficult metric to calculate, as it requires gene tree analyses.

We also used a PCA ordination approach to summarize the variation in all three metrics simultaneously and compiled data sets based on loci PCA score. This sampling strategy clearly demonstrated the utility of sampling loci based on total information content for estimating accurate species trees. The reduction in the number of loci needed to infer the reference tree when the most informative loci (as determined by the PC analysis) are used is striking; only four to six loci are needed in comparison to 15 or 20, using *BEAST or STELLS, respectively. Nonetheless, the ascertainment bias of using this sampling strategy should be considered, as both branch lengths and population size estimates are certainly affected when discarding loci with particular attributes (Fig. 4–7).

The discordance between the STEM topology and the topologies found using all other methods could result from several factors. First, in simulation studies, STEM has been shown to perform well when gene trees are not noisy (Wu 2012; Kubatko et al. 2009), and when branch lengths are accurate. Under these conditions, the program has low error rates and finds species trees with low inference error compared to other ML methods. Our data set, however, did not fit these criteria—our gene trees were poorly resolved and had highly variable branch lengths (Fig. S4). Furthermore, the failure of STEM to find the correct topology with the top tier loci could be due to increased phylogenetic noise, which would be less significant in the second tier loci. However, we have no clear evidence for this phenomenon, as it predicts that the least variable loci from the PCA will best resolve the correct tree.

Second, this discordance may be due, in part, to the equal weights given by STEM to both well supported, fully resolved and

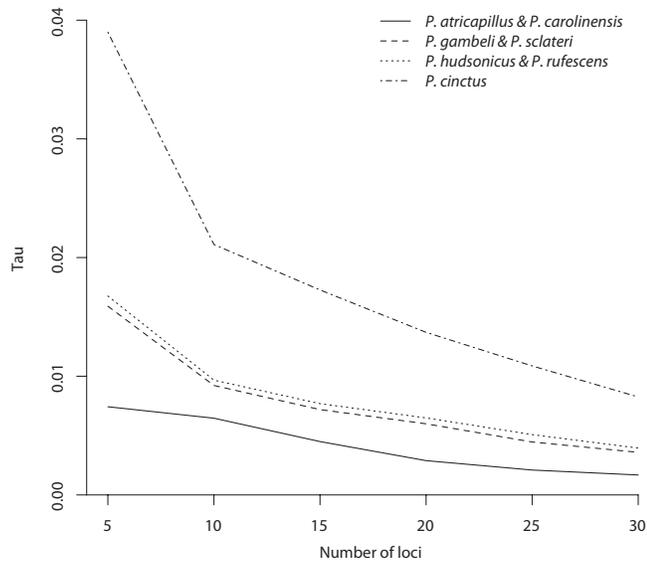


Figure 6. The variance in divergence time (τ) estimates for each species as given by the 5, 10, 15, 20, 25, and 30 randomly sampled loci data sets analyzed by *BEAST. *Parus palustris* is not included, as only one individual of this species was included in the study. Variance is given as the average standard deviation of the posterior distribution of trees.

unsupported, unresolved gene trees. We discovered that STEM was unable to determine the relationships within a clade of only three species, as it found every variation of that tree with equal ML score—thus, based on STEM a polytomy is the most accurate representation of relationships within this clade. STEM should perform best when the true gene trees are fully resolved; however, our data did not meet this criterion. We attempted to mitigate this issue by using a bootstrapping methodology (see also Leaché and Rannala 2011). In the regular implementation of STEM, the robustness of the results from ML methods is difficult to assess, as neither STEM nor STELLS provides nodal support values; rather, they compute the log likelihood of a given topology, which cannot be compared directly to support metrics derived from other methods (Salter and Pearl 2001). Poorly resolved gene trees and branch length uncertainty are not unique to our data set, and our results show that more consistent results in species tree estimation are attainable from methods that account for uncertainty in gene tree estimates, rather than using methods that rely on single-point estimates of gene tree topology (Leaché and Rannala 2011). We did not test whether adding individuals instead of loci could have ameliorated this effect. The results of the STEM analyses do not lend themselves to easy interpretation.

STELLS, by contrast, increased in accuracy as more loci were added to the analyses. The STELLS algorithm uses a coalescent summary model that allows for error in gene trees. STEM instead assumes accurately estimated gene trees and, when such trees are input, is able to infer the correct species tree with fewer genes

(Wu 2012). Our results agree with Wu's (2012) conclusion that STELLS performs better than other ML methods with noisy gene trees and increased sampling.

POPULATION GENETIC ESTIMATES

Theoretical studies suggest that increasing the number of loci directly impacts the accuracy of parameter estimates of genetic diversity (Felsenstein 2006) and population growth rates (Kuhner et al. 1998). Therefore, an optimal range of loci exists—one that decreases sampling effort, but also promises to provide accurate parameter estimates. In our study, we extend the idea to τ and θ . Our results demonstrate empirically how the precision of parameter estimation varies with the number of loci sampled. The variance in both θ and τ estimates decreased by more than a half when the number of loci analyzed increased from 5 to 30 (Figs. 4, 6). A previous study demonstrated a decrease in the error of τ estimates up to 10 loci, followed by a plateau in error (Jennings and Edwards 2005). Our finding of marked decrease in variance for τ beyond 10 loci suggests that even more loci are required to obtain accurate population parameter estimates.

In addition to the number of loci, our results show that the estimation of population genetic parameters is highly dependent on the information content of loci. θ and τ point estimates from the 10 most parsimony informative loci do not track estimates from the full data set and sometimes perform worse than the randomly sampled data sets with equal numbers of loci. Decreasing the information content generally leads to an increase in θ estimates, whereas τ point estimates increase. The top five informative loci do not outperform the 10 most parsimony informative loci, suggesting that five loci are not enough to accurately infer population genetic parameters, despite resolving the reference tree. In sum, our findings suggest that choosing loci with high information content can increase the accuracy of population genetic estimates and, in turn, require fewer loci. From our τ estimates, we see that the top five informative loci have a drastically higher τ estimate for all species groups except *P. gambeli* and *P. sclateri*. The increase in τ estimates with information content makes sense, because coalescent times are expected to follow an exponential distribution (Kingman 1982; Hudson 1990). We show that analyses conducted with the same number of loci, but with less variation, produce shorter trees. In our empirical study, five loci were insufficient for estimating τ with accuracy or precision.

NEW WORLD CHICKADEE RELATIONSHIPS

Our analyses allowed us to re-assess the phylogenetic relationships of the New World chickadees. Considering all data sets and analyses, the only clade that was consistently supported by all species tree methods is *P. rufescens* and *P. hudsonicus*. The sister relationships between *P. gambeli* and *P. sclateri* and between *P. carolinensis* and *P. atricapillus* were also consistently recovered

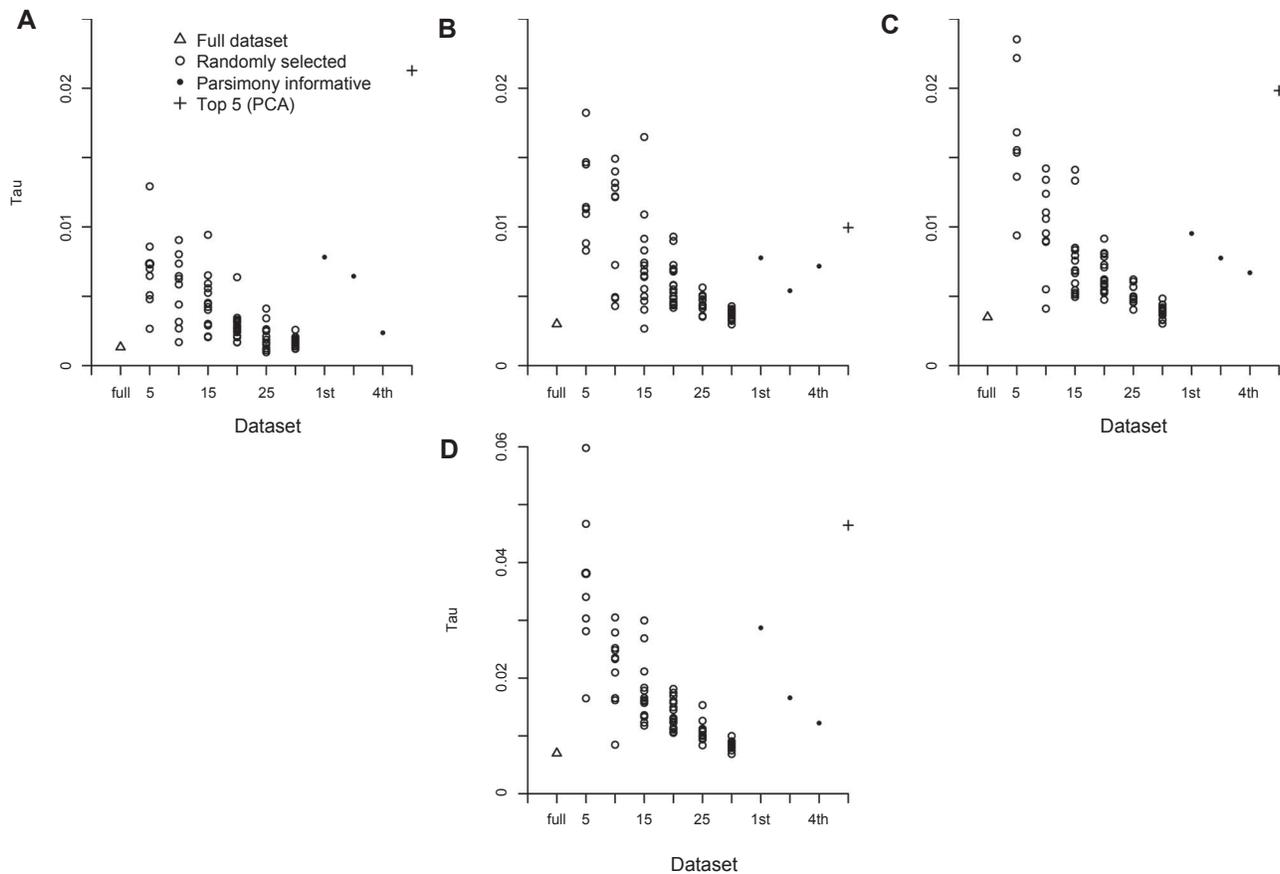


Figure 7. Point estimate of τ for each species group: (A) *Poecile atricapillus* and *Poecile carolinensis*, (B) *Poecile gambeli* and *Poecile sclateri*, (C) *Poecile rufescens* and *Poecile hudsonicus*, and (D) *Poecile cinctus*. Only those data sets that found the topology of the reference tree are shown. Multiple points reflect estimates from each of the 16 replicates of the randomly selected data sets. *Parus palustris* is not included, as only one individual of this species was included in the study.

by the nuclear data, except in the STEM analyses that produced equivocal results using all data sets (Fig. 1). More inclusive clades are also consistent across the mitochondrial and nuclear data sets (Fig. S5).

The morphological similarity between *P. carolinensis* and *P. atricapillus*, and their extensive hybridization in areas of sympatry, suggest that these two species may be sister taxa. Therefore, the highly supported sister relationship between *P. atricapillus* and *P. gambeli* inferred from the mitochondrial data, both by Gill et al. (2005) and for the substantially longer set of mitochondrial sequences we analyzed here (Fig. S5), is surprising. The discordance between mitochondrial and nuclear topologies is likely due to a past episode of mitochondrial introgression via hybridization, or to incomplete mitochondrial lineage sorting.

In closely related taxa, gene flow may limit the accurate inference of the species tree (Eckert and Carstens 2008). In this study, we detected small but significant unidirectional gene flow from *P. carolinensis* into *P. atricapillus*. Our sampling scheme is not ideal for making conclusive statements about the rate of

gene flow, as we purposefully sampled far from the hybrid zone. Thus, low gene flow may be more of an artifact of sampling than what may be occurring in nature. However, our finding is consistent with studies showing that the hybrid zone is moving northward, into areas where previously only pure *P. atricapillus* were found (Reudink et al. 2007). Although there is certainly gene flow between these species, theoretical and simulation studies suggest that less than one migrant per generation will not influence species tree topologies (Spieth 1974; Leaché et al. 2013).

ACKNOWLEDGMENTS

For comments on the manuscript, the authors thank A. Leaché, C. Linkem, S. Taylor, K. Wagner, A. Chavez, J. Grummer, M. McElroy, A. Camargo, R. Glor, M. Hart, and three anonymous reviewers. The authors thank L. Stenzler and A. Talaba for their assistance with lab work and training. The Burke Museum at the University of Washington and the Louisiana State University Museum of Natural Science provided samples. Part of this work was carried out by using the resources of the Computational Biology Service Unit from Cornell University. Support for this research was provided by the Cornell Lab of Ornithology.

LITERATURE CITED

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Bronson, C. L., T. C. Grubb, G. D. Sattler, and M. J. Braun. 2001. Mate preference: a possible causal mechanism for a moving hybrid zone. *Anim. Behav.* 65:489–500.
- Camargo, A., L. J. Avila, M. Morando, and J. W. Sites Jr. 2012. Accuracy and precision of species trees: effects of locus, individuals, and base-pair sampling on inference of species trees in lizards of the *Liolaemus darwini* group. *Syst. Biol.* 61:272–288.
- Carling, M. D., and R. T. Brumfield. 2007. Gene sampling strategies for multi-locus population estimates of genetic diversity (theta). *PLoS ONE* 2:e160.
- Chaplin, S. B. 1974. Daily energetics of Black-Capped chickadee, *Parus atricapillus*, in winter. *J. Comp. Physiol.* 89:321–330.
- Curry, R. L. 2005. Hybridization in chickadees: much to learn from familiar birds. *Auk* 122:747–758.
- Eckert, A. J., and B. C. Carstens. 2008. Does gene flow destroy phylogenetic signal? The performance of three methods for estimating species phylogenies in the presence of gene flow. *Mol. Phylogenet. Evol.* 49:832–842.
- Edwards, S. V. 2009. Is a new and general theory of molecular systematics emerging? *Evolution* 64:1–19.
- Edwards, S. V., and P. Beerli. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. *Evolution* 54:1839–1854.
- Edwards, S. V., L. Liu, and D. K. Pearl. 2007. High-resolution species trees without concatenation. *PNAS* 104:5936–5941.
- Felsenstein, J. 2006. Accuracy of coalescent likelihood estimates: do we need more sites, more sequences, or more loci? *Mol. Biol. Evol.* 23:691–700.
- Gill, F. B., A. Mostrom, and A. L. Mack. 1993. Speciation in North American chickadees. I. Patterns of mtDNA genetic divergence. *Evolution* 47:193–212.
- Gill, F. B., B. Silkas, and F. H. Sheldon. 2005. Phylogeny of titmice (Paridae). II. Species relationships based on sequences of the mitochondrial cytochrome-b gene. *Auk* 122:121–143.
- Gould, K. L., S. W. Newman, E. M. Tricomi, and T. J. DeVoogd. 2001. The distribution of substance P and neuropeptide Y in four songbird species: a comparison of food-storing and non-storing birds. *Brain Res.* 918:80–95.
- Healy, S. D., and J. R. Krebs. 1996. Food storing and the hippocampus in Paridae. *Brain Behav. Evol.* 47:195–199.
- Heled, J., and A. J. Drummond. 2010. Bayesian inference of species trees from multi-locus data. *Mol. Biol. Evol.* 27:570–580.
- Hey, J. 2010b. Isolation with migration models for more than two populations. *Mol. Biol. Evol.* 27:905–920.
- Hill, R. W., D. L. Beaver, and J. H. Veghte. 1980. Body-surface temperatures and thermoregulation in the Black-Capped chickadee (*Parus atricapillus*). *Physiol. Zool.* 53:305–321.
- Huang, H., Q. He, L. S. Kubatko, and L. Knowles. 2010. Sources of error inherent in species-tree estimation: impact of mutational and coalescent effects on accuracy and implications for choosing among different models. *Syst. Biol.* 59:573–583.
- Hudson, M. E. 2008. Sequencing breakthroughs for genomic ecology and evolutionary biology. *Mol. Ecol. Res.* 8:3–17.
- Hudson, R. R. 1990. Gene genealogies and the coalescent processes. *Oxford Surveys Evol. Biol.* 7:1–44.
- Huelsenbeck, J. P., and F. Ronquist. 2001. MrBayes: Bayesian inference and phylogenetic trees. *Bioinformatics* 17:747–755.
- Jennings, W. B., and S. V. Edwards. 2005. Speciation history of Australian grass finches (*Poephila*) inferred from thirty gene trees. *Evolution* 59:2033–2047.
- Johansson, U. S., J. Ekman, R. C. K. Bowie, P. Halvarsson, J. I. Ohlson, T. D. Price, and P. G. P. Ericson. 2013. A complete multilocus species phylogeny of the tits and chickadees (Aves: Paridae). *Mol. Phylogenet. Evol.*
- Kingman, J. F. C. 1982. The coalescent. *Stochastic processes and their applications* 13:235–248.
- Knowles, L. L. 2009. Estimating species trees: methods of phylogenetic analysis when there is incongruence across genes. *Syst. Biol.* 58:463–467.
- Kubatko, L. S. 2009. Identifying hybridization events in the presence of coalescence via model selection. *Syst. Biol.* 58:478–488.
- Kubatko, L. S., B. C. Carstens, and L. L. Knowles. 2009. STEM: species tree estimation using maximum-likelihood for gene trees under coalescence. *Bioinformatics* 25:971–973.
- Kubatko, L. S., H. L. Gibbs, and E. W. Bloomquist. 2011. Inferring species-level phylogenies and taxonomic distinctiveness using multilocus data in *Sistrurus rattlesnakes*. *Syst. Biol.* 60:393–409.
- Kuhner, M. K., P. Beerli, J. Yamato, and J. Felsenstein. 2000. Usefulness of single nucleotide polymorphism data for estimating population parameters. *Genetics* 156:439–447.
- Leaché, A. D. 2009. Species tree discordance traces to phylogenetic clade boundaries in North American fence lizards (*Sceloporus*). *Syst. Biol.* 58:547–559.
- Leaché, A. D., and B. Rannala. 2011. The accuracy of species tree estimation under simulation: a comparison of methods. *Syst. Biol.* 60:126–137.
- Leaché, A. D., R. B. Harris, B. Rannala, and Z. Yang. 2013. The influence of gene flow on Bayesian species tree estimation: a simulation study. *Syst. Biol.* *In Press*.
- Lee, J. Y., L. Joseph, and S. V. Edwards. 2012. A species tree for the Australo-Papuan fairy-wrens and allies (Aves: Maluridae). *Syst. Biol.* 61:253–271.
- Lerner, H. R. L., and R. C. Fleischer. 2010. Prospects for the use of next-generation methods in ornithology. *The Auk* 127:4–15.
- Linnen, C. R., and B. D. Farrell. 2008. Comparison of methods for species-tree inference in the sawfly genus *Neodiprion* (Hymenoptera: Diprionidae). *Syst. Biol.* 57:876–891.
- Liu, L. 2008. BEST: Bayesian estimation of species trees under the coalescent model. *Bioinformatics* 24:2542–2543.
- Liu, L., and D. K. Pearl. 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. *Syst. Biol.* 56:504–514.
- Lovette, I. 2004. Mitochondrial dating and mixed support for the “2% rule” in birds. *The Auk* 121:1–6.
- Maddison, W. P. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–536.
- McCormack, J. E., H. Huang, and L. L. Knowles. 2009. Maximum likelihood estimates of species trees: how accuracy of phylogenetic inference depends upon the divergence history and sampling design. *Syst. Biol.* 58:1–8.
- Milne, I., D. Linder, M. Bayer, D. Husmeiner, G. McGuire, D. F. Marshall, and F. Wright. 2009. TOPALi v2: a rich graphical interface for evolutionary analyses of multiple alignments on HPC clusters and multi-core desktops. *Bioinformatics* 25:26–127.
- Olson, J. R., S. J. Cooper, D. L. Swanson, M. J. Braun, and J. B. Williams. 2010. The relationship of metabolic performance and distribution in Black-Capped and Carolina chickadees. *Physiol. Biochem. Zool.* 83:263–275.
- Posada, D. 2008. jModelTest: phylogenetic model averaging. *Mol. Biol. Evol.* 25:1253–1256.
- Rambaut, A. 2009. FigTree, version 1.3.1. Published by author.
- Rambaut, A., and A. J. Drummond. 2007. Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>. Accessed September 15, 2011.

- Reudink, M. W., S. G. Mech, S. P. Mullen, and R. L. Curry. 2007. Structure and dynamics of the hybrid zone between black-capped chickadee (*Poecile atricapillus*) and Carolina chickadee (*P. carolinensis*) in southeastern Pennsylvania. *Auk* 124:463–478.
- Robbins, M. B., M. J. Braun, and E. A. Tobey. 1986. Morphological and vocal variation across a contact zone between the chickadees *Parus atricapillus* and *P. carolinensis*. *Auk* 103:655–666.
- Ronquist, F., and J. P. Huelsenbeck. 2003. MRBAYES 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Salter, L., and D. Pearl. 2001. A stochastic search strategy for estimation of maximum likelihood phylogenetic trees. *Syst. Biol.* 50:7–17.
- Schliep, K. P. 2011. phangorn: phylogenetic analysis in R. *Bioinformatics* 27:592–593.
- Slatkin, M., and W. P. Maddison. 1989. A cladistics measure of gene flow inferred from the phylogeny of alleles. *Genetics* 123:603–613.
- Spiehl, P. T. 1974. Gene flow and genetic differentiation. *Genetics* 78:961–965.
- Stamatakis, A., T. Ludwig, and H. Meier. 2005. Raxml-iii: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics* 21:456–463.
- Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A fast bootstrapping algorithm for the RAxML web-servers. *Syst. Biol.* 57:758–771.
- Stephens, M., N. Smith, and P. Donnelly. 2001. A new statistical method for haplotype reconstructions from population data. *Am. J. Hum. Genet.* 68:978–989.
- Steel, M. A., and D. Penny. 1993. Distributions of tree comparisons metrics—some new results. *Syst. Biol.* 42:126–141.
- Swofford, D. L. 2002. Paup*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
- Than, C., D. A. Ruths, and L. Nakhleh. 2008. PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. *BMC Bioinform.* 9:322.
- Wilgenbusch, J. C., D. L. Warren, and D. L. Swofford. 2004. AWTY: A system for graphical exploration of MCMC convergence in Bayesian phylogenetic inference.
- Wu, Y. 2012. Tree topologies under incomplete lineage sorting by maximum likelihood. *Evolution* 66:763–775.
- Yang, Z. 2002. Likelihood and Bayes estimation of ancestral population sizes in Hominoids using data from multiple loci. *Genetics* 162:1811–1823.
- Yu, Y., J. H. Degnan, and L. Nakhleh. 2012. The probability of a gene tree topology within a phylogenetic network with applications to hybridization detection. *PLoS Genet.* 8:e1002660.

Associate Editor: M. Hart

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Avian DNA sequences used to build character matrices are available from NCBI Genbank (accession HQ719474–HQ720138). Character matrices and tree files for analyses are available from TreeBase (<http://purl.org/phylo/treebase/phyloids/study/TB2:S13400?x-access-code=cetcb32d3f14a97d24682621836fae1f&format=html>).

Figure S1. Map of ingroup (*Poecile*) species ranges in North America and sampling localities.

Figure S2. Diagrams showing the workflow used in this study to sample, partition, and analyze loci.

Figure S3. Principal component analysis conducted in R using the number of parsimony informative sites, variable sites, and nodes resolved.

Figure S4. Individual nuclear RAxML gene trees with circles at nodes indicating bootstrap values: black > 95%, gray > 75%, and white < 75%.

Figure S5. The gene tree given by the concatenated mitochondrial DNA with branch lengths and posterior probabilities from the MrBayes analysis.

Figure S6. The top most likely species networks conducted in PhyloNet accounting for both incomplete lineage sorting and gene flow.

Figure S7. The bootstrap version of the STEM analysis using 40 nuclear loci.

Table S1. Sample identification numbers, museums, and collecting locations of samples.

Table S2. Nuclear loci GenBank IDs, primer sequences, and annealing temperatures.

Table S3. List of molecular markers and descriptions.

Table S4. Composition of randomly sampled loci for each data set.

Table S5. Composition of loci for data sets based on ranked attributes.