Comparison of Target-Capture and Restriction-Site Associated DNA Sequencing for

Joseph D. Manthey[1]*, Luke C. Campillo[1], Kevin J. Burns[2], and Robert G. Moyle[1]

[1]*Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, KS 66045, USA and*
[2]*Department of Biology, San Diego State University, San Diego, CA 92182, USA*
Correspondence to be sent to: Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Dyche Hall, 1345 Jayhawk Blvd., Lawrence, KS 66045-7561, USA; Email: jdmanthey@gmail.com

Abstract.—Restriction-site associated DNA sequencing (RAD-seq) and target capture of specific genomic regions, such as ultraconserved elements (UCEs), are emerging as two of the most popular methods for phylogenomics using reduced-representation genomic data sets. These two methods were designed to target different evolutionary timescales: RAD-seq was designed for population-genomic level questions and UCEs for deeper phylogenetics. The utility of both data sets to infer phylogenies across a variety of taxonomic levels has not been adequately compared within the same taxonomic system. Additionally, the effects of uninformative gene trees on species tree analyses (for target capture data) have not been explored. Here, we utilize RAD-seq and UCE data to infer a phylogeny of the bird genus *Piranga*. The group has a range of divergence dates (0.5–6 myr), contains 11 recognized species, and lacks a resolved phylogeny. We compared two species tree methods for the RAD-seq data and six species tree methods for the UCE data. Additionally, in the UCE data, we analyzed a complete matrix as well as data sets with only highly informative loci. A complete matrix of 189 UCE loci with 10 or more parsimony informative (PI) sites, and an approximately 80% complete matrix of 1128 PI single-nucleotide polymorphisms (SNPs) (from RAD-seq) yield the same fully resolved phylogeny of *Piranga*. We inferred non-monophyletic relationships of *Piranga lutea* individuals, with all other *a priori* species identified as monophyletic. Finally, we found that species tree analyses that included predominantly uninformative gene trees provided strong support for different topologies, with consistent phylogenetic results when limiting species tree analyses to highly informative loci or only using less informative loci with concatenation or methods meant for SNPs alone. [Birds; phylogenomics; *Piranga*; restriction-site associated DNA sequencing; species tree methods; target capture; ultraconserved elements.]

Restriction-site associated DNA sequencing (RAD-seq; Miller et al. 2007) and target capture of specific genomic regions, such as ultraconserved elements (UCEs; Faircloth et al. 2012), are emerging as two of the most useful methods for phylogenomics. RAD-seq was initially developed for identifying and studying recent population divergences, and has been successfully utilized for phylogeographic studies in a variety of organisms, including *Sarracenia* pitcher plants (Zellmer et al. 2012), *Lycaeides* butterflies (Gompert et al. 2010), and an assortment of bird species (McCormack et al. 2012). At deep timescales, RAD-seq may recover few orthologous loci among divergent lineages, with potential rate bias in the loci that do overlap (i.e., they are identified as orthologous because of slow mutation rate; Rubin et al. 2012). Conversely, target-capture using UCEs was developed for studying deep divergences and difficult phylogenetic questions, and has since been utilized to infer higher-level phylogenies in a diverse set of taxa, including Hymenoptera (Faircloth et al. 2014), ray-finned fishes (Faircloth et al. 2013), and birds (McCormack et al. 2012). In contrast to RAD-seq, the drawbacks of using UCEs are at recent timescales, due to insufficient time to accumulate informative sites next to the conserved target regions (e.g., Smith et al. 2013).

Although many studies have investigated systematic or phylogeographic questions using one of these methods, few have explored both methods in the same study system. Harvey and colleagues (2013) used UCEs and RAD-seq to study phylogeographic patterns in a widespread Neotropical songbird (*Xenops minutus*) with shallow divergences (<5 myr); they identified concordant phylogenetic relationships between data sets. Conversely, Leaché and colleagues (2015) used sequence capture and RAD-seq to study deep systematics (15–55 myr) of phrynosomatid lizards; here, the authors found different results between data sets depending on which pipeline and filtering steps were used.

In addition to the inherent differences between target-capture and RAD-seq data sets, other issues may influence species tree inference within target-capture data sets alone. Multiple simulation studies have identified that species tree inference is negatively influenced by inclusion of weak—or largely uninformative—loci (Lanier et al. 2014; Liu et al. 2015). With the rapid expansion of target-capture data for phylogenetic inference, it is necessary to assess the impacts of uninformative loci (and their estimated gene trees) on species tree inference across methodologies with empirical data sets. In contrast, because RAD-seq loci are inherently short—often less than 100 bp—phylogeneticists have generally not attempted species tree reconstruction using the underlying gene trees. Therefore, we are limiting our assessment of low information loci on species tree inference to target capture, as we already presume that gene trees generated from RAD loci will be mostly unresolved.

Here, we extend previous comparisons of sequence capture and RAD-seq for phylogenetics in a genus of songbirds (*Piranga*, Aves: Cardinalidae) with a range of divergence dates [0.5–6 myr based on mitochondrial DNA (mtDNA); Burns 1998] among 11 recognized species and recognized intraspecific divergences within three of the species (Table 1; Burns 1998, IOC World Bird List v5.1). With most of the intragenus relationships unresolved from mtDNA alone (Burns 1998, Shepherd and Burns 2007), *Piranga* provide an empirical system that is an appropriate size and age to examine the relative utility of phylogenetic inference using RAD-seq- and UCE-based data sets.

We used Illumina sequencing to obtain RAD-seq and UCE data sets for 28 individuals across 11 species in *Piranga* (Table 1). With these data, we used seven species tree estimation methods to investigate the following questions: 1) Do RAD-seq and UCE data sets perform similarly—that is, provide sufficient phylogenetic signal—across a range of timescales (0.5–6 myr) common in many phylogenetic studies? 2) Do species tree analyses of RAD-seq and UCE data sets estimate comparable species relationships? 3) How do UCE loci with few or no informative sites influence species tree estimation?

METHODS

Sampling and Sample Preparation

Fresh tissue or blood samples of 28 *Piranga* individuals, two each per recognized species-level taxonomic unit (Burns 1998, IOC World Bird List v5.1) as well as each genetically distinct lineage (e.g., Shepherd and Burns 2007), were obtained from natural history museums (Table 1). Two samples of *Cardinalis cardinalis* were used as outgroup individuals (Klicka et al. 2007). Multiple intraspecific lineages were included in analysis for *Piranga bidentata*, *Piranga ludoviciana*, and *Piranga rubra* (Table 1), for which multiple clades have been identified or hypothesized. We used a QIAGEN DNeasy blood and tissue kit to extract genomic DNA from each individual. We used Qubit Fluorometric Quantitation (Life Technologies) to quantify and standardize DNA concentrations of all samples.

Sequence Capture Laboratory Procedures and Bioinformatics

We performed target-capture with the Tetrapods-UCE-5Kv1 probe set (available at ultraconserved.org), which targets 5060 UCE loci. Initially, genomic DNA (∼500 ng) was fragmented using a Covaris S220

TABLE 1. Samples used in this study, their museum voucher numbers, locality to state and country, number of RAD-seq single-end reads, and number of UCE paired-end reads (read 1, read 2, and singletons), both following quality control

Species (group)	Museum number	Locality	Number of RAD reads	Number of UCE reads
Piranga bidentata (mainland)	MZFC 15161	Hidalgo, Mexico	4,241,415	3,832,093
Piranga bidentata (mainland)	MZFC 17737	Hidalgo, Mexico	4,946,952	4,970,928
Piranga bidentata (Isla Tres Marias)	MZFC 19262	Nayarit, Mexico	1,385,346	4,233,936
Piranga bidentata (Isla Tres Marias)	MZFC 19257	Nayarit, Mexico	4,330,638	3,788,097
Piranga erythrocephala	FMNH 343370	Jalisco, Mexico	625,648	1,070,726
Piranga erythrocephala	FMNH 343371	Jalisco, Mexico	4,031,787	2,821,347
Piranga flava	KU 90809	Upper Takutu, Guyana	2,726,534	4,536,216
Piranga flava	LSU B15408	Santa Cruz, Bolivia	6,798,099	2,809,217
Piranga hepatica	KU 4970	Morazon, El Salvador	7,562,397	2,982,228
Piranga hepatica	KU 9084	Morazon, El Salvador	344,750	2,384,144
Piranga leucoptera	FMNH 481795	Jalisco, Mexico	8,007,493	3,061,556
Piranga leucoptera	LSU B7783	Pichincha, Ecuador	6,000,079	4,998,833
Piranga ludoviciana (mtDNA Clade A)	SDSU 2208	California, USA	3,051,629	1,935,630
Piranga ludoviciana (mtDNA Clade A)	SDSU 2650	California, USA	957,011	1,810,446
Piranga ludoviciana (mtDNA Clade B)	SDSU 2385	California, USA	3,629,473	7,222,535
Piranga ludoviciana (mtDNA Clade B)	SDSU 2648	California, USA	4,983,759	2,993,931
Piranga lutea (east)	KU 89864	Upper Takutu, Guyana	4,762,028	2,989,652
Piranga lutea (west)	LSU B5400	San Martin, Peru	4,725,109	3,120,335
Piranga olivacea	KU 2699	Missouri, USA	2,644,667	4,064,639
Piranga olivacea	KU 4672	Missouri, USA	3,861,784	3,582,762
Piranga roseogularis	KU 2049	Campeche, Mexico	1,137,659	2,390,649
Piranga roseogularis	KU 2141	Campeche, Mexico	8,396,195	2,352,851
Piranga rubra (east)	KU 26592	Kansas, USA	2,590,769	3,243,467
Piranga rubra (east)	KU 7046	Kansas, USA	7,823,448	2,731,446
Piranga rubra (west)	SDSU 2635	Arizona, USA	5,305,901	4,438,132
Piranga rubra (west)	SDSU 2654	California, USA	7,575,459	4,420,569
Piranga rubriceps	LSU B265	Cajamarca, Peru	7,578,165	3,686,237
Piranga rubriceps	LSU B34818	Cajamarca, Peru	5,812,555	5,511,602
Cardinalis cardinalis	KU 21828	Florida, USA	2,598,037	4,465,045
Cardinalis cardinalis	KU 25393	Kansas, USA	3,610,110	4,308,623

Notes: KU, University of Kansas Natural History Museum; LSU, Louisiana Museum of Natural Science; FMNH, Field Museum of Natural History; SDSU, San Diego State University Museum of Biodiversity; MZFC, Museo de Zoología de la Facultad de Ciencias at Universidad Nacional Autónoma de México.

focused-ultrasonicator (175 W peak incident power, 2% duty factor, and 200 cycles per burst for 45 s). Sonicator conditions were based on optimized settings for unrelated UCE projects at KU using other avian samples. We then prepared sequencing libraries from the fragmented DNA using Kapa Biosystems Library Prep Kits (KBLPK; Kit: KK8201). Samples were subject to end repair and A-tailing following KBLPK instructions. We ligated adapters compatible with the iTruStub dual indexing system (available at baddna.org). Samples were subjected to limited-cycle (10 cycles) PCR to index and amplify individual samples prior to UCE enrichment, followed by pooling of indexed samples (16 samples per pooled library, including two samples from an unrelated project). We enriched the pooled libraries for the 5Kv1 probe set using a Mycroarray MyBaits Kit, followed by a brief (16–17 cycles) PCR reaction to amplify the small amount of each library enriched for UCEs. Although the protocol recommended 12–16 cycles at this step, we ran 17 cycles of PCR on one library to compensate for a lower yield in the UCE enrichment step. The pooled libraries were tested for quality and quantity of DNA using quantitative PCR and the Agilent Tapestation at the University of Kansas Genome Sequencing Core Facility, followed by sequencing of 100 bp paired-end reads on a partial (31.25%) lane of an Illumina HiSeq2500.

Bioinformatics of UCE data largely used the PHYLUCE software package of Python v2.7 scripts (Faircloth 2015). The *illumiprocessor.py* script was used to trim adapter contamination and check quality of sequencing reads. Next, we used Trinity (version: rnaseq_r2013_08_14; Grabherr et al. 2011) to assemble contigs for each individual. We matched our assembled contigs to the UCE probe set using the *match_contigs_to_probes.py* script. We aligned each UCE locus using the *seqcap_align_2.py* script, which batch calls the use of MAFFT (Katoh and Standley 2013). Finally, custom R (R Development Core Team 2012) and Python scripts were used to obtain summary statistics of all loci and convert file types for various phylogenetic analyses (see Data Accessibility and Supplementary Information, available on Dryad at http://dx.doi.org/10.5061/dryad.j5n06).

RAD-seq Laboratory Procedures and Bioinformatics

We performed a single digest modified RAD-seq (Miller et al. 2007; Andolfatto et al. 2011) protocol to obtain many anonymous genetic loci. Genomic DNA samples (∼50 ng per sample) were initially digested with the restriction enzyme NdeI as a first step to produce a reduced representation genomic library. We used NdeI because of positive results from previous projects (e.g., Manthey and Moyle 2015) and the required overhang for the associated indexing barcodes (Andolfatto et al. 2011). We ligated custom adapters with attached barcodes for multiplexing, followed by pooling of samples. Each sample was barcoded in triplicate (Supplementary Table S1, available on Dryad)

to minimize sample bias from individual barcodes. We used a Pippin Prep electrophoresis cassette (Sage Science) to size select fragments in the range of 450–600 bp. Following size selection, samples were subjected to a brief polymerase chain reaction (PCR) of the pooled samples in duplicate, which added standard Illumina sequencing indices (i.e., additional indices to dual-index; Andolfatto et al. 2011). The pooled libraries were tested for quality and quantity of DNA using quantitative PCR and the Agilent Tapestation at the University of Kansas Genome Sequencing Core Facility, followed by sequencing of 100 bp single-end reads on a single lane of an Illumina HiSeq2500.

We used the STACKS (Catchen et al. 2011) pipeline to assemble loci de novo from the fastQ files obtained from the Illumina sequencing run. We assigned sequence reads to individuals and removed reads of poor quality using the *process_RADtags* script included in the STACKS pipeline. For inclusion, sequencing reads were required to have an average phred score of 10 in sliding windows of 15 bp. We removed sequences lacking the restriction site or containing possible adaptor contamination. Next, we used the *ustacks*, *cstacks*, and *sstacks* modules of STACKS with the default settings, with the exception of allowing up to five mismatches between individuals (94% sequence identity between individuals' stacks). The mismatches setting was changed to increase the number of loci recovered across individuals, as the default value of 2 would have required a 98% sequence identity across each RAD locus. Finally, we created a SNP data set using the *populations* module of STACKS with the following restrictions: minimum stack depth of five, a minimum of one individual per lineage [as informed from a RAxML (Stamatakis 2014) tree produced from concatenated UCE data (Table 1; Fig. 1)], minor allele frequency of 0.05, and observed heterozygosity across all individuals no greater than 0.5 (to reduce inclusion of paralogous loci). To assess the effect of stack depth, we reran the *populations* module with varying minimum stack depths ($m = 5$, 10, 15); here, the number of loci changed among data sets, while pairwise comparisons of genetic differentiation (F_{ST}) remained the same ($R^2 > 0.98$) and phylogenetic relationships did not change (Supplementary Information, available on Dryad) with the exception of the Hepatic Tanager clade (see "Results" section).

Phylogenetic Inference Using UCEs

We used two phylogenetic methods with the concatenated complete matrix (all individuals sampled; 734 loci) data set. First, we used maximum likelihood in RAxML v8.1.18 (Stamatakis 2014) with a GTR+G model of sequence evolution. Support was assessed using 1000 rapid bootstrap replicates. Second, we estimated a species tree using SVDquartets (Chifman and Kubatko 2014) implemented in PAUP* v4a142 (Swofford 2003). SVDquartets samples quartets of individuals' sequences and infers an unrooted phylogeny, followed by species

* = Full Support

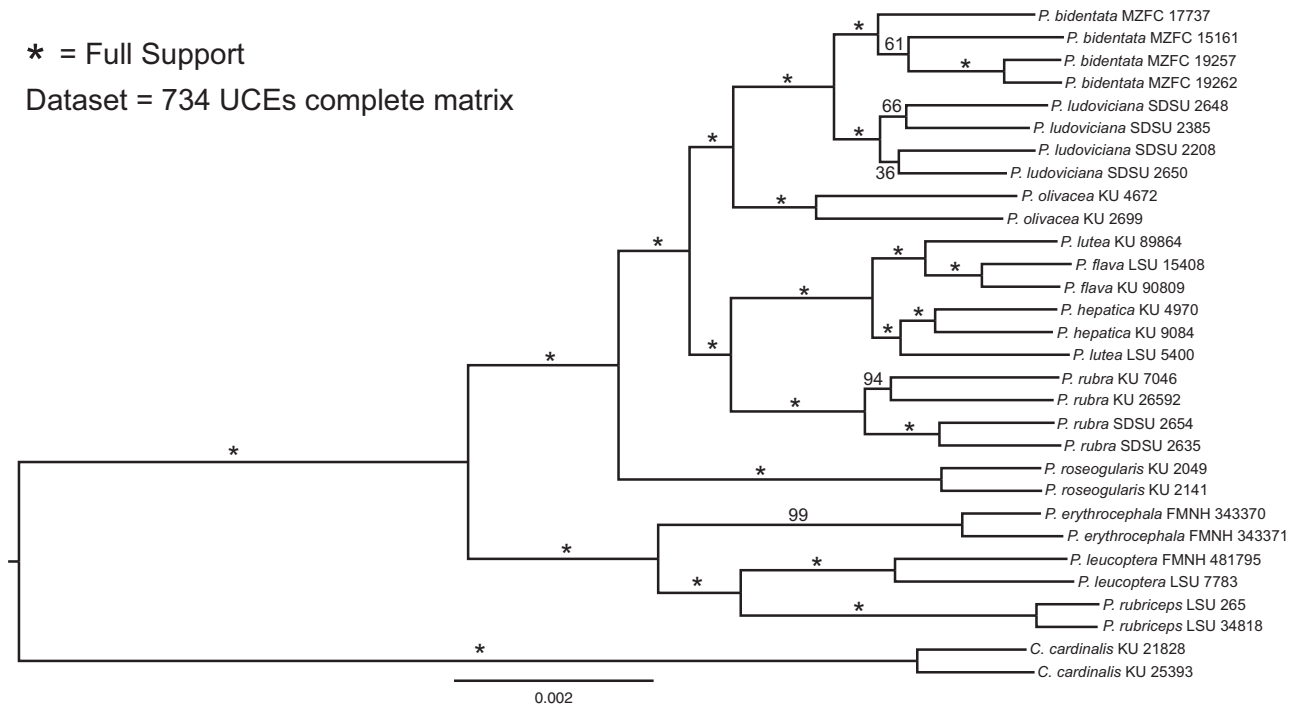**Dataset = 734 UCEs complete matrix**

FIGURE 1.   Maximum-likelihood phylogeny of the complete concatenated UCE data set. Support values are based on 1000 quick bootstrap replicates.

tree inference using all sampled quartets. With this method, we exhaustively sampled quartets from the data set and inferred a species tree from the quartets.

In addition to analyses utilizing concatenated data sets, we inferred species trees using four methods that sample individual gene trees to infer a species tree. For these species tree analyses, we used three data sets: 1) the complete data matrix (all individuals sampled; 734 loci), 2) all loci in the complete data matrix with 10 or more parsimony informative sites (PIS; 189 loci), and 3) all loci in the complete data matrix with 20 or more PIS (70 loci). This allowed us to assess the impact of uninformative gene trees on species tree reconstruction methods, as uninformative loci have been shown to influence summary-based species tree methods (Lanier et al. 2014; Liu et al. 2015).

To estimate gene trees for each locus, we used the Cloudforest pipeline (available at github.com/ngcrawford/CloudForest), which implements MrAIC to estimate the best substitution model for each locus, and then uses PhyML (Guindon et al. 2010) to estimate gene trees. For each data set, 100 multilocus bootstraps (Seo 2008) were performed; each bootstrap resamples loci within the data set, as well as bases within each locus. Because this process resamples loci across the entire data set, the bootstrapping procedure was performed separately for each data set (based on PIS; see Data Accessibility and Supplementary Information, available on Dryad). To ensure that species trees estimated from PhyML gene trees were robust, we also created multilocus bootstrapped gene tree data

sets using RAxML (GTR+Gamma model of sequence evolution). Again, as for the PhyML gene trees, each locus analyzed with RAxML was subject to multilocus bootstrapping.

Using the gene trees as input, we used three methods of species tree inference: 1) Species trees from average ranks of coalescence (STAR; Liu et al. 2009), 2) accurate species tree algorithm (ASTRAL; Mirarab et al. 2014), and 3) maximum pseudo-likelihood of estimating species trees (MP-EST; Liu et al. 2010) These three species tree methods were performed for both PhyML and RAxML generated sets of gene trees.

We also used Bayesian Evolutionary Analysis by Sampling Trees [*BEAST (Heled and Drummond 2010), implemented in BEAST v.2 (Bouckaert et al. 2014)]. *BEAST simultaneously infers gene trees for each locus and estimates a most-probable species tree given the gene trees from the multi-individual, multilocus sequence data (Heled and Drummond 2010). Tree models were left unlinked across all partitions. We used a Yule process species tree prior, and relaxed lognormal molecular clocks for each locus with relative rates estimated, and an HKY model of sequence evolution. To assess the effects of changing clock settings or model of sequence evolution, we attempted to run *BEAST with a JC model of sequence evolution and a strict clock; with these settings, we were unable to achieve convergence after more than 8 weeks of analysis time. We assessed convergence and stationarity using TRACER (Rambaut and Drummond 2007) and AWTY (Nylander et al. 2008). In TRACER,

we assessed stationarity by plateauing of parameter estimates and sufficient sampling when effective sample sizes (ESSs) reached 200 (parameters: posterior, likelihood, prior, speciescoalescent, birthrate.t:Species, and YuleModel.t:Species). In AWTY, we used the *cumulative* and *compare* utilities to visualize the cumulative posterior probabilities of tree bipartitions across replicate runs and posterior probabilities of tree bipartitions of individual runs through time. Each data set was run for multiple—but varying in number—runs (Supplementary Table S1, available on Dryad) of two billion generations, with tree and parameter estimates sampled every 1,000,000 MCMC generations. Burn-in was determined when stationarity was reached based on visualization in AWTY and TRACER.

Finally, we used the coalescent-based program SNAPP (Bryant et al. 2012), a likelihood species-tree method that uses only SNPs and is executable in BEAST v.2 (Bouckaert et al. 2014). We used SNAPP to have an additional method (other than SVDquartets) to compare results between target capture and RAD-seq data sets. For SNAPP, we extracted a single PI SNP from each UCE locus, which resulted in a final data set of 605 SNPs. In SNAPP, we used empirical estimates of major and minor allele base frequencies to inform mutation rate priors. We used gamma rate priors for the alpha and beta parameters, with all other priors for theta left as defaults implemented in SNAPP. Convergence and stationarity were assessed using the *compare* and *cumulative* utilities implemented in AWTY (Nylander et al. 2008). Two replicates of SNAPP were run for 100,000 burn-in iterations, followed by an additional two million MCMC generations. Tree and parameter estimates were sampled every 1000 MCMC generation. Burn-in was determined when stationarity was reached based on visualization in AWTY.

Phylogenetic Inference Using RAD-seq Data

We used two methods to estimate a phylogeny for *Piranga* using RAD-seq data. First, we estimated a species tree using SVDquartets (Chifman and Kubatko 2014) with identical methods to those used for the UCE data. For use in SVDquartets, we used whole sequence reads, to include both variant and invariant sites. Second, we used the coalescent-based program SNAPP (Bryant et al.

2012), with identical methods to those used for the UCE data.

Because of multiple potential species tree topologies in the hepatic tanager clade (*Piranga flava, Piranga hepatica, Piranga lutea*; see "Results" section), we used STRUCTURE (Pritchard et al. 2000) to assess potential admixture between recognized taxonomic units using a complete RAD-seq SNP data matrix for the clade (2960 SNPs represented in every sample; i.e., the final step of STACKS was rerun to identify SNPs present in every individual of the Hepatic Tanager clade). In STRUCTURE, we initially inferred lambda by estimating the likelihood of one population ($k = 1$) and allowing lambda to converge. Successive STRUCTURE runs used a fixed lambda and the admixture model. Five replicates of STRUCTURE were run for up to four populations ($k = 1$–4), using 50,000 burn-in steps followed by 100,000 MCMC iterations.

RESULTS

The mean numbers of RAD-seq and UCE reads per sample were 4.4 million (2.4 million SD) and 3.6 million (1.3 million), respectively (Table 1). RAD sequencing resulted in a mean of 40,267 (SD = 11,125) RAD-tags per individual (Supplementary Table S1, available on Dryad). Among samples, RAD-tag recovery was highly variable, resulting in a data set of 1374 loci, with a mean coverage of 109.7 reads (SD = 59.8) per individual for each included locus. Target capture resulted in approximately 80% recovery of UCE loci for each individual (mean = 4148 UCE loci, SD = 259.0), with an average UCE locus length of 874 bp across samples (SD = 159.3; Supplementary Table S1, available on Dryad). RAD loci were each 91 bp. A complete matrix, which included data at each locus for all sampled individuals, contained 734 UCE loci. Both data sets had high variability in number of PIS across loci (Table 2) as well as a similar number of total PIS. The RAxML phylogeny of the concatenated UCE data set identified strongly supported relationships in all but two clades (*P. bidentata* and *P. ludoviciana*; Fig. 1). With the exception of *P. lutea*, all *a priori* species were monophyletic. In *P. lutea*, a western individual (Peru) was identified as sister to *P. hepatica*, whereas an eastern individual (Guyana) was identified as sister

TABLE 2. Summary of each data set, including coverage and number of PI single-nucleotide polymorphisms (SNPs)

Method	UCE—Complete	UCE—10 PIS	UCE—20 PIS	RAD-seq
Number of loci	734	189	70	1374
Number of PI SNPs	5052	3520	1901	6926
Median PI SNPs per locus (SD)	4 (8.36)	17 (8.09)	25 (6.85)	5 (2.83)
Range PI SNPs per locus	0–50	10–50	20–50	0–19
Coverage (% individuals × loci)	100	100	100	79.71
Median length locus (SD)	918 (200.3)	1072 (137.5)	1124 (111.5)	91 (0)

Notes: Data sets include the UCE complete matrix, UCEs with a minimum of 10 PIS, UCEs with a minimum of 20 PIS, and the RAD-seq data set (minimum of one individual per species for a locus to be included).
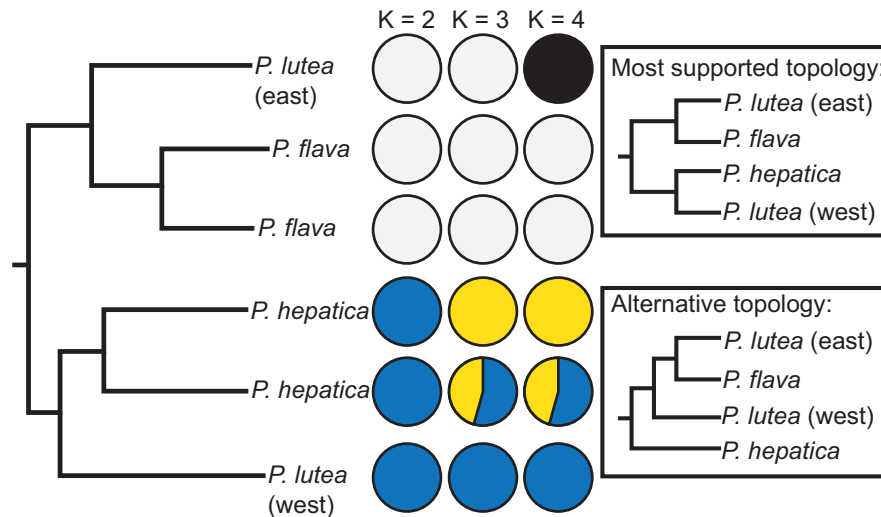
FIGURE 2.   RAxML topology of the concatenated UCE data (734 loci) for the Hepatic Tanager clade (from Fig. 1). At the tips are STRUCTURE results from RAD-seq data (complete matrix of 2960 SNPs) for two, three, and four inferred populations (proportion of color informs inferred probability of population ancestry). On the right are possible species tree topologies within this clade. Three analyses exhibited strong support for relationships in the most supported topology [*BEAST and ASTRAL (UCEs), and SNAPP (RAD-seq)]. SVDquartets (RAD-seq) was the only analysis to strongly support the alternative topology. All other analyses identified a polytomy ($<75\%$ bootstrap support) of *P. hepatica*, *P. lutea* (west), and *P. flava* + *P. lutea* (east).

to *P. flava* (Fig. 1). Results from STRUCTURE analyses corroborated patterns seen in the RAxML phylogenetic estimates of the concatenated UCE data set (Fig. 2) by identifying *P. lutea* eastern and western individuals most genetically similar to *P. flava* and *P. hepatica*, respectively. The hypothesized intraspecific genetic structure in *P. ludoviciana*, *P. bidentata*, and *P. rubra* were all supported (Fig. 1). Within *P. ludoviciana* and *P. rubra*, previously recognized mtDNA clades were confirmed (Shepherd and Burns 2007), while in *P. bidentata*, individuals from Isla Tres Marias grouped together within the mainland *P. bidentata* clade.

Comparison of RAD-seq and UCE Species Tree Methods and Effects of Uninformative UCE Loci on Species Tree Analyses

Species tree analyses of SNPs from the RAD-seq and UCE data sets in SNAPP (1128 and 605 SNPs, respectively) yielded identical results (Fig. 3) to those identified in the analyses of concatenated UCE data (Fig. 1). When limited to loci with a minimum of 10 PIS or 20 PIS, all UCE species tree methods identified the same relationships as the SNAPP analyses (Fig. 3, Table 3, Supplementary Information, available on Dryad). A strong correlation existed between branch lengths recovered in RAD-seq SNAPP and UCE *BEAST analyses ($R^2 = 0.864$, $P < 0.001$; Inset of Fig. 3). Although branch lengths were highly correlated, the RAD-seq tree had consistently shorter branch lengths than the UCE tree, particularly noticeable in terminal branches. Branch lengths may have been shorter in the RAD tree due to only using PIS, thereby reducing the amount of intraspecific and intraindividual genetic variation (i.e., autapomorphies). For both UCEs and RAD-seq data,

SVDquartets analyses identified identical topologies to SNAPP and *BEAST analyses, with the exception of relationships within the Hepatic Tanager clade (*P. hepatica P. lutea P. flava*; Fig. 2). Varying minimum stack depth (RAD-seq data sets) changed the level of support for the two Hepatic Tanager clade topologies (Fig. 2) in SVDquartets analyses, but did not affect relationships among any other taxa (Supplementary Information, available on Dryad).

The complete UCE matrix (734 loci) yielded three different topologies from different analyses. The ASTRAL method—using RAxML gene trees, but not PhyML gene trees—was the only analysis to produce concordant species trees between the complete data set and the data sets limited to highly informative loci. The ASTRAL analysis was also the only case in which the input gene trees (RAxML vs PhyML) had any impact on species tree topology (aside from the Hepatic Tanager clade). All other trees inferred from STAR, ASTRAL (PhyML only), and MP-EST with the full dataset were identical to one another. These analyses all identified *P. ludoviciana* as paraphyletic with *P. bidentata* monophyletic but nested within *P. ludoviciana*. The *BEAST analysis with the complete data set yielded the most disparate—but strongly supported—relationships within *Piranga*. Here, eight species and three clades contained different sister relationships (Supplementary Information, available on Dryad). Differences between the *BEAST tree and other analyses occurred in parts of the tree with short internodes (~2% or less of total tree height). For example, *Piranga leucoptera* was found sister to *Piranga erythrocephala* rather than *Piranga rubriceps* (as in all other analyses). Additionally, *P. rubra* was not found to be sister to the Hepatic Tanager clade, but rather sister to a clade of *P. bidentata*, *P. ludoviciana*, *P. olivacea*, *P. flava*,
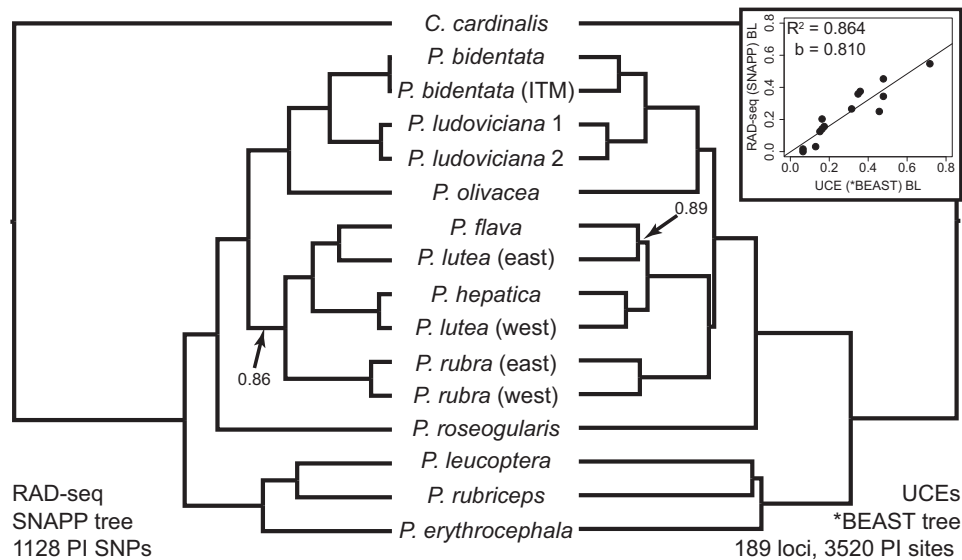
FIGURE 3. Identical species tree topologies from RAD-seq (SNAPP analysis) and UCE (*BEAST analysis) data sets. Most analyses, of both RAD-seq and UCE data, identified these relationships, with the exception of some analyses using gene trees with many uninformative UCE loci (Table 3). All unlabeled branches received greater than 0.95 posterior support values. The RAD-seq data set included only parsimony informative (PI; i.e., no singletons) sites. The UCE data set analyzed here included only highly informative loci (10 or more PI sites). The inset shows the comparison of relative branch lengths (BLs) between trees, with a slope (*b*) of approximately 0.8.

TABLE 3. Summary of RAD-seq and UCE-data set phylogeny reconstruction methods used in this study.

Method	Data set(s)		
Ultraconserved elements	Complete (734 loci)	10 PIS (189 loci)	20 PIS (70 loci)
SNAPP	+[a]	NA	NA
RAxML	+[b]	NA	NA
SVDquartets	+[b]	NA	NA
STAR	−/−	+/+	+/+
ASTRAL	−/+	+/+	+/+
MP-EST	−/−	+/+	+/+
*BEAST	−	+	+
RAD-seq	Complete (1374 loci)		
SNAPP	+[a]		
SVDquartets	+[b]		

Notes: Species tree methods were used on three UCE data sets: (1) complete matrix [734 loci, 5052 PIS], (2) loci with a minimum of 10 PIS (189 loci, 3520 PIS), and (3) loci with a minimum of 20 PIS (70 loci, 1901 PIS). If the estimated phylogeny matched the hypothesized best tree [Fig. 3; excluding relationships within Hepatic Tanager clade (*P. hepatica P. lutea P. flava*)], it is marked with a (+). Conversely, an estimated phylogeny that did not match the best tree is marked with a (−). For STAR, ASTRAL, and MP-EST, the results are labeled for analyses using PhyML gene trees or RAxML gene trees on the left/right, respectively.
[a]SNAPP was run using a data set with only SNPs (1128 and 605 SNPs for the RAD-seq and UCE data sets, respectively).
[b]RAxML and SVDquartets were run on a concatenated data set with all loci, including variant and invariant sites.

P. hepatica, and *P. lutea*. These differences highlight the effect of uninformative UCE loci on species tree analyses (especially *BEAST; Table 3), especially in situations with short internode lengths.

DISCUSSION

Here, we provide a comparison of RAD-seq and UCE data for phylogenetic inference using a suite of species tree inference methods. We fully resolved the phylogeny of *Piranga* and found nonmonophyly of *P. lutea*. However, we found that less informative UCE loci—and their respective gene trees—negatively impact some species tree methods. The bias caused by less informative loci is not likely to be restricted to UCE data sets, but biologists have generally not used gene trees for RAD-seq loci, likely due to limited information content per RAD locus. Indeed, analyses using SNPs extracted from UCE data—including all loci—recovered the hypothesized best topology, suggesting the negative impacts simply stem from using gene trees from many UCE loci with little information content.

Utility of RAD-seq and Target Capture for Phylogenetic Inference with Mixed Levels of Divergence

The utility of RAD-seq data for use in inferring interspecific phylogenies with deep divergences has been shown using *in silico* genome digests (e.g., Cariou et al. 2013) and empirical data (e.g., Cruaud et al. 2014). However, the nature of RAD-seq methodologies introduces genealogical biases in obtaining orthologous genetic markers (Arnold et al. 2013). Because the number of comparable RAD markers among species deteriorates through time because of mutations at restriction sites, estimating deep divergences using RAD-seq data may be problematic because cut sites preserved across deep phylogenies likely occur in regions that evolve at lower

rates than other parts of the genome (Rubin et al. 2012). This potential bias raises the question: Can RAD-seq data adequately estimate a phylogeny across a moderately divergent taxonomic group? Here, we compared the ability of RAD-seq and UCE data to estimate a phylogeny in *Piranga*. We found that analysis of RAD-seq data identified a phylogenetic topology identical to that derived from UCE data (Fig. 3). This result suggests that the method of data collection (RAD-seq vs. UCEs) does not necessarily influence the results of species tree analyses at intermediate levels of divergence.

Two recent studies investigated the utility of RAD-seq and target capture for phylogenetic estimation at different timescales. In a shallow (<5 myr) phylogeographic study, Harvey et al. (2013) found concordant results between data sets. Conversely, in a deeper systematic study (15–55 myr), Leaché and colleagues (2015) found conflict between methods depending on how the data were analyzed. Using target capture data, Leaché et al. (2015) consistently found the same species tree, using both concatenation and coalescent-based methods. Conversely, when using RAD-seq data, no species trees matched that of the species tree from target capture data (Leaché et al. 2015). Additionally, using only the RAD-seq data, Leaché and colleagues (2015) identified different topologies between concatenation and species tree methods, as well as differences within a method (e.g., using only concatenation) when different clustering thresholds were used to group RAD loci.

We provide here a view at an intermediate timescale (0.5–6 myr), with denser sampling (multiple individuals per phylogenetic tip) that allowed a broader range of species tree inference methods (Table 3), and find concordant results among data sets (Fig. 3; Table 3). Unfortunately, our study system and design differed from Leaché et al. (2015) in multiple respects. First, we used a more recent timescale, which may have facilitated the use of both target-capture- and RAD-seq-based analyses. Second, different restrictions of coverage across loci (i.e., minimum depth of sequence reads), in both target capture and RAD-seq data sets, could potentially influence species tree inference. Finally, sampling different numbers of individuals per phylogenetic tip (i.e., population or species) could influence the power to resolve relationships.

Evolutionary History and Taxonomy

We identified a strongly supported phylogeny, which was mostly consistent with previous mtDNA phylogenies (Burns 1998, Shepherd and Burns 2007), with a few important differences. Most mtDNA phylogenies identified *Piranga roseogularis* as the sister taxon to all other *Piranga*; however, we find a clade of *P. erythrocephala*, *P. leucoptera*, and *P. rubriceps* as sister to all other *Piranga*. In addition, the positions of *P. rubra* and *P. olivacea* with respect to other *Piranga* were unclear in the mtDNA studies, but are clarified with our data. *Piranga rubra* is sister to the Hepatic Tanager clade (*P. flava*, *P. hepatica*, *P. lutea*), and *P. olivacea* is sister to a clade including *P. bidentata* and *P. ludoviciana*. A recent phylogenetic study of Emberizoidea, which includes *Piranga* and many other genera, identified many of the relationships we found (Barker et al. 2015). The only discrepancy between Barker and colleagues' (2015) phylogeny and ours was the position of *P. olivacea* within the clade containing *P. olivacea*, *P. bidentata*, *P. ludoviciana*, *P. flava*, *P. lutea*, *P. hepatica*, and *P. rubra*.

Comparison of three closely related lineages (as identified in Fig. 4) reveals that only one contains complete allopatric replacements. This pattern suggests that allopatric replacement is prevalent within *Piranga* during early divergence, soon followed by renewed sympatry. For example, *P. hepatica*, *P. lutea*, and *P. flava* are all allopatric replacements, with partial sympatry with sister species *P. rubra*. This pattern is repeated in clade 3 (Fig. 4); here, *P. leucoptera* and *P. rubriceps* are allopatric replacements of one another, with their sister species, *P. erythrocephala*, partially sympatric with both. These patterns suggest that early stages of differentiation are due to geographic separation, lending support to allopatric speciation as the main mechanism of diversification in *Piranga*.

We found that one currently recognized species— *P. lutea*—is not monophyletic. A sample of *P. lutea* (subspecies *haemelea*) from the Guiana Shield is sister to *P. flava*, whereas a sample of western *P. lutea* (subspecies *lutea*) from the Andes is sister to *P. hepatica*. Although Burns (1998) did not find this relationship, their results do not contradict these relationships of *P. lutea*. Burns (1998) found western *P. lutea* sister to *P. hepatica*; however, the disjunct eastern range of *P. lutea* was not sampled in their study. These results may suggest that the Hepatic Tanager clade may be more than three distinct species because the Highland Tanager (*P. lutea*) has two distinct phylogenetic groups that are sister to different recognized species within the clade. The two lineages of *P. lutea* represent geographically and phenotypically differentiated subspecies of *P. lutea* (Hilty 2016); additionally, the other two taxa within the Hepatic Tanager group—*P. flava* and *P. hepatica*— show phenotypic differences from *P. lutea* (Hilty 2016). Differential morphologies, geographic ranges, and genetic histories lend support to each of the four lineages being considered distinct species under the evolutionary species concept (Wiley and Lieberman 1981), where all of the four taxa are on distinct evolutionary trajectories. Alternatively, these results could be interpreted as suggesting that the four groups of Hepatic Tanager are not each valid species, and that only one species exists. This alternative is plausible because the divergences within this clade are similar to those within other *Piranga* species with intraspecific sampling (Fig. 1) Additionally, the phenotypic differences between the taxa are complicated due to large seasonal and age plumage variation (Hilty 2016), for which we lacked solid geographical sampling because of the scope of
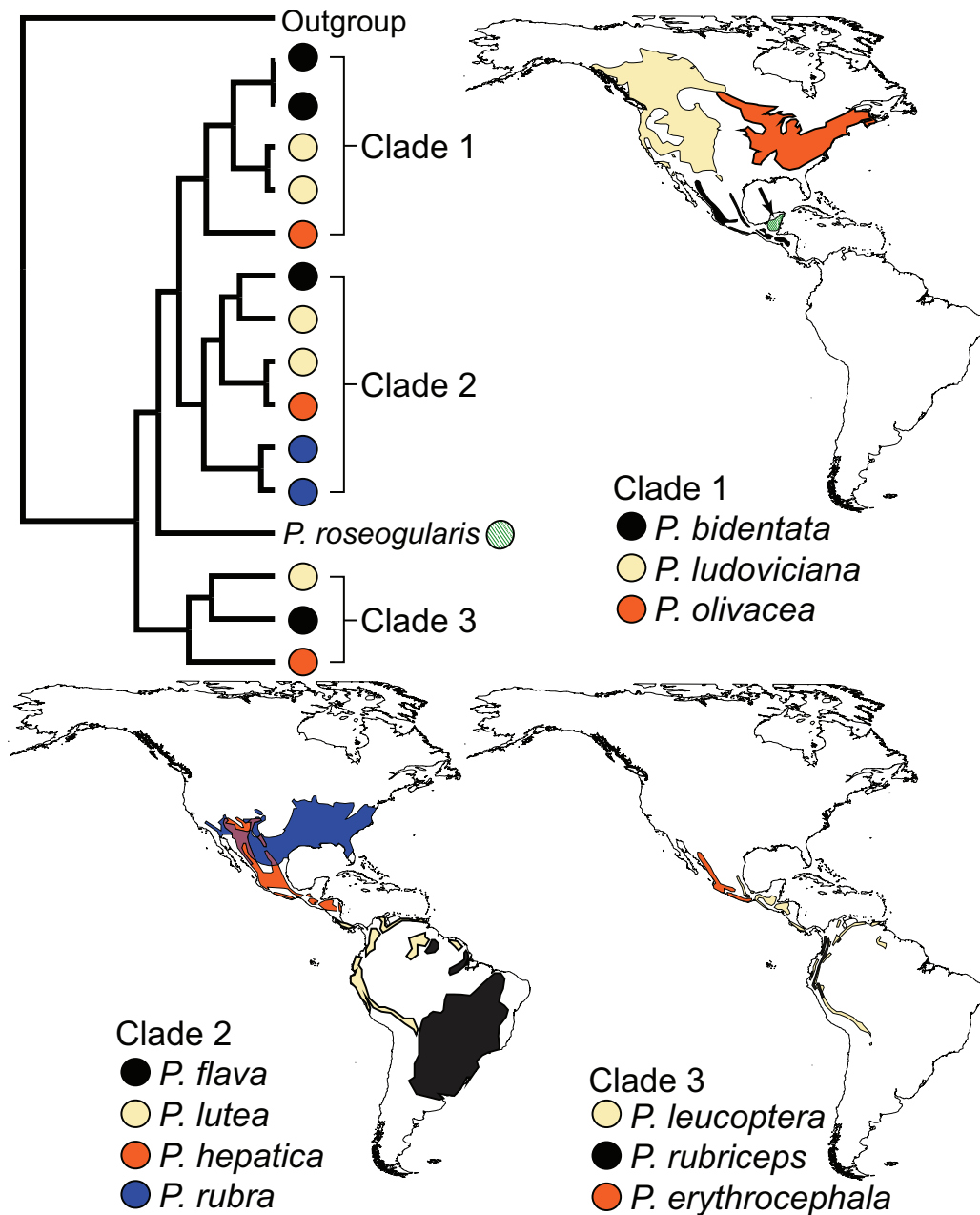
FIGURE 4. Allopatric replacement in one of three main clades in *Piranga*. All distributions represent breeding range. *Piranga roseogularis* (diagonal stripes and indicated with an arrow, shown with Clade 1) is allopatric to all other species and present only in the Yucatan Peninsula of Mexico. Note that *P. lutea* (Clade 2) has a disjunct distribution, where an eastern individual is sister to *P. flava*, and a western individual is sister to *P. hepatica*. The only case of strict allopatric replacement within a clade is exhibited in Clade 1.

this study. This clade warrants future phylogeographic investigation.

Import of Informative Loci for Species Tree Estimation

Although concatenated and SNP-based analyses of the overall UCE data set produced results concordant with methods using RAD-seq data, species tree analyses—those that utilize gene trees of each locus—that included low-information loci appeared to produce unreliable results (also see Hosner et al. 2016). Species-tree analyses of the complete UCE matrix (including uninformative loci) produced multiple distinct topologies; most species tree analyses recovered a slightly changed topology whereas the *BEAST analysis identified a much different topology. Conversely, species-tree analyses of data sets that included only informative loci (i.e., strong loci) produced the same species tree estimate for every method. Interestingly, longer loci generally had more PIS (Table 2), suggesting that researchers should strive to use

library preparation methods that produce longer loci. Differences here are likely due to how different species tree methods handle gene trees from uninformative and low information loci. For example, our complete UCE matrix contained 129 loci with no PIS, and a total of 448 loci with 5 or fewer PIS. It is unlikely that these loci with few informative sites provide much phylogenetic signal for most relationships in the phylogeny. The methods used for species tree analysis with the RAD-seq data (SNAPP and SVDquartets) potentially circumvent information-content issues by considering each site to have an independent genealogy under the coalescent (Bryant et al. 2012; Chifman and Kubatko 2014) and not relying on underlying gene tree topologies for species tree estimation (such as in gene tree-based coalescent methods). As such, some studies have used SNP data extracted from UCEs to estimate species trees (e.g., this study and McCormack et al. 2016) to potentially circumvent information content issues of some loci. Using these two methods (i.e., SVDquartets and SNAPP) with SNPs extracted from UCE loci, we identified the phylogeny recovered from RAD-seq data and concatenated UCE data (Figs. 1 and 3). As such, incorporating less informative loci had no negative impact when analyzed in a concatenated data set, or in a method utilizing only SNPs (Table 3).

When testing the performance of MP-EST for species tree inference in a six-taxon simulation, Liu et al. (2015) found that increasing the number of low-information loci did not increase the probability of obtaining the correct species tree (e.g., >80 loci), whereas high-information loci produced the correct species tree 100% of the time with as few as 30 loci. Similarly, Lanier et al. (2014) identified that high- and low-variation loci yielded tremendous differences in accuracy of species-tree methods and recommended that quality, and not just quantity of loci be considered for phylogenetic studies. Here, we provide an empirical example of this situation, where our data set of 70 high-information UCE loci (with 20 PIS or more) was more consistent across species tree analyses—and obtained our hypothesized best species tree—compared with the much larger complete data set (734 UCE loci) which included many low-information UCE loci.

Conclusions

We created RAD-seq and UCE data sets to infer the phylogeny of the bird genus *Piranga*, with the following goals: 1) Assess the utility and comparability of RAD-seq and UCE for interspecific comparisons, 2) fully resolve taxonomic relationships in *Piranga*, and 3) assess the impacts of uninformative UCE loci on species tree reconstruction. We identified a fully resolved phylogeny of *Piranga*, with concordant topologies between RAD-seq and UCE data. We identified novel relationships in *Piranga*, relative to a previous study, and found that one currently recognized species (*P. lutea*) is not monophyletic. Finally, we found conflicting results from species tree analyses that included gene trees from uninformative UCE loci, with completely congruent results when limiting species tree analyses to data sets with only highly informative loci or only utilizing less informative loci with concatenation or methods meant for SNPs alone.

References

Andolfatto P., Davison D., Erezyilmaz D., Hu T.T., Mast J., Sunayama-Morita T., Stern D.L. 2011. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. Genome Res. 21:610–617.

Arnold B., Corbett-Detig R.B., Hartl D., Bomblies K. 2013. RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. Mol. Ecol. 22:3179–3190.

Barker F.K., Burns K.J., Klicka J., Lanyon S.M., Lovette I.J. 2015. New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. Auk 132:333–348.

Bouckaert R., Heled J., Kühnert D., Vaughan T., Wu C.H., Xie D., et al. 2014. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput. Biol. 10:e1003537.

Bryant D., Bouckaert R., Felsenstein J., Rosenberg N.A., RoyChoudhury A. 2012. Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol. Biol. Evol. 29:1917–1932.

Burns K.J. 1998. Molecular phylogenetics of the genus *Piranga*: Implications for biogeography and the evolution of morphology and behavior. Auk 115:621–634.

Cariou M., Duret L., Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An *in silico* assessment and optimization. Ecol. Evol. 3:846–852.

Catchen J.M., Amores A., Hohenlohe P., Cresko W., Postlethwait J.H. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3: Genes, Genomes, Genetics. 1:171–182.

Chifman J., Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. Bioinformatics 30:3317–3324.

Cruaud A., Gautier M., Galan M., Foucaud J., Sauné L., Genson G., Dubois E., Nidelet S., Deuve T., Rasplus J.Y. 2014. Empirical assessment of RAD sequencing for interspecific phylogeny. Mol. Biol. Evol. 31:1272–1274.

Faircloth B.C., McCormack J.E., Crawford N.G., Harvey M.G., Brumfield R.T., Glenn T.C. 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. Syst. Biol. 61:717–726.

Faircloth B.C., Sorenson L., Santini F., Alfaro M.E. 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLoS One 8:e65923.

Faircloth B.C. 2015 PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics. btv646.

Faircloth B.C., Branstetter M.G., White N.D., Brady S.G. 2014. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among *Hymenoptera*. Mol. Ecol. Res. 15:489–501.

Gompert Z., Forister M.L., Fordyce J.A., Nice C.C., Williamson R.J., Buerkle C.A. 2010. Bayesian analysis of molecular variance in pyrosequences quantifies population genetic structure across the genome of *Lycaeides* butterflies. Mol. Ecol. 19:2455–2473.

Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N., Regev A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Guindon S., Dufayard J.F., Lefort V., Anisimova M., Hordijk W., Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: Assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Harvey M.G., Smith B.T., Glenn T.C., Faircloth B.C., Brumfield R.T. 2013. Sequence capture versus restriction site associated DNA sequencing for phylogeography. arXiv preprint arXiv:1312.6439.

Heled J., Drummond A.J. 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27:570–580.

Hilty S. 2016. Highland Hepatic-tanager (Piranga lutea). In: del Hoyo J., Elliott A., Sargatal J., Christie D.A. de Juana E., editors. Handbook of the birds of the world alive. Barcelona: Lynx Edicions. Retrieved from http://www.hbw.com/node/61823 on 9 January 2016.

Hosner P.A., Faircloth B.C., Glenn T.C., Braun E.L., Kimball R.T. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). Mol. Biol. Evol. doi: 10.1093/molbev/msv347.

Katoh K., Standley D.M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. Mol. Biol. Evol. 30:772–780.

Klicka J., Burns K., Spellman G.M. 2007. Defining a monophyletic Cardinalini: A molecular perspective. Mol. Phylogenet. Evol. 45:1014–1032.

Lanier H.C., Huang H., Knowles L.L. 2014. How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. Mol. Phylogenet. Evol. 70:112–119.

Leaché A.D., Chavez A.S., Jones L.N., Grummer J.A., Gottscho A.D., Linkem C.W. 2015. Phylogenomics of phrynosomatid lizards: Conflicting signals from sequence capture versus restriction site associated DNA sequencing. Genome Biol. Evol. 7:706–719.

Liu L., Yu L., Kubatko L., Pearl D.K., Edwards S.V. 2009. Coalescent methods for estimating phylogenetic trees. Mol. Phylogenet. Evol. 53:320–328.

Liu L., Yu L., Edwards S.V. 2010. A maximum pseudo-likelihood approach for estimating species trees under the coalescent model. BMC Evol. Biol. 10:302.

Liu L., Yu L. 2011. Estimating species trees from unrooted gene trees. Syst. Biol. 60:661–667.

Liu L., Xi Z., Wu S., Davis C., Edwards S.V. 2015. Estimating phylogenetic trees from genome-scale data. arXiv preprint arXiv:1501.03578.

Manthey J.D., Moyle R.G. 2015. Isolation by environment in White-breasted Nuthatches (*Sitta carolinensis*) of the Madrean Archipelago sky islands: A landscape genomics approach. Mol. Ecol. 24: 3628–3638.

McCormack J.E., Maley J.M., Hird S.M., Derryberry E.P., Graves G.R., Brumfield R.T. 2012. Next-generation sequencing reveals phylogeographic structure and a species tree for recent bird divergences. Mol. Phylogenet. Evol. 62:397–406.

McCormack J.E., Harvey M.G., Faircloth B.C., Crawford N.G., Glenn T.C., Brumfield R.T. 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. PLoS One 8:e54848.

McCormack J.E., Tsai W.L., Faircloth B.C. 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Res. doi: 10.1111/1755-0998.12466.

Miller M.R., Dunham J.P., Amores A., Cresko W.A., Johnson E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17:240–248.

Mirarab S., Reaz R., Bayzid M.S., Zimmermann T., Swenson M.S., Warnow T. 2014. ASTRAL: Genome-scale coalescent-based species tree estimation. Bioinformatics 30:i541–i548.

Nylander J.A., Wilgenbusch J.C., Warren D.L., Swofford D.L. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583.

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

R Development Core Team. 2012. R: A language for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rambaut A., Drummond, A.J. 2007. http://beast.bio.ed.ac.uk/Tracer.

Rubin B.E., Ree R.H., Moreau C.S. 2012. Inferring phylogenies from RAD sequence data. PLoS One 7:e33394.

Seo T.K. 2008. Calculating bootstrap probabilities of phylogeny using multilocus sequence data. Mol. Biol. Evol. 25:960–971.

Shepherd T.M., Burns K.J. 2007. Intraspecific genetic analysis of the summer tanager *Piranga rubra*: Implications for species limits and conservation. J. Avian Biol. 38:13–27.

Smith B.T., Harvey M.G., Faircloth B.C., Glenn T.C., Brumfield R.T. 2013. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63:83–95.

Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

Swofford D.L. 2003. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Version 4.

Wiley E.O., Lieberman B.S. 1981. Phylogenetics: Theory and practice of phylogenetic systematics. In: Species and speciation. 2nd ed. p. 23–65.

Zellmer A.J., Hanes M.M., Hird S.M., Carstens B.C. 2012. Deep phylogeographic structure and environmental differentiation in the carnivorous plant *Sarracenia alata*. Syst. Biol. 61:763–777.