



RESEARCH ARTICLE

Target enrichment of thousands of ultraconserved elements sheds new light on early relationships within New World sparrows (Aves: Passerellidae)

Robert W. Bryson, Jr.,^{1*} Brant C. Faircloth,² Whitney L. E. Tsai,³ John E. McCormack,³ and John Klicka¹

¹ Department of Biology and Burke Museum of Natural History and Culture, University of Washington, Seattle, Washington, USA

² Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, Louisiana, USA

³ Moore Laboratory of Zoology, Occidental College, Los Angeles, California, USA

* Corresponding author: brysonjr@uw.edu

Submitted January 28, 2016; Accepted March 16, 2016; Published May 25, 2016

ABSTRACT

Sparrows in the nine-primaried oscine family Passerellidae represent an attractive model for studying avian diversification across North and South America. However, the lack of phylogenetic resolution at the base of the New World sparrow tree has hampered the use of the existing sparrow phylogeny to test questions about the evolution of sparrow traits. We generated phylogenomic data from 1,063 ultraconserved elements to estimate phylogenetic relationships among the major clades of New World sparrows. Concatenated and species-tree analyses of 271,830 base pairs of sequence data converged on a well-supported phylogeny that differs from previous estimates. The resolved backbone of the sparrow phylogeny provides new insight into the biogeography of this radiation by suggesting both a tumultuous biogeographic history, with many colonizations of South America, and several independent ecological transitions to different habitat types.

Keywords: nine-primaried oscines, phylogenetics, species tree, UCEs

El enriquecimiento dirigido de miles de elementos ultra conservados brinda una nueva mirada sobre la relación temprana adentro de los gorriones del Nuevo Mundo (Aves: Passerellidae)

RESUMEN

Los gorriones de nueve primarias de la familia de oscinos Passerellidae representan un modelo atractivo para el estudio de la diversificación de las aves a través de América del Norte y del Sur. Sin embargo, la falta de resolución filogenética en la base del árbol de los gorriones del Nuevo Mundo ha obstaculizado el uso de la filogenia existente de los gorriones para analizar la evolución de los rasgos de los gorriones. En este estudio generamos datos filogenéticos a partir de 1,063 elementos ultra conservados para estimar las relaciones filogenéticas entre los clados principales de los gorriones del Nuevo Mundo. Los análisis concatenados y del árbol de especies de 271,830 pares de bases de datos de secuencias convergieron en una filogenia robusta que difiere de las estimaciones previas. La resolución de la columna vertebral de la filogenia de los gorriones brinda una nueva mirada sobre la biogeografía de esta radiación, sugiriendo tanto una historia biogeográfica tumultuosa con muchas colonizaciones de América del Sur, como varias transiciones ecológicas independientes hacia diferentes tipos de hábitat.

Palabras clave: árbol de especies, filogenias, oscinos de nueve primarias

INTRODUCTION

Sparrows in the nine-primaried oscine family Passerellidae are one of the largest songbird radiations in the New World. The family consists of 26 genera and 129 species distributed across North and South America (Barker et al. 2013, Klicka et al. 2014, Slager and Klicka 2014). Although most species inhabit temperate grasslands in the Americas, New World sparrows also occupy arid tropical deciduous thornscrub, montane forests, and a variety of other habitat types (Klicka and Spellman 2007, Rising et al. 2011). Many temperate

species are long-distance migrants that winter in the Neotropics, and the repeated evolution of seasonal shifts in distribution among Passerellidae may have led to the high species diversity within the family (Winger et al. 2014).

The broad distribution of New World sparrows makes them an attractive model for studying the relative impacts of geological events, paleoclimate change, migration, and ecological shifts on avian diversification across North and South America. However, the lack of phylogenetic resolution at the base of the New World sparrow tree has hampered these types of investigations (Klicka et al. 2014).

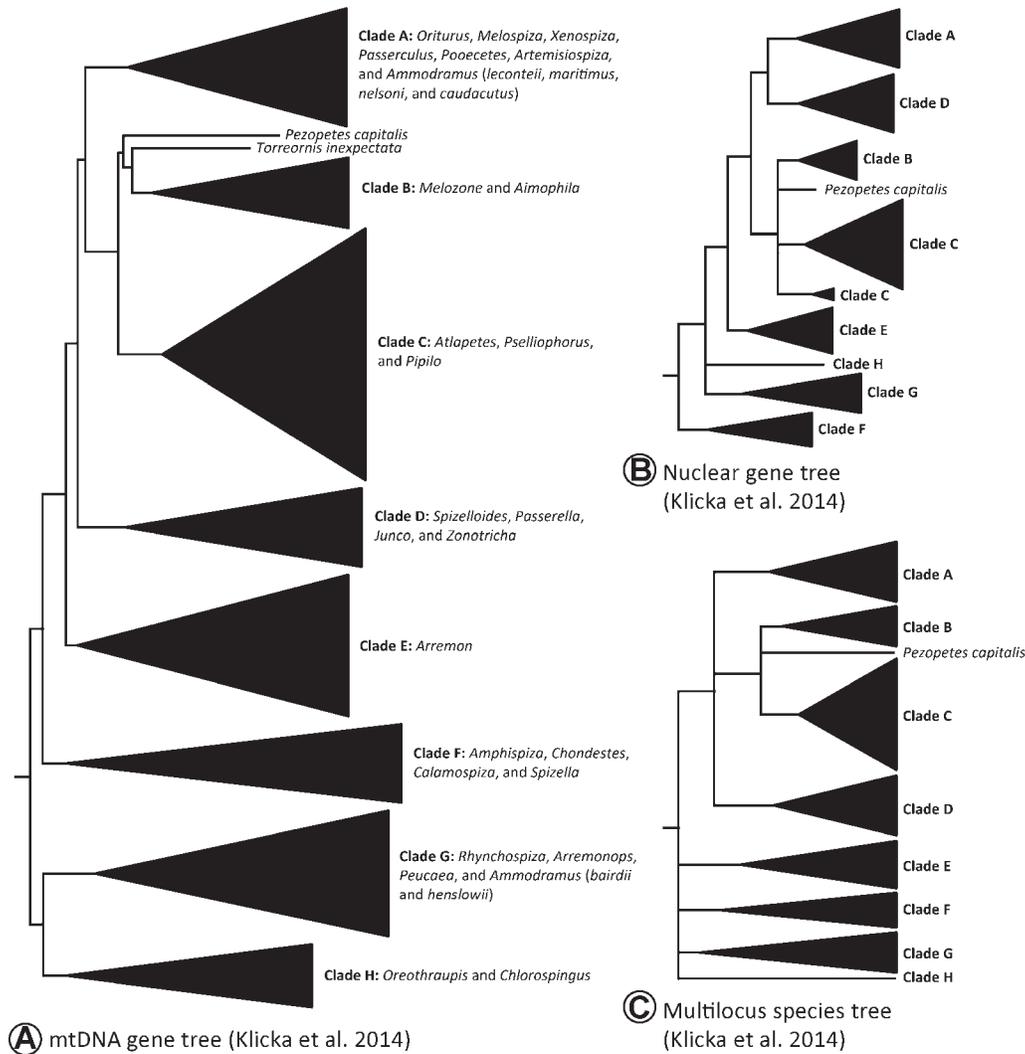


FIGURE 1. Phylogenetic hypotheses for New World sparrows from Klicka et al. (2014). **(A)** “Best” estimate of phylogenetic relationships based on complete species-level sampling and 2,184 bp of mtDNA. **(B)** Phylogenetic relationships based on 5,705 bp of DNA from 4 nuclear genes. **(C)** Multilocus species tree reconstructed using 7,889 bp from 2 mtDNA and 4 nuclear genes. In all trees, poorly supported nodes (<70 bootstrap, <0.95 Bayesian posterior probability) are collapsed.

Phylogenetic analyses of multilocus data from all described species suggest that nearly all sparrows can be placed within 8 well-supported clades (Klicka et al. 2014; Figure 1). The relationships among these clades appear to be well resolved on the basis of mitochondrial DNA (mtDNA; Figure 1A). In contrast, multilocus data revealed poorly supported relationships, especially among the older sparrow lineages (Figure 1B, 1C). From these results, Klicka et al. (2014) concluded that the handful of genes used (2 mtDNA and 5 nuclear) were insufficiently informative to disentangle questions of sparrow relationships.

The advent of large-scale phylogenomic datasets has transformed the field of phylogenetics. It is now common for researchers to sequence hundreds to thousands of loci in nonmodel taxa. Target enrichment of ultraconserved

elements is one of several next-generation sequencing approaches (Faircloth et al. 2012) that can produce large DNA datasets. Ultraconserved elements (UCEs) are a class of highly conserved and abundant nuclear markers found throughout the genomes of a variety of animals (Bejerano et al. 2004, Stephen et al. 2008, Faircloth et al. 2012, 2013, 2015). Sequence data from UCEs are relatively easy to obtain and, together with DNA adjacent to UCE locations (flanking DNA), can resolve both shallow-level (<5 mya; e.g., Smith et al. 2014) and deep-level (>100 mya; e.g., McCormack et al. 2013, Faircloth et al. 2015) relationships.

Here, we use phylogenomic data from >1,000 UCEs to infer phylogenetic relationships among the major clades of New World sparrows. We expected that the 100-fold increase in loci over the 7 genes used in Klicka et al. (2014)

TABLE 1. Taxa used in this study, with specimen source, locality information, and clade membership. Clade designations follow Klicka et al. (2014), as shown in Figure 1. Institution abbreviations: JFBM = James Ford Bell Museum of Natural History; LSUMNS = Louisiana State University Museum of Natural Science; MVZ = Museum of Vertebrate Zoology, University of California, Berkeley; and UWBM = University of Washington, Burke Museum.

Species	Specimen [tissue]	Locality	Clade
<i>Aimophila rufescens</i>	UWBM 93611 [DHB 3531]	Honduras: Copan	B
<i>Ammodramus leconteii</i>	UWBM 80599 [CSW 6816]	USA: North Dakota	A
<i>A. savannarum</i>	JFBM 43678 [JK 94-056]	USA: Montana	G
<i>Amphispiza bilineata</i>	UWBM 96901 [TKA 098]	USA: Nevada	F
<i>Arremonops aurantirostris</i>	UWBM 93635 [JK 00-053]	Honduras: Copan	E
<i>A. brunneinucha</i>	UWBM 69340 [DAB 1706]	Nicaragua: Granada	E
<i>Arremonops rufivirgatus</i>	JFBM 43282 [X6828]	USA: Texas	G
<i>Artemisospiza belli</i>	MVZ 170007 [20306]	USA: California	A
<i>Atlappetes citrinellus</i>	UWBM 92260 [B8332]	Argentina: Tucuman	C
<i>A. gutturalis</i>	UWBM 93636 [GAV 1374]	Honduras: Copan	C
<i>Calamospiza melanocorys</i>	JFBM 43748 [JK 94-058]	USA: Montana	F
<i>Chlorospingus ophthalmicus</i>	UWBM 114696 [JK 99-074]	Honduras: Copan	H
<i>Chondestes grammacus</i>	UWBM 103584 [DHB 3910]	USA: Arizona	F
<i>Junco phaeonotus</i>	UWBM 107920 [GMS 596]	Mexico: Jalisco	D
<i>Melospiza melodia</i>	JFBM 43741 [JK 94-084]	USA: Montana	A
<i>Melospiza aberti</i>	UWBM 98275 [DHB 2352]	USA: Nevada	B
<i>M. leucotis</i>	UWBM 109906 [JK 02-053]	Guatemala: Quetzaltenango	B
<i>Oriturus superciliosus</i>	FMNH 343328	Mexico: Morelos	A
<i>Passerculus sandwichensis</i>	UWBM 110768 [JK 03-341]	Canada: Newfoundland	A
<i>Passerella iliaca</i>	JFBM [DNA extract FOSP91]	USA: California	D
<i>Peucaea ruficauda</i>	UWBM 69323 [DAB 1680]	Nicaragua: Rivas	G
<i>Pezopetes capitalis</i>	LSUMZ 138633 [B-16232]	Costa Rica: San Jose	-
<i>Pipilo chlorurus</i>	UWBM 77977 [VGR 711]	USA: Arizona	C
<i>Pooecetes gramineus</i>	JFBM 43743 [JK 94-063]	USA: Montana	A
<i>Rhynchospora strigiceps</i>	UWBM 92416 [DHB 2425]	Argentina: Salta	G
<i>Spizella atrogularis</i>	UWBM 109557 [JK 01-210]	USA: Nevada	F
<i>Spizelloides arborea</i>	UWBM 66284 [GKD 126]	USA: Alaska	D
<i>Zonotrichia atricapilla</i>	UWBM 80383 [EVL 740]	USA: Alaska	D
<i>Geothlypis trichas</i>	UWBM 77569 [EVL 782]	USA: Arizona	Outgroup
<i>Icterus gularis</i>	UWBM 114470 [JK 11-007]	Mexico: Chiapas	Outgroup

would better resolve relationships within the New World sparrow tree and, in doing so, provide researchers with the evolutionary framework needed to understand the evolution of these songbirds across North and South America.

METHODS

Taxon Sampling

We collected tissues from 28 individuals representing all 8 major clades of New World sparrows (Klicka et al. 2014; Table 1) and the monotypic species *Pezopetes capitalis*. Seven clades were represented by ≥ 2 samples. We also included 2 samples, *Geothlypis trichas* and *Icterus gularis*, to serve as outgroups (Barker et al. 2013, Klicka et al. 2014), and we harvested UCE data from the genome sequence of *Geospiza fortis* (described below), which we used as an additional set of outgroup loci.

Library Preparation, UCE Target Enrichment, and Sequencing

We extracted genomic DNA from vouchered bird tissues using Qiagen DNeasy Blood & Tissue Kits (Qiagen, Valencia,

California, USA), and we prepared sequencing libraries for Illumina sequencing using Nextera XT Library Preparation Kits (Illumina, San Diego, California; Adey et al. 2010). We generally followed the sequence-capture workflow for Nextera-prepared libraries detailed in Faircloth et al. (2012), but we modified this protocol by using one-quarter aliquots of the required Nextera kit reagents and genomic DNA, and we pooled indexed libraries into groups of 8 libraries, prior to enrichment. We enriched each pool using a set of 5,472 custom-designed probes (MYbaits; MYcroarray, Ann Arbor, Michigan, USA) targeting 5,060 UCE loci (Faircloth et al. 2012) following an open-source protocol (for the full protocol, see <http://www.ultraconserved.org>). Following enrichment and an 18-cycle PCR-recovery, we combined pooled libraries in equimolar ratios to a final concentration of 10 μ M, and we sent the pooled libraries to the UCLA Genoseq Facility for 250 bp (base pair) paired-end sequencing using 2 runs of an Illumina MiSeq.

Data Extraction from *Geospiza fortis*

In addition to the data we collected from *Geothlypis trichas* and *Icterus gularis*, we harvested UCE loci from the existing

Geospiza fortis genome assembly to serve as a third set of outgroup data. We downloaded a *G. fortis* assembly from the UCSC Genome Browser (geoFor1; downloaded August 16, 2012), and we identified areas in the genome assembly matching the UCE probe set with the “run_multiple_lastzs_sqlite” program that is part of PHYLUCÉ (Faircloth 2016). After identifying UCE loci, we sliced FASTA sequences from around each UCE locus (± 500 bp) using the “slice_sequence_from_genomes” program, and we screened the resulting FASTA sequences for putative paralogs using the “match_contigs_to_probes” program. We input these loci to the taxon-set creation, described below.

Sequence Data Processing

After sequencing, we followed the standard PHYLUCÉ pipeline for processing target-enriched UCE data (Faircloth 2016). Briefly, we trimmed reads for adapter contamination and low-quality bases, and we assembled reads using Trinity (trinityrnaseq-r2013-02-25; Grabherr et al. 2011). We screened the resulting assemblies for those contigs matching enriched UCE loci using the “match_contigs_to_probes” program, and we discarded putative paralogs using the standard PHYLUCÉ algorithm. We created a taxon set containing our ingroup and outgroup taxa (including the loci harvested from *G. fortis*), and we used this taxon set to create a FASTA file containing all data for all taxa. We input this FASTA file to a program that parallelizes MAFFT (Katoh and Standley 2013) alignments across all loci, turning off the standard PHYLUCÉ algorithm for alignment trimming and outputting the resulting alignments in FASTA format. We trimmed the alignments with a parallelized wrapper around Gblocks (Castresana 2000, Talavera and Castresana 2007). Following alignment, we removed locus names from each alignment, and we created a 75% complete dataset (each locus contains data for 75% of the total number of individuals). We concatenated alignments in the 75% complete dataset using a helper program from the PHYLUCÉ package, and we output these concatenated alignments in PHYLIP format for downstream use.

Phylogenetic Analyses

We performed unpartitioned concatenated maximum-likelihood (ML) analyses of the 75% complete dataset using RAxML 8.0.19 (Stamatakis 2014) by conducting 20 searches of the data with the GTR GAMMA site-rate substitution model for the best ML tree. We assessed support for the best ML topology by performing nonparametric bootstrapping using the autoMRE option in RAxML with the GTR GAMMA site-rate substitution model, and we reconciled the best ML tree with the bootstrap replicates using RAxML. We performed unpartitioned concatenated Bayesian analyses of the 75% complete dataset with ExaBayes 1.2.1 (Aberer et al. 2014) using 4 independent runs of 1 million iterations with

heated chains turned off. We assessed convergence of the posterior distribution using the default thinning for ExaBayes (500 iterations) after discarding 25% of the posterior sample as “burn-in” by (1) computing the standard deviation of split frequencies, (2) examining the ESS (effective sample size) and potential scale reduction factor (PSRF) values of the parameter estimates output by the “postProcParam” ExaBayes helper program, and (3) visually assessing posterior trace distributions using Tracer 1.6.0 (Rambaut et al. 2014). We computed the majority rules extended (MRE) consensus of the thinned, post-burn-in posterior distribution using the “consense” helper program that is part of ExaBayes.

We also analyzed a subset of the loci in the 75% complete dataset using a gene-tree-based coalescent approach. Although coalescent approaches offer a promising alternative to concatenated analyses, which can mislead inferences under certain conditions (Kubatko and Degnan 2007, Liu and Pearl 2007, Linkem et al. 2016), coalescent methods that are applicable to phylogenomic datasets can be negatively affected by poorly resolved gene trees (Gatesy and Springer 2014, Lanier et al. 2014, Xi et al. 2015, Meiklejohn et al. 2016). Poorly resolved gene trees often result when alignments input to tree-inference programs contain reduced phylogenetic signal—one characteristic of conserved loci, particularly when the phylogenetic analysis spans relatively recent divergence times. To address this potential problem, several filtering approaches have been proposed to exclude loci that may bias species-tree analyses (e.g., Harris et al. 2014, Hosner et al. 2016, Meiklejohn et al. 2016). Here, we computed the number of parsimony-informative sites in each locus of our 75% complete dataset, and we created a subset of loci having a number of parsimony-informative sites in the upper quartile of the range. We sampled this reduced dataset by loci, with replacement, to create 100 resampled datasets, and we performed nonparametric bootstrapping by sites of each locus in each dataset using RAxML with a GTR GAMMA model of site-rate substitution. After running the gene-tree estimation procedure, we sorted the bootstrapped gene trees to ensure that each resampled dataset was resampled by loci and by sites (Seo 2008). We input the 100 resampled datasets to ASTRAL-II 4.7.8 (Mirarab and Warnow 2015), which is a summary-based species-tree-inference program that is reasonably robust to the effects of incorrectly estimated gene trees (Simmons and Gatesy 2015, Meiklejohn et al. 2016). We computed the extended majority-rule consensus of the resulting species trees using RAxML, and we collapsed branches with bootstrap support values $< 70\%$.

RESULTS

Sequence Data

After trimming reads for adapter contamination and quality, we collected an average of 958,240 sequence reads

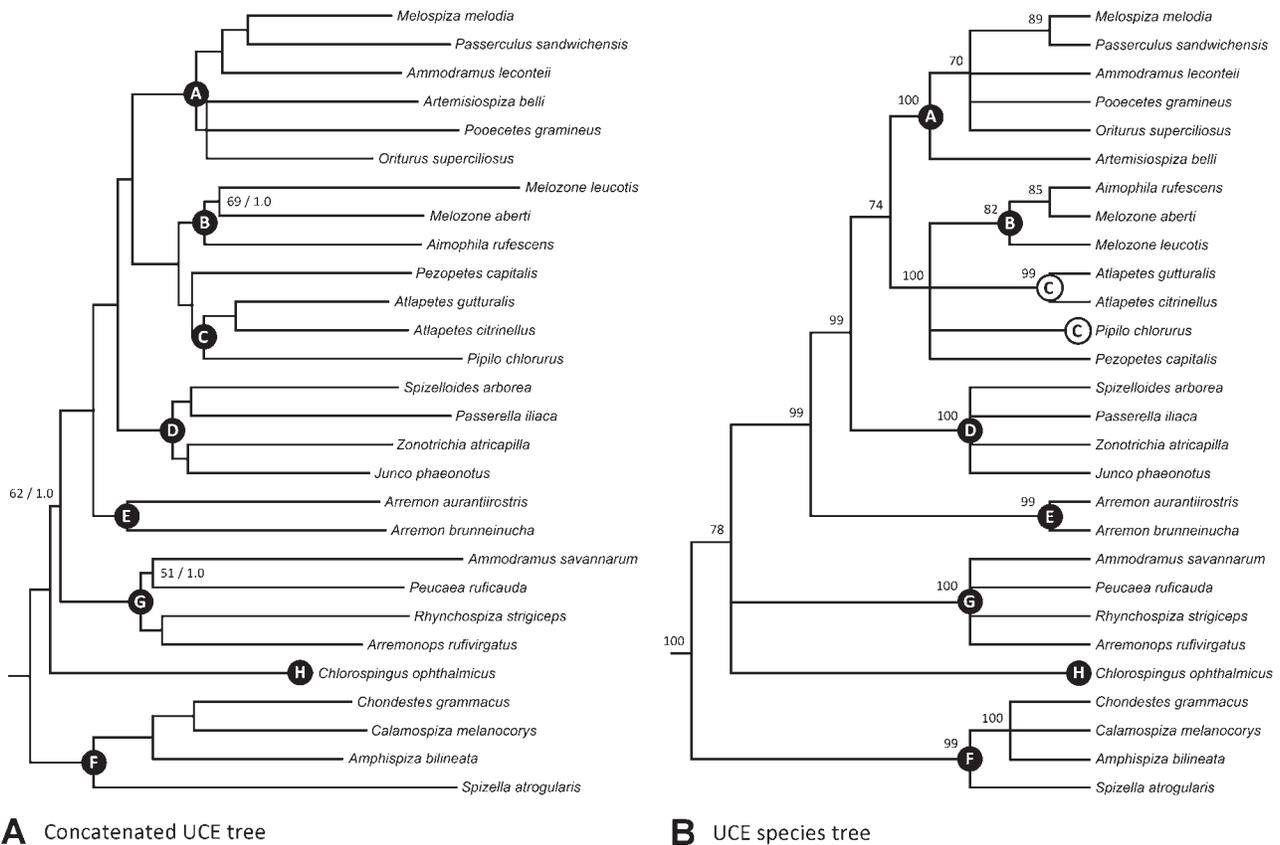


FIGURE 2. Phylogenetic relationships of New World sparrow clades estimated from 271,830 bp of UCE sequence data. Clade designations follow Klicka et al. (2014), as shown in Figure 1. **(A)** Maximum-likelihood tree of 1,063 concatenated UCE loci. A Bayesian analysis also supported the same tree topology. Support values are shown for nodes receiving $<70\%$ bootstrap but ≥ 0.95 posterior probability support (as bootstrap/posterior probability); nodes supported by $<70\%$ bootstrap and 0.95 posterior probability support are collapsed. **(B)** Species tree estimated from the gene trees of the 329 UCE loci with ≥ 3 parsimony-informative sites. Support values are shown for nodes receiving $\geq 70\%$ bootstrap support; nodes supported by $<70\%$ bootstrap support are collapsed.

(95% confidence interval [CI]: 140,934) from each library, with an average length of 136.6 bp (95% CI: 4.9 bp; [Supplemental Material Appendix A](#)), and we assembled these reads into an average of 32,245 contigs (95% CI: 4,110) with an average length of 258.5 bp (95% CI: 2.1 bp; [Supplemental Material Appendix B](#)). After UCE detection, we enriched an average of 1,403 contigs (95% CI: 95) representing UCE loci with an average length of 343 bp (95% CI: 8.5) and an average sequencing coverage of $6\times$ (95% CI: 0.4; [Supplemental Material Appendix C](#)). After taxon-set creation, FASTA extraction, alignment, and alignment trimming, the 75% complete dataset contained 1,063 loci with an average length of 256 bp (95% CI: 2.7 bp), an average of 27 taxa locus⁻¹ (95% CI: 0.2), and an average of 2.4 parsimony-informative sites locus⁻¹ (95% CI: 0.2). The concatenated supermatrix of the 75% complete dataset contained 271,830 bp, 36,816 alignment patterns, and 2,538 parsimony-informative sites. The upper quartile of parsimony-informative loci in the 75%

complete dataset were those having ≥ 3 parsimony-informative sites ($n = 329$).

Phylogenetic Estimation

Maximum-likelihood and Bayesian analyses of the concatenated UCE loci converged on a strongly supported topology (Figure 2A). Monophyly of the major clades (A–H) in Klicka et al. (2014) represented by ≥ 2 samples in our analyses was strongly supported, and only one node in the tree received $<70\%$ bootstrap and 0.95 posterior probability support. This node (collapsed in Figure 2A) weakly supported a sister relationship between *Artemisiospiza belli* and *Poocetes gramineus*. The species tree reconstructed from the screened subset of UCE loci was less well supported (Figure 2B) but generally matched the concatenated UCE topology. Ten nodes received $<70\%$ bootstrap support, mostly involving splits within clades. Phylogenetic relationships among clades were strongly supported, with one exception near the base of the tree. Monophyly of all

but one of the major sparrow clades represented by ≥ 2 samples in our analyses was strongly supported.

DISCUSSION

Phylogenetic analysis of 1,063 UCE loci and 271,830 bp of aligned sequence data produced a strongly supported backbone phylogeny for New World sparrows. Of the internal nodes defining clades A–H, only the placement of clade G was not well resolved (Figure 2). By contrast, the most reliable estimate of sparrow relationships in Klicka et al. (2014), based on mtDNA and few nuclear markers, yielded only 2 resolved internal nodes, and none of the focal clades (A–H) could be placed with confidence (Figure 1C). In Klicka et al. (2014), the problematic clade G was also unresolved in the mtDNA, nuclear DNA, and species-tree analyses.

The UCE phylogeny suggests that prior phylogenetic hypotheses based on mtDNA, as well as those based on small numbers of nuclear genes, might each be misleading in their own ways, with biologically interesting implications. Perhaps surprisingly, in one important instance the UCE tree supported a relationship indicated by the previous mtDNA tree rather than one indicated by the nuclear DNA tree. In Klicka et al. (2014), the mtDNA tree shows clade D as sister to the clade comprising clades A, B, and C (Figure 1A). This relationship contrasts with that suggested by the nuclear gene tree from the same study, which instead places clade D as sister to clade A (Figure 1B). The relationship between clades A and D is biologically interesting because these 2 clades comprise the grassland and brushland sparrows. Thus, the prior nuclear data support a single origin for this ecological trait and a degree of phylogenetic niche conservatism, whereas the mtDNA data support more ecological transitions through time. The UCE data strongly support the latter hypothesis, with the grassland and brushland sparrows not coming together as sister clades. Rather, each is a member of a broader group containing lower-latitude species like those in the speciose genus *Atlapetes*.

In another instance, the UCE tree conflicts with the prior mtDNA tree. At their bases, the UCE tree (Figure 2) and the prior nuclear gene tree (Figure 1B) both support a sister relationship between clade F and the remaining sparrow clades. The mtDNA tree instead strongly supports clades G and H as sister to the rest of the sparrows, with clade F nested well within the New World sparrow tree (Figure 1A). This discrepancy is difficult to explain without careful testing and could be caused by analytical and biological factors that affect organellar and nuclear DNA differently. Clarifying clade F as the sister to the rest of the sparrow phylogeny, however, is an important component of any future attempts to analyze the evolution of New World sparrow traits through time.

The UCE trees generated from concatenated and species-tree analyses agreed on branching order of the deep nodes, whereas the species tree was less resolved than the concatenated tree with respect to more shallow divergences. The concatenated UCE tree showed only a single polytomy within clade A (Figure 2A). The UCE species tree, meanwhile, showed 10 collapsed nodes scattered among 5 of the 8 core clades (Figure 2B). For the remaining (uncollapsed) nodes, the concatenated tree and species tree differed at a single node. The species tree suggested a sister relationship between *Aimophila rufescens* and *Melospiza aberti*; the concatenated UCE tree placed *M. aberti* and *M. leucotis* as sister species. The lack of resolution of the UCE species tree toward the tips of the tree could reflect the relatively lower informative content of UCEs for very shallow timescales, especially because our library preparation method led to shorter locus lengths (see below) than comparable studies carried out more recently (e.g., Giarla and Esselstyn 2015, McCormack et al. 2016).

A resolved hypothesis of basal relationships among sparrow clades has been lacking until now. The strong support along the “backbone” of the phylogeny allows new glimpses into the sparrow radiation in the New World. For example, the UCE phylogeny suggests a tumultuous biogeographic history. Sparrows are well represented in both North and South America. Each of the 8 core clades has representation in North America, and elements of 5 clades (C, D, E, G, and H) are represented in South America. Given a probable North American origin (Barker et al. 2015), the UCE tree suggests multiple independent colonizations of South America. As mentioned above, the UCE tree also provides insight into the origins of ecological diversity in sparrows. Grassland-dwelling sparrows, for example, do not form a clade, but instead occur in 3 phylogenetically distant clades (A, F, and G), suggesting more transitions between different habitat types (e.g., grassland, brushland, and forest) over evolutionary time than were previously suspected (e.g., Paynter 1964, Dickerman et al. 1967).

Detailed analyses that involve mapping distributional, behavioral, or phenotypic characters on the sparrow phylogeny await a tree with better taxonomic sampling. We generated phylogenomic data from only a small percentage (~22%) of sparrow species to infer relationships across the backbone of the tree. Future studies should fill in the phylogeny with additional UCE data from all passerellid species. Alternatively, our UCE data could be combined with the more densely sampled 5-locus dataset of Klicka et al. (2014) in a hybrid phylogenetic–phylogenomic approach (Leaché et al. 2014). This hybrid approach could produce a strongly supported species-level phylogeny of New World sparrows amenable to biogeographic and character-state reconstructions without the need for further data collection.

Finally, testing, optimizing, and reducing costs of laboratory methods for target enrichment of UCEs has been a priority since we began using the technique to infer phylogeny, and library preparation is the single most expensive step, on a per sample basis, of the target enrichment process. Here, we prepared libraries using the Illumina Nextera enzyme in one-quarter reactions to reduce costs. Although we knew that the Nextera enzyme produces shorter insert sizes, on average, than random shearing (cf. Crawford et al. 2012, 2015), we found the reduced contig sizes an acceptable tradeoff, given the increased efficiency during library preparation that the Nextera system offers. However, dilution of Illumina Nextera XT transposase was inconsistent in our hands, particularly compared to earlier, similar dilutions of Epicentre Nextera transposase, which worked well (Crawford et al. 2012, Faircloth et al. 2012, McCormack et al. 2013). Dilution of the Nextera XT reagent appears to have affected this study by producing shorter contigs (expected) while also reducing the consistency of enrichment and sequencing coverage across enriched UCE loci (unexpected). Typically, we obtain longer average locus lengths (400–800 bp) and more consistent enrichment of UCE loci when we and others use alternative shearing and library preparation protocols (e.g., Giarla and Esselstyn 2015, McCormack et al. 2016). Nonetheless, we were still able to recover >1,000 UCE loci from 30 taxa in our 75% complete matrix despite the shorter contig lengths and inconsistent enrichment that we observed, demonstrating the power of target enrichment approaches over Sanger-based techniques even when the performance of the enrichment procedure is suboptimal.

ACKNOWLEDGMENTS

We thank the following curators, collection managers, and institutions for tissue samples critical for the completion of this study: F. Sheldon, R. Brumfield, and D. Dittmann (Louisiana State University Museum of Natural Science); N. Johnson and C. Cicero (University of California, Berkeley, Museum of Vertebrate Zoology); R. Zink (James Ford Bell Museum of Natural History); and S. Rowher and S. Birks (University of Washington, Burke Museum of Natural History). We also thank D. Crowley for help in the lab, and I. Lovette and an anonymous reviewer for suggestions that improved the manuscript.

Funding statement: This work was funded in part by National Science Foundation (NSF) grants DEB-0315469 and DEB-1257785 (to J.K.) and DEB-1258205 (to J.E.M.), and computational portions were supported by DEB-1242260 (to B.C.F.). Portions of this research were also conducted with high-performance computing resources provided by Louisiana State University (<http://www.hpc.lsu.edu>). None of the funders had any input into the content of the manuscript, nor

did they require manuscript approval before submission or publication.

Ethics statement: All federal, state, and local permits were secured prior to fieldwork.

Author contributions: (1) J.K. and J.E.M. conceived the idea; (2) R.W.B. and W.L.E.T. generated the data; (3) B.C.F. analyzed the data; (4) R.W.B., B.C.F., W.L.E.T., J.E.M., and J.K. wrote the paper; and (5) B.C.F., J.E.M., and J.K. contributed substantial materials, resources, and funding.

Data availability: Raw read data and assembled contigs are available from NCBI BioProject PRJNA318526. Aligned sequences are available at 10.6084/m9.figshare.3101359.

LITERATURE CITED

- Aberer, A. J., K. Kobert, and A. Stamatakis (2014). ExaBayes: Massively parallel Bayesian tree inference for the whole-genome era. *Molecular Biology and Evolution* 31:2553–2556.
- Adey, A., H. G. Morrison, Asan, X. Xun, J. O. Kitzman, E. H. Turner, B. Stackhouse, A. P. MacKenzie, N. C. Caruccio, X. Zhang, and J. Shendure (2010). Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biology* 11:R119.
- Barker, F. K., K. J. Burns, J. Klicka, S. M. Lanyon, and I. J. Lovette (2013). Going to extremes: Contrasting rates of diversification in a recent radiation of New World passerine birds. *Systematic Biology* 62:298–320.
- Barker, F. K., K. J. Burns, J. Klicka, S. M. Lanyon, and I. J. Lovette (2015). New insights into New World biogeography: An integrated view from the phylogeny of blackbirds, cardinals, sparrows, tanagers, warblers, and allies. *The Auk: Ornithological Applications* 132:333–348.
- Bejerano, G., M. Pheasant, I. Makunin, S. Stephen, W. J. Kent, J. S. Mattick, and D. Haussler (2004). Ultraconserved elements in the human genome. *Science* 304:1321–1325.
- Castresana, J. (2000). Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution* 17:540–552.
- Crawford, N. G., B. C. Faircloth, J. E. McCormack, R. T. Brumfield, K. Winker, and T. C. Glenn (2012). More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. *Biology Letters* 8:783–786.
- Crawford, N. G., J. F. Parham, A. B. Sellas, B. C. Faircloth, T. C. Glenn, T. J. Papenfuss, J. B. Henderson, M. H. Hansen, and W. B. Simison (2015). A phylogenomic analysis of turtles. *Molecular Phylogenetics and Evolution* 83:250–257.
- Dickerman, R. W., A. R. Phillips, and D. W. Warner (1967). On the Sierra Madre Sparrow, *Xenospiza baileyi*, of Mexico. *The Auk* 84:49–60.
- Faircloth, B. C. (2016). PHYLUCe is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32:786–788.
- Faircloth, B. C., M. G. Branstetter, N. D. White, and S. G. Brady (2015). Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Molecular Ecology Resources* 15:489–501.
- Faircloth, B. C., J. E. McCormack, N. G. Crawford, M. G. Harvey, R. T. Brumfield, and T. C. Glenn (2012). Ultraconserved elements

- anchor thousands of genetic markers spanning multiple evolutionary timescales. *Systematic Biology* 61:717–726.
- Faircloth, B. C., L. Sorenson, F. Santini, and M. E. Alfaro (2013). A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). *PLOS One* 8:e65923. doi:10.1371/journal.pone.0065923
- Gatesy, J., and M. S. Springer (2014). Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. *Molecular Phylogenetics and Evolution* 80:231–266.
- Giarla, T. C., and J. A. Esselstyn (2015). The challenges of resolving a rapid, recent radiation: Empirical and simulated phylogenomics of Philippine shrews. *Systematic Biology* 64:727–740.
- Grabherr, M. G., B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* 29:644–652.
- Harris, R. B., M. D. Carling, and I. J. Lovette (2014). The influence of sampling design on species tree inference: A new relationship for the New World chickadees (Aves: *Poecile*). *Evolution* 68:501–513.
- Hosner, P. A., B. C. Faircloth, T. C. Glenn, E. L. Braun, and R. T. Kimball (2016). Avoiding missing data biases in phylogenomic inference: An empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* 33:1110–1125.
- Katoh, K., and D. M. Standley (2013). MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular Biology and Evolution* 30:772–780.
- Klicka, J., F. K. Barker, K. J. Burns, S. M. Lanyon, I. J. Lovette, J. A. Chaves, and R. W. Bryson, Jr. (2014). A comprehensive multilocus assessment of sparrow (Aves: Passerellidae) relationships. *Molecular Phylogenetics and Evolution* 77:177–182.
- Klicka, J., and G. M. Spellman (2007). A molecular evaluation of the North American “grassland” sparrow clade. *The Auk* 124:537–551.
- Kubatko, L. S., and J. Degnan (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56:17–24.
- Lanier, H. C., H. Huang, and L. L. Knowles (2014). How low can you go? The effects of mutation rate on the accuracy of species-tree estimation. *Molecular Phylogenetics and Evolution* 70:112–119.
- Leaché, A. D., P. Wagner, C. W. Linkem, W. Böhme, T. J. Papenfuss, R. A. Chong, B. R. Lavin, A. M. Bauer, S. V. Nielsen, E. Greenbaum, M.-O. Rödel, et al. (2014). A hybrid phylogenetic–phylogenomic approach for species tree estimation in African *Agama* lizards with applications to biogeography, character evolution, and diversification. *Molecular Phylogenetics and Evolution* 79:215–230.
- Linkem, C. W., V. N. Minin, and A. D. Leaché (2016). Detecting the anomaly zone in species trees and evidence for a misleading signal in higher-level skink phylogeny (Squamata: Scincidae). *Systematic Biology* 65. In press.
- McCormack, J. E., M. G. Harvey, B. C. Faircloth, N. G. Crawford, T. C. Glenn, and R. T. Brumfield (2013). A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLOS One* 8:e54848. doi:10.1371/journal.pone.0054848
- McCormack, J. E., W. L. E. Tsai, and B. C. Faircloth (2016). Sequence capture of ultraconserved elements from bird museum specimens. *Molecular Ecology Resources* 16. doi:10.1111/1755-0998.12466
- Meiklejohn, K. A., B. C. Faircloth, T. C. Glenn, R. T. Kimball, and E. L. Braun (2016). Analysis of a rapid evolutionary radiation using ultraconserved elements (UCEs): Evidence for a bias in some multispecies coalescent methods. *Systematic Biology* 65. In press.
- Mirarab, S., and T. Warnow (2015). ASTRAL-II: Coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* 31:i44–i52.
- Paynter, R. A., Jr. (1964). Generic limits of *Zonotrichia*. *The Condor* 66:277–281.
- Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond (2014). Tracer 1.6. <http://beast.bio.ed.ac.uk/Tracer>
- Rising, J., A. Jaramillo, J. L. Copete, S. Madge, and P. Ryan (2011). Family Emberizidae (Buntings and New World Sparrows). In *Handbook of the Birds of the World, vol. 16: Tanagers to New World Blackbirds* (J. del Hoyo, A. Elliott, and D. A. Christie, Editors). Lynx Edicions, Barcelona, Spain. pp. 428–683.
- Seo, T.-K. (2008). Calculating bootstrap probabilities of phylogeny using multilocus sequence data. *Molecular Biology and Evolution* 25:960–971.
- Simmons, M. P., and J. Gatesy (2015). Coalescence vs. concatenation: Sophisticated analyses vs. first principles applied to rooting the angiosperms. *Molecular Phylogenetics and Evolution* 91:98–122.
- Slager, D. L., and J. Klicka (2014). A new genus for the American Tree Sparrow (Aves: Passeriformes: Passerellidae). *Zootaxa* 3821:398–400.
- Smith, B. T., M. G. Harvey, B. C. Faircloth, T. C. Glenn, and R. T. Brumfield (2014). Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. *Systematic Biology* 63:83–95.
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313.
- Stephen, S., M. Pheasant, I. V. Makunin, and J. S. Mattick (2008). Large-scale appearance of ultraconserved elements in tetrapod genomes and slowdown of the molecular clock. *Molecular Biology and Evolution* 25:402–408.
- Talavera, G., and J. Castresana (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Systematic Biology* 56:564–577.
- Winger, B. M., F. K. Barker, and R. H. Ree (2014). Temperate origins of long-distance seasonal migration in New World songbirds. *Proceedings of the National Academy of Sciences USA* 111:12115–12120.
- Xi, Z., L. Liu, and C. C. Davis (2015). Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. *Molecular Phylogenetics and Evolution* 92:63–71.