Molecular Phylogenetics and Evolution 202 (2025) 108235

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

A phylogenomic tree of wood-warblers (Aves: Parulidae): Dealing with good, bad, and ugly samples

Min Zhao [a,b,1], Jessica A. Oswald [b,c,1], Julie M. Allen [d], Hannah L. Owens [b,e], Peter A. Hosner [e,f], Robert P. Guralnick [b], Edward L. Braun [a], Rebecca T. Kimball [a,*]

[a] Department of Biology, University of Florida, Gainesville, FL 32611, USA
[b] Florida Museum of Natural History, University of Florida, Gainesville, FL 32611, USA
[c] U.S. Fish and Wildlife Service, National Fish and Wildlife Forensic Laboratory, Ashland, OR 97520, USA
[d] Department of Biological Sciences, Virginia Tech, Blacksburg, VA 24060, USA
[e] Center for Global Mountain Biodiversity, Section for Biodiversity, Globe Institute, University of Copenhagen, København Ø, Denmark
[f] Natural History Museum Denmark, University of Copenhagen, København Ø, Denmark

ARTICLE INFO

Keywords:
Ultraconserved elements
Phylogenomics
Historical specimens
Wood warblers

ABSTRACT

The New World warblers (Parulidae) are a model group for ecological and evolutionary analyses. However, current phylogenetic relationships across this family are based upon few loci. Here we use ultraconserved elements (UCEs) to estimate a rigorous species-level phylogeny for the family. As is true for many groups, high-quality tissues were unavailable for some taxa. Thus, we explored methods for incorporating sequences derived from historical (toe pad) samples to expand the phylogenetic datasets. We recovered an average of 4,186 UCE loci and mitochondrial bycatch data (supplemented with published mitochondrial data) from 96% of all currently recognized species. We found that the UCE phylogeny built with alignments with less than 70% of gaps and ambiguities recovered the most robust phylogenetic relationships for this family, representing 101 species. Using this phylogeny as a topological backbone and adding ten fair quality "bad" samples effectively generated an overall well supported phylogeny, representing 108 species (~90% of all species). Based on this tree, we then added in seven poor quality "ugly" samples and six of those were placed within their expected genera. We also explored the phylogenetic positions of the likely extinct *Leucopeza semperi* and the endangered *Catharopeza bishopi* where limited data was obtained. Overall, taxonomic placements in our UCE trees largely correspond to previously published studies with the recovery of all currently recognized genera as monophyletic except for *Basileuterus* which was rendered paraphyletic by *B. lachrymosus*. Our study provides insights in understanding the phylogenetic relationships of a model Passeriformes family and outlines effective practices for managing sparse genomic data sourced from historical museum specimens. Variable topological arrangements across datasets and analyses reflect the evolutionary complexity of this group and provide future topics for in-depth studies.

1. Introduction

The New World warblers (Parulidae) have long served as a model group in ecology and evolution, being used to address a wide range of questions (e.g., Baiz et al., 2021; Leroy et al., 2024; MacArthur, 1958; Mitchell et al., 2019; Simpson et al., 2020, 2021; Winger et al., 2012). The Parulidae are found across the Americas and the Caribbean in boreal to desert to tropical habitats, with species exhibiting extensive ranges (e.g., Yellow-rumped Warbler, *Setophaga coronata*) to single island endemics (e.g., Barbuda Warbler, *Setophaga subita*). Aside from their conserved small body size and thin, insect-gleaning bill, species vary greatly in traits and behavioral characteristics that have made them a focal lineage for the investigation of an array of topics ranging from, foraging ecology and seasonal migratory behavior to coloration patterns and vocal evolution, among others. The Parulidae feature many common and familiar species, but the clade also includes species that have gone extinct in the 20th century, including Bachman's Warbler, *Vermivora bachmanii*, which lived in bottomland hardwoods in the southeast United States and Cuba, and the monotypic Semper's Warbler, *Leucopeza semperi*, which was found in St. Lucia. Some range-restricted species are

* Corresponding author.
E-mail address: rkimball@ufl.edu (R.T. Kimball).
[1] Contributed equally to this work.

https://doi.org/10.1016/j.ympev.2024.108235
Received 28 March 2024; Received in revised form 26 October 2024; Accepted 10 November 2024
Available online 13 November 2024

threatened, such as the endangered Golden-cheeked Warbler *Setophaga chrysoparia* and Whistling Warbler *Catharopeza bishopi*. In addition to their relevance as a model system, an improved understanding of the group is important for species conservation.

Molecular phylogenies have been transformative in the understanding of the parulid evolution. Molecular data has altered the circumscription of what constitutes the family (e.g., Barker et al., 2013, 2015; Lovette and Bermingham, 2002), making it possible to identify a monophyletic Parulidae. Subsequent dense sampling within the Parulidae highlighted issues with generic limits (Lovette et al., 2010), leading to repeated taxonomic revisions (e.g., Chesser et al., 2011, 2019). Finally, detailed studies within species or species complexes (e.g., Barrera-Guzmán et al., 2012; Escalante et al., 2009; Gutiérrez-Pinto et al., 2012; Vilaça and Santos, 2010) have uncovered unappreciated species diversity (e.g., Chesser et al., 2016, 2018).

Until recently, DNA-based evolutionary hypotheses, such as those that have changed Parulidae taxonomy, have sourced a few nuclear introns and/or mitochondrial genes. These data may provide limited resolution or can be biased by one or a few genes (e.g., Kimball et al., 2013; Shen et al., 2017). Since many commonly used nuclear markers are too slowly evolving to elucidate relationships at shallow scales, some studies are effectively based primarily on mitochondrial data, and thus on a single locus. Over the past 15 years, growth of parallel sequencing technologies has facilitated the cost-efficient collection of thousands of loci across genomes (reviewed in Carter et al., 2023). Analysis of these large, genome-wide datasets, while confirming some previously hypothesized relationships, have also led to a myriad of well-supported novel relationships (e.g., Chen et al., 2018; Hosner et al., 2017; Zhao et al., 2023). Within vertebrates, one of the commonly used approaches is to sequence ultraconserved elements (UCEs) and their flanking regions, which can provide thousands of loci scattered throughout the nuclear genome (Faircloth et al., 2012). UCEs provide the power to investigate deep phylogenetic relationships across avian orders (McCormack et al., 2013), but can also provide sufficient information to elucidate shallow intraspecific divergences (e.g., Oswald et al., 2016; Smith et al., 2014). Thus, UCEs have provided insights among avian taxa with historically enigmatic relationships (Oliveros et al., 2019; Wang et al., 2017) and generated new phylogenetic hypotheses across taxonomic levels.

A limitation in many molecular phylogenetic studies is a lack of high-quality tissues suitable for sequencing and subsequent underrepresentation of relevant taxonomic units. Therefore, DNA is extracted from toe pads, skin punches or bone of historical specimens archived in museums (Tsai et al., 2020). The resultant historical DNA is low-concentration, fragmented, and often contaminated with exogenous content. While it is possible to collect UCEs from such samples (e.g., McCormack et al., 2016; Salter et al., 2022; Sun et al., 2014), poor quality historical DNA input can result in genomic datasets that vary significantly in locus recovery and alignment length (Smith et al., 2020). In some cases, heterogeneous data quantity and quality can affect phylogenetic estimation. For example, fragmented and gappy sequences may result in large amounts of missing data in the alignments that might increase the risk of systematic errors in phylogenetic inference. Gappy alignments can yield incorrect estimates of branch lengths and bias gene tree estimation, whereas missing taxa in certain locus alignments may bias the gene tree summary methods given that gene trees are only partially represented (Hosner et al., 2016; Sayyari et al., 2017; Simmons, 2014). Analyses of data generated from historical specimens can also lead to exceptionally long terminal branches that are not contributed by expedited evolutionary signals (e.g., Braun et al., 2024). Long branches may further be erroneously united together causing an artifact known as the long branch attraction (reviewed by Bergsten, 2005). However, the ideal analytical approaches for mitigating such problems are unclear (Gilbert et al., 2018; Smith et al., 2020).

The phylogenetic position of the Parulidae within the nine-primaried oscines (and other passerines) has been recently confirmed with phylogenomic data (Oliveros et al., 2019), but comprehensive species-level relationships within the Parulidae remain founded in mitochondrial and limited numbers of nuclear loci (Lovette et al., 2010). Here we sampled thousands of UCEs and complete mitogenomes from most species within the Parulidae to re-evaluate their phylogenetic relationships. However, available DNA was limited to older museum specimens for some taxa, necessitating exploration of methods to incorporate sample data that was variable in quality and quantity.

2. Methods

2.1. Taxon sampling

We collected 94 fresh tissue and 17 toe pad samples for 110 species (Table 1), which covers 91.7% of the species recognized by IOC World Bird List 13.1 (Gill et al., 2023) for the Parulidae family. The toe pad-only samples included two individuals of *Leucopeza semperi*, as it is likely extinct, otherwise a single toe pad represented a single species. These data were combined with available data to represent 96% of all parulid species (see 2.2 below).

Total genomic DNA from tissue samples were extracted using the Qiagen DNeasy Extraction Kit (Valencia, CA) following the manufacturer's protocol. DNA yield was quantified with a Qubit® 2.0 Fluorometer. DNA from toe pad samples was extracted using a phenol–chloroform extraction protocol (Tsai et al., 2020), but omitted SPRI bead cleanup. Extracted DNA samples were sent to Rapid Genomics (Gainesville, FL) for DNA library preparation, UCE enrichment (using the Tetrapods-UCE-5Kv1 probe set; available at ultraconserved.org), and sequencing according to the Faircloth et al. (2012) protocol.

2.2. Sequence assembly and alignment

Adapters were trimmed from raw sequence reads using illumiprocessor (Bolger et al., 2014; Faircloth, 2013). Duplicate reads were filtered using PRINSEQ-lite v0.20.4 (Schmieder and Edwards, 2011). The remaining reads were error-corrected with QuorUM v1.1.0 (Marçais et al., 2015). Reads were then assembled using SPAdes 3.15.0 (Bankevich et al., 2012) using the –sc option, to better accommodate the heterogeneity in read coverage typical of target capture data, and with a coverage cutoff of 5.

To supplement our data from target capture sequencing, we searched GenBank for published whole genome sequencing (WGS) raw reads and genome assemblies and added data for seven more parulid species (three from genomes, four from raw sequencing reads). Among them, *Vermivora bachmanii* was represented by five separate toe pad samples, see Table 1 and also Smith et al., (2021). We also identified outgroup genomes of four species from Passerellidae and three species from Icteridae (Table 1). We used PHYLUCE (Faircloth, 2016) to extract UCE loci from genome assemblies based on the 5K UCE probe set. Upstream and downstream 500 bp from the UCE probe were extracted. To reduce the amount of data needed for assembly from published WGS reads of three species (all but *V. bachmanii*), we made a file of UCEs from the three assembled parulid genomes (comprising 14,184 UCE sequences, with most UCE loci represented three times). This was used as a pseudo-reference genome in Bowtie2 v2.4.2 (Langmead and Salzberg, 2012) to extract loci from WGS raw reads that matched our UCEs. The extracted loci were then assembled using SPAdes as was done for the other sequence data.

From the SPAdes assemblies, we used PHYLUCE (Faircloth, 2016) to extract UCE loci based on the 5K UCE probe set. When combined with the genome data, this led to a total of 122 parulid samples representing 115 of 120 Parulidae species (Table 1) and seven outgroup species. A total of 4,969 UCEs were obtained and each locus was aligned using MAFFT 7 (Katoh and Standley, 2013).

Table 1
Taxon sampling for the UCE dataset. All newly collected sequence data is deposited at NCBI under BioProject PRJNA1177964. For published genome or SRA data, we provided the NCBI GenBank identification. "*" denotes fair quality samples and "**" denotes poor quality samples.

Species[1]	UCE loci based on unfiltered dataset	Sites recovered	% Gap/ ambiguity[2]	Museum code[3]	Tissue or voucher number (tissue first if both listed)	Sample type	Collection year
Basileuterus belli	4733	4,689,659	7.81	FMNH	343,398	Tissue	1989
Basileuterus culicivorus	4776	4,756,287	6.85	FMNH	343,401	Tissue	1989
*Basileuterus ignotus**	2174	545,829	78.41	AMNH	811,716	Toe Pad	1975
Basileuterus lachrymosus	4815	4,693,900	7.76	KU	5024 (93716)	Tissue	2001
Basileuterus melanogenys	4785	4,699,623	7.43	UWBM	107,393	Tissue	2006
Basileuterus punctipectus	4783	4,268,591	13.43	AMNH	DOT-11807	Tissue	2000
Basileuterus tacarcunae	4081	1,432,079	65.06	USNM	484,831	Toe Pad	1964
Basileuterus trifasciatus	4819	4,616,314	8.64	MSB	35,142	Tissue	2010
Basileuterus tristriatus	4817	4,943,237	5.18	LSUMZ	B-44485	Tissue	?
Cardellina canadensis	4806	5,052,930	4.42	UF	46,305	Tissue	2007
Cardellina pusilla	4831	4,917,411	5.26	UF	47,869	Tissue	2010
Cardellina rubra	4795	4,890,523	5.62	UWBM	111,423	Tissue	2004
Cardellina rubrifrons	4796	4,874,301	5.67	UWBM	114,366	Tissue	2010
Cardellina versicolor	4762	4,640,531	8.16	UWBM	105,934	Tissue	2002
*Catharopeza bishopi***	301	73,810	77.19	ROM	109,695	Toe Pad	1971
Geothlypis aequinoctialis	4751	4,826,075	6.17	UF	47,472	Tissue	2009
*Geothlypis auricularis**	2618	651,603	78.65	UF	20,456	Toe Pad	1970
Geothlypis beldingi	4816	4,808,672	6.29	AMNH	DOT-8593	Tissue	1988
Geothlypis chiriquensis	4761	4,752,154	6.97	AMNH	DOT-8951	Tissue	1989
Geothlypis flavovelata	4798	4,833,300	5.90	AMNH	DOT-8423	Tissue	1987
Geothlypis formosa	4815	4,982,823	5.02	UF	50,067	Tissue	2013
Geothlypis nelsoni	4804	4,956,664	5.08	UWBM	112,866	Tissue	2006
Geothlypis philadelphia	4823	4,620,703	8.67	UF	46,234	Tissue	2007
Geothlypis poliocephala	4783	3,414,026	28.44	KU	25,093 (127865)	Tissue	2007
Geothlypis rostrata	4755	4,741,371	7.08	UF	44,191	Tissue	2003
Geothlypis semiflava	4790	4,862,374	5.85	UF	46,339	Tissue	2007
Geothlypis speciosa	4798	4,924,148	5.20	AMNH	DOT-8302	Tissue	1987
Geothlypis tolmiei	4821	4,982,825	4.78	UF	49,778	Tissue	2011
Geothlypis trichas	4689	5,272,355	3.27		GCA_009764595.1	GENOME	2013
Geothlypis velata	4758	4,793,728	6.54	LSUMZ	B-25722	Tissue	?
Helmitheros vermivorum	4741	4,274,174	13.22	UF	49,595	Tissue	2012
Leiothlypis celata	4763	4,715,230	7.23	UF	45,689	Tissue	2007
Leiothlypis crissalis	4751	4,708,217	7.46	FMNH	394,186	Tissue	1989
Leiothlypis luciae	4797	4,934,931	5.36	MSB	28,542	Tissue	2009
Leiothlypis peregrina	4772	4,732,675	7.10	UF	46,246	Tissue	2007
Leiothlypis ruficapilla	4794	4,791,518	6.66	UF	49,581	Tissue	2011
Leiothlypis virginiae	4799	4,889,749	5.61	KU	30,091 (128792)	Tissue	2012
*Leucopeza semperi***	36	9244	75.36	USNM	355,856	Toe Pad	1934
*Leucopeza semperi***	61	16,148	77.99	AMNH	507,561	Toe Pad	1901
Limnothlypis swainsonii	4790	4,697,407	7.28	UF	45,493	Tissue	2006
Mniotilta varia	4809	5,015,504	4.66	UF	42,459	Tissue	2001
Myioborus albifacies	4796	4,735,900	6.75	AMNH	DOT-4910	Tissue	1995
Myioborus albifrons	4767	4,513,044	9.34	LSUMZ	B-25313	Tissue	?
Myioborus brunniceps	4815	4,778,565	7.14	KU	9730 (96860)	Tissue	2005
Myioborus cardonai	4768	4,534,296	9.39	LSUMZ	B-25322	Tissue	?
Myioborus castaneocapilla	4761	4,702,589	7.37	KU	4017 (92239)	Tissue	2001
*Myioborus flavivertex***	270	65,681	78.70	USNM	388,068	Toe Pad	1946
Myioborus melanocephalus	4790	4,873,124	5.55	LSUMZ	B-44407	Tissue	?
Myioborus miniatus	4780	4,443,414	10.97	AMNH	DOT-16297	Tissue	2009
Myioborus ornatus	3812	1,371,190	65.05	ROM	94,444	Toe Pad	1959
Myioborus pictus	4829	5,121,571	4.06	UWBM	114,032	Tissue	2009
Myioborus torquatus	4784	4,850,644	5.79	UWBM	107,406	Tissue	2006
*Myiothlypis basilica**	3166	914,790	73.24	USNM	388,057	Toe Pad	1946
Myiothlypis bivittata	4811	5,125,383	4.09	FMNH	430,138	Tissue	2001
*Myiothlypis chlorophrys***	547	130,247	80.14	ROM	62,551	Toe Pad	1898

(*continued on next page*)

Table 1 (*continued*)

Species[1]	UCE loci based on unfiltered dataset	Sites recovered	% Gap/ ambiguity[2]	Museum code[3]	Tissue or voucher number (tissue first if both listed)	Sample type	Collection year
Myiothlypis chrysogaster	4801	4,664,581	7.87	MSB	27,264	Tissue	2007
*Myiothlypis cinereicollis***	42	10,834	77.82	AMNH	388,404	Toe Pad	1952
*Myiothlypis conspicillata***	502	117,871	79.69	USNM	388,101	Toe Pad	1945
Myiothlypis coronata	4820	4,806,461	6.53	LSUMZ	B-44043	Tissue	?
Myiothlypis flaveola	4809	5,045,332	4.53	FMNH	334,601	Tissue	1987
Myiothlypis fraseri	4825	4,664,470	8.10	LSUMZ	B-66111	Tissue	?
*Myiothlypis fulvicauda**	3329	1,056,904	69.41	UF	39,028	Toe Pad	1993
*Myiothlypis griseiceps**	1582	370,391	80.19	FMNH	92,358	Toe Pad	1932
Myiothlypis leucoblephara	4796	4,959,696	5.21	KU	3738 (91645)	Tissue	2000
*Myiothlypis leucophrys***	208	40,932	84.04	AMNH	270,394	Toe Pad	1929
Myiothlypis luteoviridis	4831	5,137,264	3.99	LSUMZ	B-43954	Tissue	?
Myiothlypis nigrocristata	4789	4,971,785	4.94	MSB	35,159	Tissue	2010
Myiothlypis rivularis	4800	4,904,088	5.38	KU	5730 (94871)	Tissue	2002
Myiothlypis roraimae	4802	4,994,915	4.69	KU	4079 (94406)	Tissue	2001
Myiothlypis signata	4389	1,975,300	56.06	FMNH	398,567	Tissue	1999
Oporornis agilis	4770	4,851,226	6.14	UF	46,800	Tissue	2007
Oreothlypis gutturalis	4771	4,758,875	6.78	AMNH	DOT-3663	Tissue	1982
Oreothlypis superciliosa	4838	4,719,639	7.83	FMNH	394,178	Tissue	1989
Parkesia motacilla	4792	4,896,929	5.43	UF	49,965	Tissue	2011
Parkesia noveboracensis	4802	4,633,590	8.41	UF	46,355	Tissue	2008
Protonotaria citrea	4777	4,755,347	6.95	UF	49,571	Tissue	2012
Seiurus aurocapilla	4802	4,686,800	7.51	UF	42,497	Tissue	2001
Setophaga adelaidae	4777	4,974,896	4.98	UWBM	65,259	Tissue	1997
Setophaga aestiva	4792	4,673,162	7.46	UF	46,243	Tissue	2007
Setophaga americana	4785	4,635,508	8.16	UF	48,359	Tissue	2011
Setophaga angelae	4760	4,506,824	9.60	LSUMZ	B-11455	Tissue	?
Setophaga auduboni	4814	4,279,310	13.38	UF	44,458	Tissue	2000
Setophaga caerulescens	4756	3,971,049	17.97	UF	44,598	Tissue	2004
Setophaga castanea	4788	4,857,060	5.90	UF	45,495	Tissue	2006
Setophaga cerulea	4747	4,798,644	6.62	KU	2462 (89178)	Tissue	1998
Setophaga chrysoparia	4817	5,091,755	4.25	TCWC	22,841	Tissue	2012
Setophaga citrina	4704	4,581,436	9.12	UF	50,506	Tissue	2013
Setophaga coronata	4664	5,206,052	3.59		GCA_001746935.2	GENOME	2005
Setophaga delicata	3783	1,302,178	66.30	ROM	111,701	Toe Pad	1971
Setophaga discolor	4793	4,797,988	6.30	UF	48,170	Tissue	2002
Setophaga dominica	4678	3,287,331	30.74	UF	42,493	Tissue	1999
Setophaga flavescens	4816	4,962,436	4.93	FMNH	397,152	Tissue	2000
Setophaga fusca	4797	4,921,971	5.27	UF	45,550	Tissue	2006
Setophaga graciae	4739	4,553,751	9.20	UWBM	77,516	Tissue	2004
Setophaga kirtlandii	4510	2,101,649	54.01	UF	44,179	Tissue	2004
Setophaga kirtlandii	4744	5,322,706	3.60		GCA_013399655.1	GENOME	1988
Setophaga magnolia	4758	4,809,627	6.49	UF	45,700	Tissue	2006
Setophaga nigrescens	4791	5,037,715	4.54	MSB	26,805	Tissue	2008
Setophaga occidentalis	4815	4,209,365	14.24	UF	48,094	Tissue	2005
Setophaga palmarum	4806	4,995,323	4.76	UF	48,317	Tissue	2011
Setophaga pensylvanica	4820	4,881,800	5.84	UF	46,293	Tissue	2007
Setophaga petechia	4820	4,967,327	4.94	UF	48,314	Tissue	2011
Setophaga pharetra	3962	1,749,336	59.82	FMNH	331,125	Tissue	1987
Setophaga pinus	4744	4,517,910	9.67	UF	47,524	Tissue	2010
Setophaga pitiayumi	4781	4,721,942	7.39	KU	18,797 (113606)	Tissue	2009
Setophaga pityophila	4766	4,627,676	8.62	FMNH	397,154	Tissue	2000
*Setophaga plumbea**	3349	969,873	73.94	TCWC	12,760	Toe Pad	1990
Setophaga ruticilla	4799	4,682,283	7.86	UF	46,696	Tissue	2008
Setophaga striata	4774	4,842,855	5.90	UF	50,272	Tissue	2013
*Setophaga subita***	1005	239,848	78.69	USNM	572,215	Toe Pad	1983
Setophaga subita	4581	4,438,901	13.10		SRR13216240	SRA (Tissue)	?
Setophaga tigrina	2727	1,644,934	40.85		SRR13176648	SRA (Tissue)	1992
Setophaga townsendi	4807	4,851,880	5.75	MSB	22,490	Tissue	1996
Setophaga virens	4785	4,571,346	9.28	UF	49,596	Tissue	2009

(*continued on next page*)

Table 1 (*continued*)

Species[1]	UCE loci based on unfiltered dataset	Sites recovered	% Gap/ ambiguity[2]	Museum code[3]	Tissue or voucher number (tissue first if both listed)	Sample type	Collection year
Setophaga vitellina	4518	4,315,241	14.31		SRR13216255	SRA (Tissue)	?
*Vermivora bachmanii***	596	142,676	77.93		SRR12612992	SRA (Toe Pad)	1890
*Vermivora bachmanii**	1711	423,974	76.05		SRR12612994	SRA (Toe Pad)	1911
*Vermivora bachmanii**	2143	548,653	77.01		SRR12612997	SRA (Toe Pad)	1906
*Vermivora bachmanii**	2809	727,692	74.82		SRR12612993	SRA (Toe Pad)	1915
*Vermivora bachmanii**	2904	768,274	75.00		SRR12612999	SRA (Toe Pad)	1905
Vermivora chrysoptera	4773	4,777,790	6.64	KU	31,965 (123588)	Tissue	2014
Vermivora cyanoptera	4798	4,834,249	6.22	KU	3545 (91054)	Tissue	2000
Outgroup							
Spizella passerina	3696	3,689,900	13.45		GCA_013401375.1	GENOME	
Zonotrichia albicollis	4731	5,244,646	4.41		GCA_000385455.1	GENOME	
Quiscalus mexicanus	4803	5,362,236	3.71		GCA_013399035.1	GENOME	
Junco hyemalis	4427	4,962,549	3.59		GCA_003829775.2	GENOME	
Melospiza melodia	4777	5,366,483	3.44		GCA_011057915.1	GENOME	
Agelaius phoeniceus	4800	5,379,390	3.42		GCA_013398535.1	GENOME	
Molothrus ater	4806	5,400,333	3.29		GCA_013401155.1	GENOME	

[1] We reconciled the names attributed to each specimen with the IOC World Bird List 13.1; names used by individual museums may differ.
[2] The proportion of gap and ambiguities (i.e., "?", "N" and "-") was calculated across all UCE alignments using IQ-TREE2.
[3] Museum codes: AMNH = American Museum of Natural History; FMNH = Field Museum of Natural History; KU = University of Kansas Biodiversity Institute; LSUMZ = Louisiana State Museum of Natural History; MSB = Museum of Southwestern Biology; ROM = Royal Ontario Museum; TCWC = Texas A&M University Biodiversity Research and Teaching Collections; UF = Florida Museum of Natural History; USNM = National Museum of Natural History, Smithsonian Institution; UWBM = University of Washington Burke Museum.

2.3. Data filtering and phylogenetic analyses

We ran a partitioned concatenation analysis using the 4,969 UCE alignments in IQ-TREE v2.2.2 (Minh et al., 2020) with 1,000 ultrafast bootstrap replicates (-B 1000) (Hoang et al., 2018) and used the implemented ModelFinder (Kalyaanamoorthy et al., 2017) to select the best-fit partitioning scheme (-m TESTMERGE) while allowing each partition to have its own rate (-p) and reducing the computational burden with relaxed hierarchical clustering algorithm (–rcluster 10). This tree rendered a few expected genera to be non-monophyletic and contained some exceptionally long branches (Fig. 1). Therefore, we decided to perform the following data filtering steps.

We calculated the proportion of gap and ambiguity ("-", "?" or "N") for every taxon in each locus alignment and averaged them across all UCE alignments for each taxon (Table 1). We selected 102 taxa that had gaps and ambiguities lower than 70%. After re-aligning the UCEs, we trimmed the two ends of each alignment until there was a site with at least 50% of the taxa sampled in that locus present. We then retained 4,653 UCE alignments with a minimum of 20 taxa and 500 bp in length for downstream phylogenetic analysis.

Based on these alignments, we first ran a concatenation analysis in IQ-TREE2 with partitioned models (-m TESTMERGE) and 1,000 ultrafast bootstrap replicates (-B 1000) using the same parameter settings as described above. For each locus alignment, we also estimated a gene tree using IQ-TREE2 with the implemented ModelFinder to determine the best-fit model (-m TEST) by considering the Gamma rate heterogeneity and invariable site and assessed the branch support with 1,000 ultrafast bootstrap replicates (-B 1000). These gene trees were combined to estimate a species tree using weighted ASTRAL hybrid (Zhang and Mirarab, 2022) in ASTRAL v5.15.5 (Zhang et al., 2018) with a maximum branch support value of 100 and a minimum support of 50. For nodes near the base of Parulidae, we used DiscoVista (Sayyari et al., 2018) to visualize the relative frequency of the gene trees that support the topology present in the weighted ASTRAL tree or alternative topologies. We then examined whether conditions of incomplete lineage sorting were met.

2.4. Adding in bad and ugly samples

For taxa with an average proportion of gap and ambiguity greater than 70%, we divided them into four groups (Table 2): 1) fair-quality samples (the "bad"), including 10 taxa with locus sampling > 1,500 loci; 2) seven poor-quality samples (the "ugly"), including seven taxa with locus sampling < 1,500 loci. The next two groups include monotypic genera with very poor-quality samples. Due to the difficulty of placing these taxa (e.g., they do not have a congener to group with), we treated these samples separately and placed them in the last two identified groups; 3) the two poor-quality samples of *Leucopeza semperi*; and 4) the only poor-quality sample of *Catharopeza bishopi*. For these last two groups, we explored their phylogenetic positions in the Parulidae separately (see below). For the remaining 17 bad and ugly samples, we added them to the phylogeny by running two subsequent constraint tree searches.

First, we generated a strict consensus tree from the concatenation tree and the ASTRAL species tree built based on the relatively better-sampled "good" taxonset (gap and ambiguity < 70%) in the previous section using SumTrees from DendroPy (Sukumaran and Holder, 2010). This consensus tree contained polytomies for unresolved nodes and would serve as the constraint tree in the first constraint tree search. We then extracted UCE alignments that included only the 102 taxa plus the 10 bad samples from the original unfiltered dataset and realigned each locus using MAFFT. We trimmed the two ends of each alignment until there was a site where at least 50% of the bad samples were present and at least 50% of all sampled taxa in that locus were present. This led to a total of 3,116 UCE alignments for 112 taxa (the "good + bad" taxonset). We concatenated these alignments to run model test using IQ-TREE v2.2.2 (-TESTMERGEONLY), and then used the best-fit partitioning scheme to run a constraint tree analysis in IQ-TREE v.1.6.12 (Nguyen et al., 2015) using the unrooted strict consensus tree as the topological constraint tree (-g) and the ultrafast bootstrap to examine support (-bb

Fig. 1. Concatenated tree of 122 parulid samples based on the **unfiltered** UCE dataset, highlighting clustering of poorly performing samples. The phylogenetic positions for bad samples (tip labels in orange) and ugly samples (tip labels in red) appeared problematic (except for *Vermivora bachmanni*) and were explored in additional analyses. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 2
Datasets used in this study. Complete mitogenomes include all 13 protein coding genes and both ribosomal RNA genes.

Data type	Dataset	Samples	Composition	Definition	Species
UCEs	Unfiltered/Full	122	All parulid samples	–	115
	The "good" taxonset	102	Good quality samples	Gaps & ambiguities < 70%	101
	The "bad"	10	Fair quality samples	Gaps & ambiguities > 70%; loci > 1500	7
	The "good + bad" taxonset	112	The good taxonset + 10 fair quality samples	–	108
	The "ugly"	10	2 *Leucopeza*, 1 *Catharopeza*, and 7 others poor quality samples	Gaps & ambiguities > 70%; loci < 1500	9
	The "good + bad + ugly" taxonset	119	The good + bad taxonset + 7 poor quality samples	–	113
	Dataset for *Leucopeza*	103	The good taxonset + 2 *Leucopeza* merged	79 loci recovered for *Leucopeza*	102
	Dataset for *Catharopeza*	103	The good taxonset + 1 *Catharopeza*	296 loci recovered for *Catharopeza*	102
Mitochondria	Mitogenomes	122	70 complete and 52 partial mitogenomes	–	115

1000). A different version of IQ-TREE was applied because the constraint tree estimation was unstable in IQ-TREE2.

Second, we extracted UCE alignments that included the 102 taxa, 10 bad samples plus the seven ugly samples from the original unfiltered dataset and realigned each locus. We trimmed the two ends of each alignment until there was a site where at least 10% of the poor-quality samples were present and at least 10% of all sampled taxa in that locus were present. This resulted in a total of 2,213 UCE alignments for 119 taxa (the "good + bad + ugly" taxonset). We then concatenated these alignments to run a model test in IQ-TREE2 and a constraint tree search in IQ-TREE v1.6.12 using the resulting unrooted concatenated tree from the previous constraint analysis (good + bad taxonset) as the constraint tree.

2.5. Exploring phylogenetic position for Leucopeza and Catharopeza

The two toe pad samples of *Leucopeza semperi* that were sequenced recovered little data (Table 1). Therefore, we combined the UCE loci extracted from the two samples to represent this species. For *Leucopeza* and *Catharopeza* respectively, we created a supermatrix to include just the UCE alignments recovered for *Leucopeza* or *Catharopeza*. For each locus, we trimmed the ends until there was a site where the focal taxa were present. We then checked the alignments individually by eye in Geneious Prime 2024.0.2 (https://www.geneious.com) to further trim the two ends and remove any problematic loci that likely resulted from contamination (see Supplementary Figure S1 for examples). This led to 79 unique loci for *Leucopeza*. Similarly, the toe pad sample from the endangered *Catharopeza bishopi* recovered 296 UCE loci. We then ran a model test in IQ-TREE2 and a constraint tree search in IQ-TREE v1.6.12 using the consensus tree based on the good taxonset as the constraint tree.

2.6. Mitogenome data collection and tree estimation

We used MitoFinder v1.4 (Allio et al., 2020) to assemble mitogenomes and extract individual mitochondrial regions from the SPAdes assemblies, raw reads from the NCBI SRA database and mitogenomes from GenBank. For taxa that lacked published mitogenome data (as of December 2023) or failed the MitoFinder extraction due to low coverage of sequencing data, we downloaded individual mitochondrial regions from GenBank nucleotide database to ensure that mitochondrial dataset had the same taxon sampling as the UCE dataset. When using published data, we limited our selection to sequences from the same voucher specimen (see Supplementary Table S1 for accessions).

We obtained a total of 15 mitochondrial regions, including 13 protein coding genes and two ribosomal RNA regions (Supplementary Table S1); for 70 taxa we included all 15 regions, while another 52 taxa were represented by partial mitogenome data. Alignments were built with MAFFT and TranslatorX (Abascal et al., 2010) was used to guide alignments of the protein coding genes based on corresponding amino acid translations (https://translatorx.co.uk/). A concatenated NEXUS data block was built using phyutility 2.7.1 (Smith and Dunn, 2008). We used IQ-TREE2 to identify the best-fit partitioning scheme (-m TESTMERGE) with protein coding genes partitioned by codon positions (i.e., 1st, 2nd, and 3rd codon positions) and estimated a phylogeny with 1,000 ultrafast bootstrap replicates (-B 1000).

3. Results

3.1. Taxon sampling

We sampled UCE data for a total of 122 parulid samples, representing all 18 genera and 115 out of the 120 Parulidae species recognized by IOC World Bird List 13.1 (Table 1). These parulid samples on average had 4,186 UCE loci, and only 10 samples had a locus sampling lower than 1,500 loci. Genome assemblies and tissue samples, as expected, recovered more sites than toe pad samples from historical museum specimens (Supplementary Figure S2). We analyzed a dataset of 102 well-sampled taxa, representing 101 unique species, 112 taxa (including 10 bad samples; Table 2) and 119 taxa (including seven ugly samples).

3.2. Phylogenetic relationships based on UCEs

For the concatenated tree with the unfiltered dataset (Fig. 1), we observed samples with fair or poor sequence recovery had long branches and clustered together (excluding the five samples of *Vermivora bachmanii*), which rendered *Myiothlypis*, *Setophaga*, *Myioborus*, *Geothlypis* and *Basileuterus* non-monophyletic. All other taxa (excluding the monotypic genera) were grouped together with their expected congeners, except for *Basileuterus lachrymosus*. Given that the issues all appeared to be due to low-quality data, we filtered the dataset based on data quality and explored processes to add in low-quality samples that could help us understand whether their positions were an artifact of low-quality samples or might reflect the actual underlying patterns of diversification.

In the concatenated ML tree for the good taxonset (well-sampled, with gaps and ambiguities < 70%), all genera were monophyletic except *Basileuterus* (which was non-monophyletic because *B. lachrymosus* was sister to the clade uniting *Myioborus*, *Cardellina* and other *Basileuterus*; Fig. 2). Nodes in this tree overall received high ultrafast bootstrap support (BS > 90). Most taxa had similar terminal branch lengths and the small number of taxa with unusually-long terminal branches corresponded with samples with relatively high proportions of gaps and ambiguities (e.g., *Setophaga delicata*, 66%; *Basileuterus tacarcunae*, 65%; *Myioborus ornatus,* 65%; *Setophaga pharetra,* 60%; and *Myiothlypis signata*, 56%; Table 1).

The weighted ASTRAL tree based on gene trees estimated from the same UCE dataset (Fig. 3) also recovered all genera as monophyletic, with the exception of *Basileuterus* (which was also rendered non-monophyletic by *B. lachrymosus*, as in the concatenated ML tree). However, topologies from the concatenated tree and the ASTRAL tree differed in several places, including relationships within *Setophaga*,

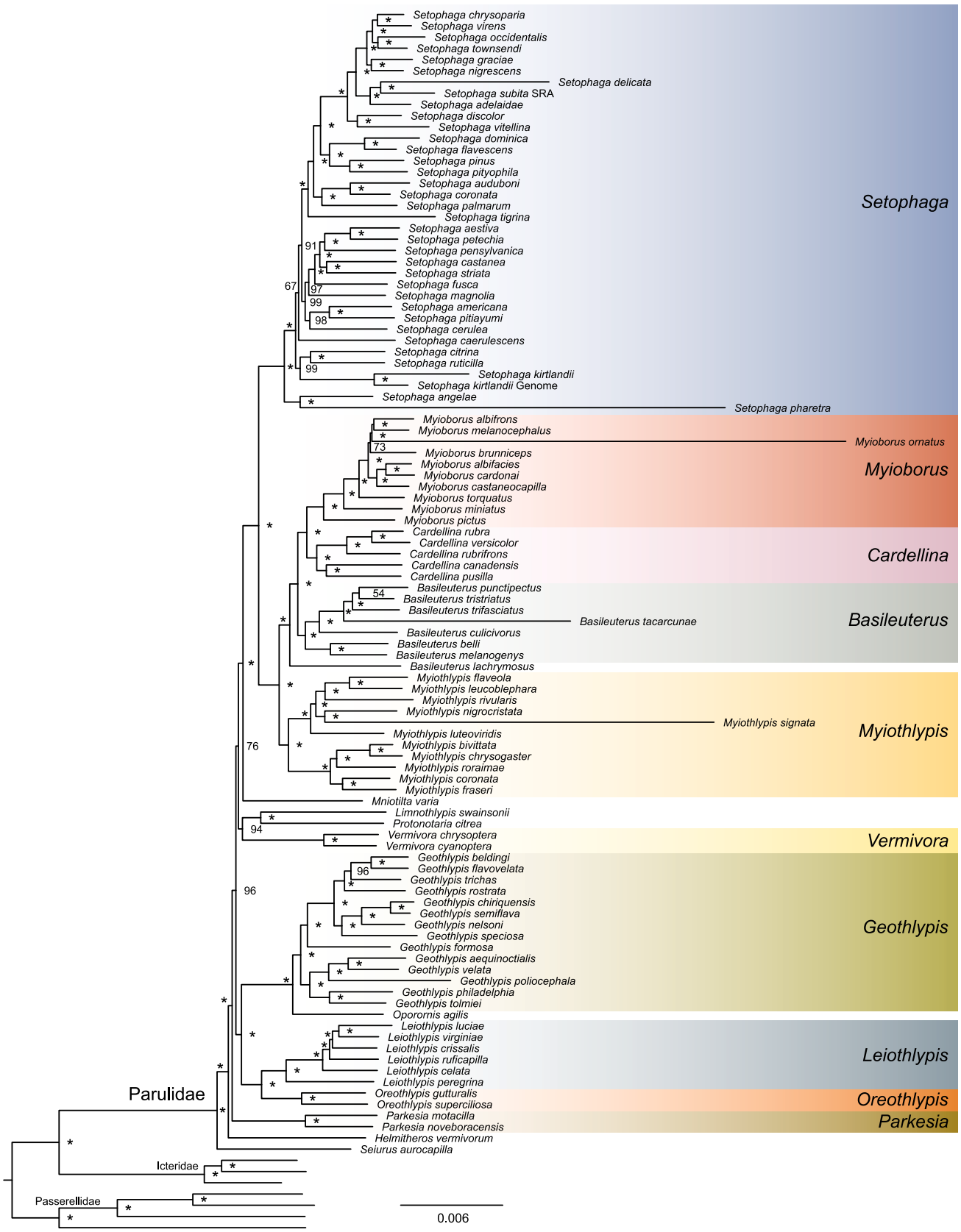

Fig. 2. Concatenated tree of the good taxonset based on the filtered UCE dataset that has an average proportion of gaps and ambiguities lower than 70%. Values at nodes are ultrafast bootstrap support values (* indicates full support).

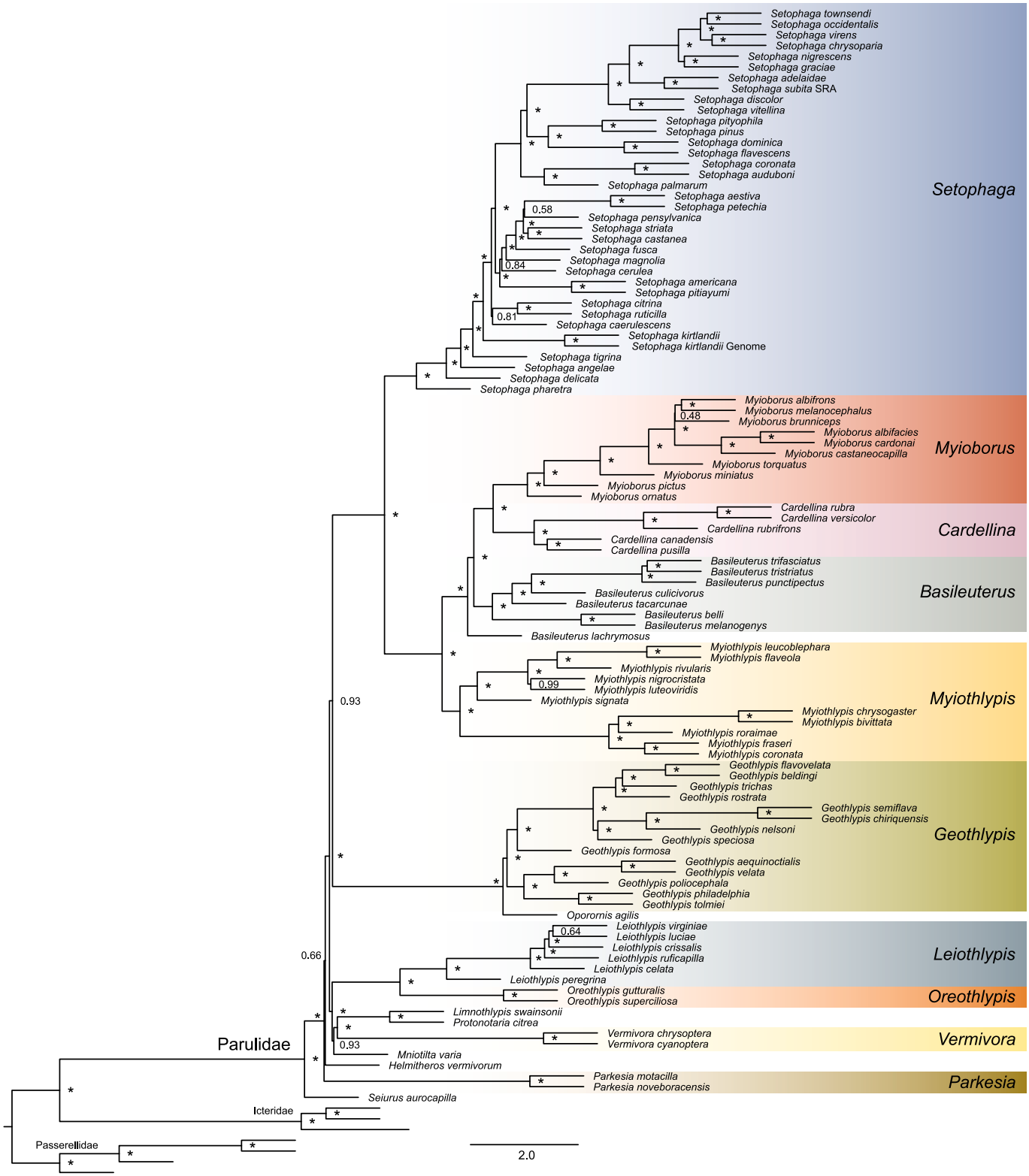

Fig. 3. Weighted ASTRAL tree of the good taxonset based on gene trees estimated from filtered UCE dataset that has average proportion of gaps and ambiguities lower than 70%. Values at nodes are quartet support values (* indicates full support).

Myioborus, *Myiothlypis* and *Basileuterus*, and arrangements among some genera (Fig. 4). The branches that conflict between the concatenated ML tree and the weighted ASTRAL tree near the base of Parulidae were associated with substantial conflict among the estimated gene trees (Fig. 5). However, there did not appear to be much asymmetry in the support for minority quartets, suggesting that incomplete lineage sorting is likely to be a major source of conflict among gene trees in this part of the species tree. Despite the generally high support for these branches, the branch lengths in this part of the tree are remarkably short (regardless of whether they are measured in substitutions per site or coalescent units). Given these very short branch lengths we present a strict consensus tree where topological conflicts are shown as

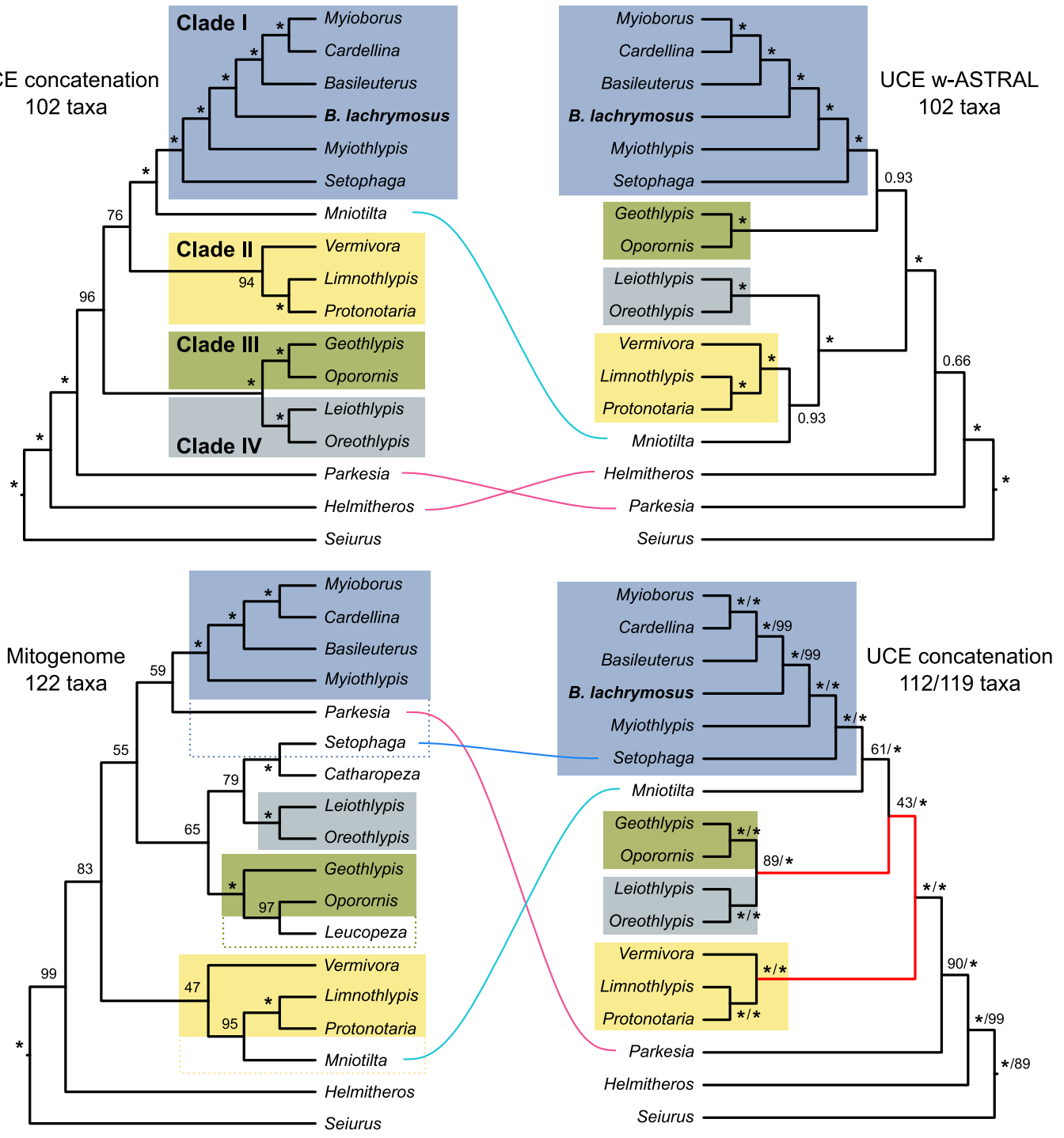

Fig. 4. Topological comparisons of relationships among genera across different analyses in our study. We labeled four major clades in the concatenated UCE tree based on the good taxonset (102 taxa). In the mitogenome tree, *B. lachrymosus* is nested within *Basileuterus*, and *Leucopeza* and *Catharopeza* are only present in this tree. Clades I, II and III are not monophyletic in the mitogenome tree, therefore dotted lines were used. The concatenated tree based on the good + bad taxonset (112 taxa) yields the same relationships among genera as those in the tree using the good + bad + ugly taxonset (119 taxa), except that *Myiothlypis cinereicollis* was sister to the rest of Parulidae for latter (pruned here, see Supplementary Figure S4 for complete tree). Tree branches in red denote topological incongruence between the concatenated tree using the good taxonset and the good + bad/good + bad + ugly taxonsets. Values near nodes are support values. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

polytomies to facilitate visualization of these conflicts (Supplementary Figure S3).

The phylogeny estimated by adding in the 10 bad samples to the good taxonset strict consensus tree (Supplementary Figure S3) also yielded a tree in which all genera were monophyletic, except for *Basileuterus* (Fig. 6). Long terminal branches in the tree all corresponded with bad samples, and these were often placed with high support. However, of the seven ugly samples (added to produce the good + bad + ugly taxonset), only six were placed within their genus. *Myiothlypis cinereicollis*, which was only sampled for 42 UCE loci, was placed sister to all other Parulidae species (Supplementary Figure S4). Otherwise, relationships among genera remained the same as those using the good +

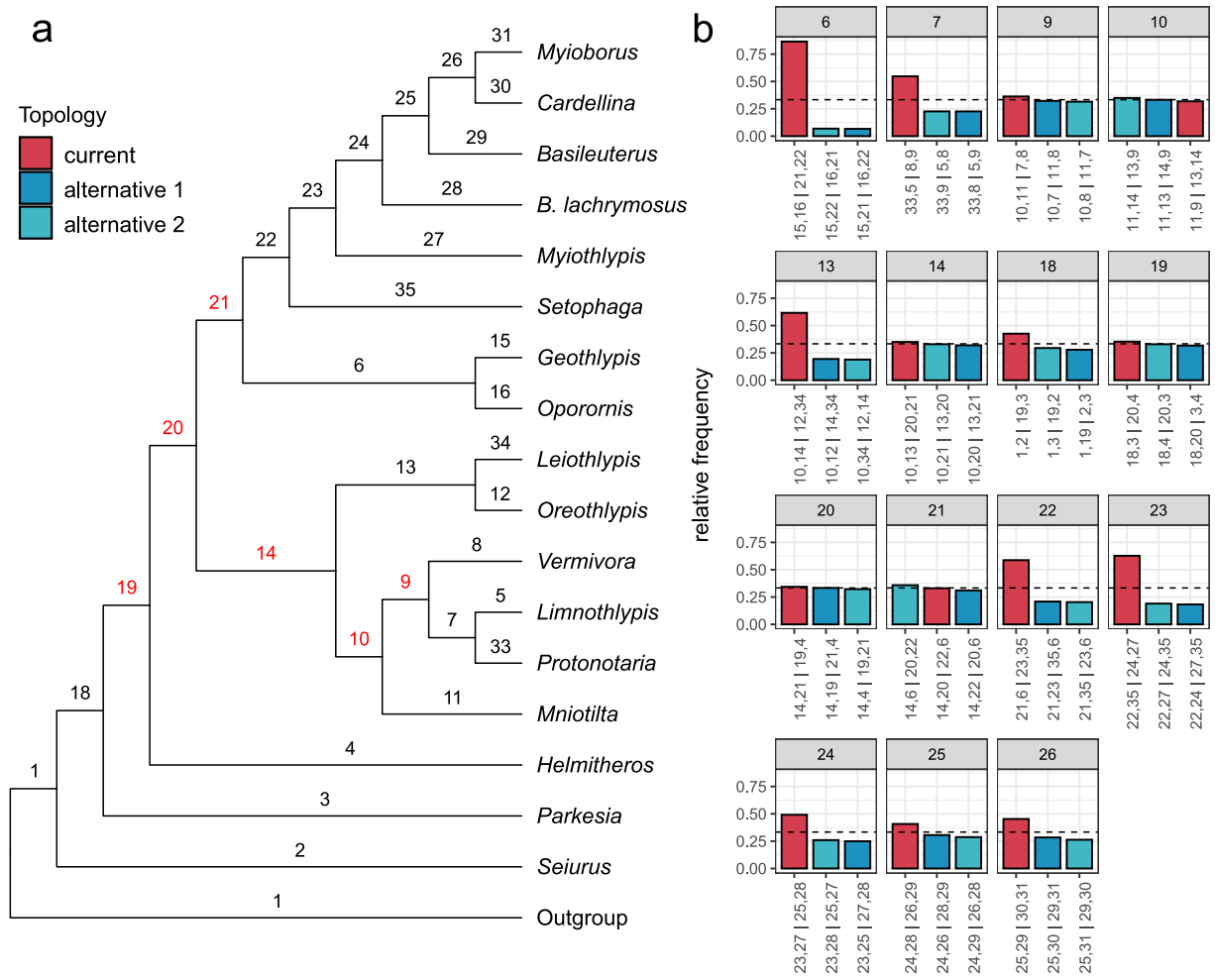

Fig. 5. Relative frequency of the gene trees that support each of the three possible topologies around a focal branch near the base of Parulidae. a. Topology of the weighted ASTRAL tree collapsed to the genus level, with branch number labeled. b. The relative frequency of the gene trees that support a specific topology for each internal branch. Red bar denotes the topology present in the weighted ASTRAL tree, whereas blue and teal bars represent the two alternative topologies. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

bad taxonset (Fig. 4).

All UCE trees (concatenation and ASTRAL; Fig. 4) consistently recovered Clade I (*Myioborus*, *Cardellina*, *Basileuterus*, *Myiothlypis,* and *Setophaga*), Clade II (*Vermivora*, *Limnothlypis,* and *Protonotaria*), Clade III (*Geothlypis* and *Oporornis*) and Clade IV (*Leiothlypis* and *Oreothlypis*). The exceptionally short internal branches connecting these lineages (Figs. 2 & 3) suggest rapid radiation early in their evolutionary history. Although the relative positions among these clades differed across analyses, the concatenated UCE tree based on the good taxonset (Fig. 2) overall received better support for the deep nodes.

3.3. Phylogenetic position of the monotypic Leucopeza and Catharopeza

We explored the phylogenetic position for *Leucopeza semperi* by adding it into the good taxonset and only using the 79 UCE loci present in the combined *Leucopeza* dataset (Supplementary Figure S5). The concatenated tree showed that *Leucopeza* was sister to *Myioborus* + *Cardellina* + *Basileuterus* + *Myiothlypis* with moderate support (BS = 86).

Similarly, we explored the phylogenetic position of *Catharopeza bishopi* using a subset of the supermatrix that contained 296 UCE loci (Supplementary Figure S6). The concatenated tree showed that *C. bishopi* was nested within *Setophaga* with moderate support (BS = 79). The two insular species, *S. angelae,* and *S. pharetra*, which formed a clade that first diverged from *Setophaga*, were sister to the clade uniting the insular *Catharopeza* and the remaining *Setophaga*.

3.4. Phylogenetic relationships based on mitogenomes

The mitogenome tree found all genera as monophyletic, including *Basileuterus* (Fig. 7). *Leucopeza semperi*, which was only sampled for NADH dehydrogenase subunit 2 (ND2) sequences, was sister to the monotypic *Oporornis*, and together were sister to *Geothlypis*. *Catharopeza bishopi* was sister to all *Setophaga*. However, well-supported relationships among genera in the UCE trees were sometimes rearranged (e.g., Fig. 4) in the mitogenome tree. For relationships within the more speciose genera, we discussed in detail below.

4. Discussion

Our well-sampled Parulidae phylogeny generally confirms previously identified relationships (e.g., Lovette et al., 2010), but with support from genome-wide sampling. However, we also identified differences that may have been driven by conflicting mitochondrial signals and/or stochastic differences due to sampling few loci. The impact of sampling a limited number of loci in earlier studies was likely exacerbated by the rapid radiation near the base of the family (note the branch lengths in Figs. 2 and 3 and the gene tree conflict evident in Fig. 5). Sample quality had an obvious impact on our initial estimates of phylogeny. However, by focusing on estimating a phylogeny with the highest-quality samples, and then iteratively adding in samples where sequence recovery was poor, we were able to overcome some of the

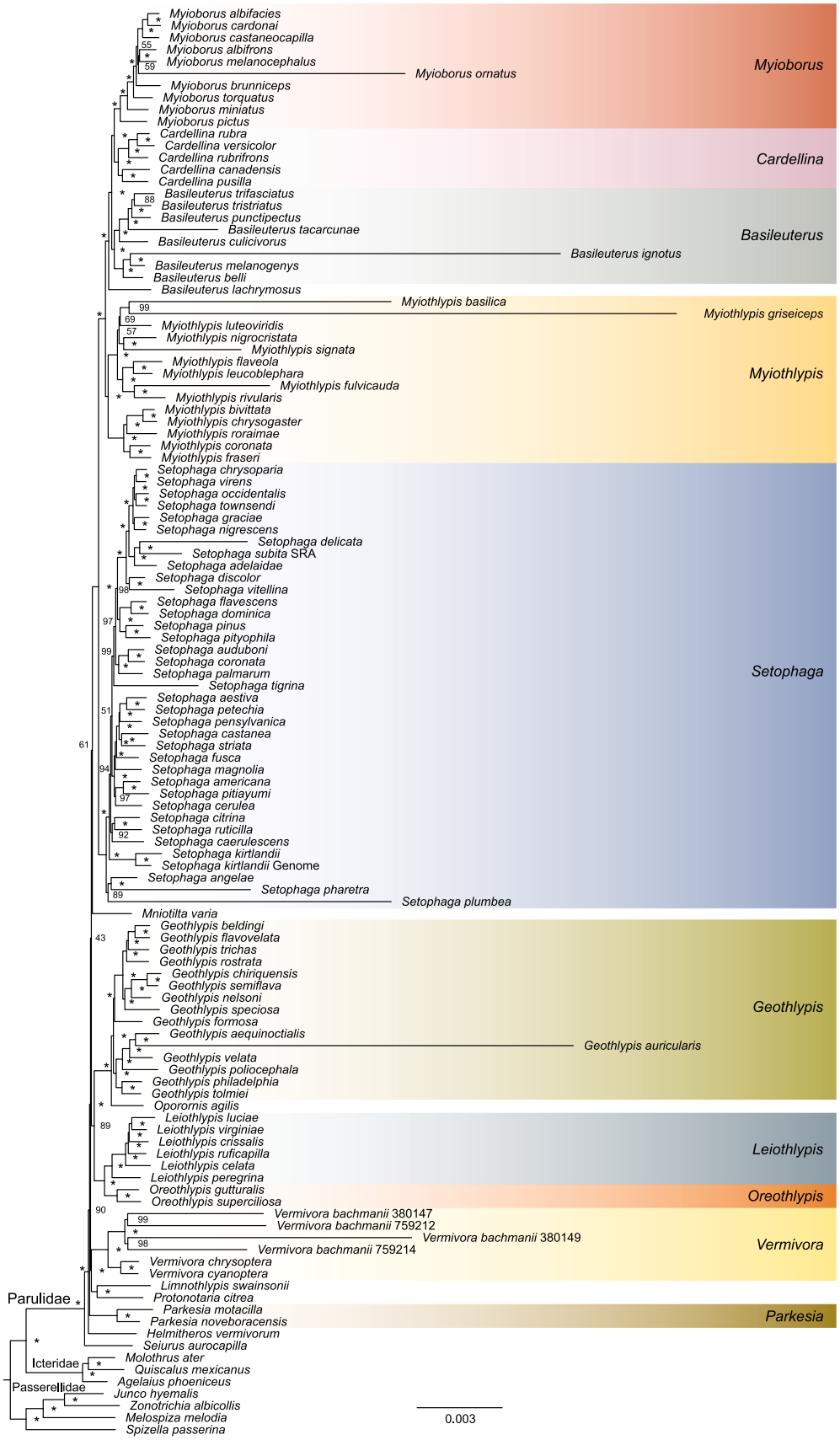

Fig. 6. Concatenated UCE tree of the good + bad taxonset (102 well sampled taxa plus 10 fair-quality samples) estimated with topological constraint. Values at nodes are ultrafast bootstrap support values (* indicates full support).

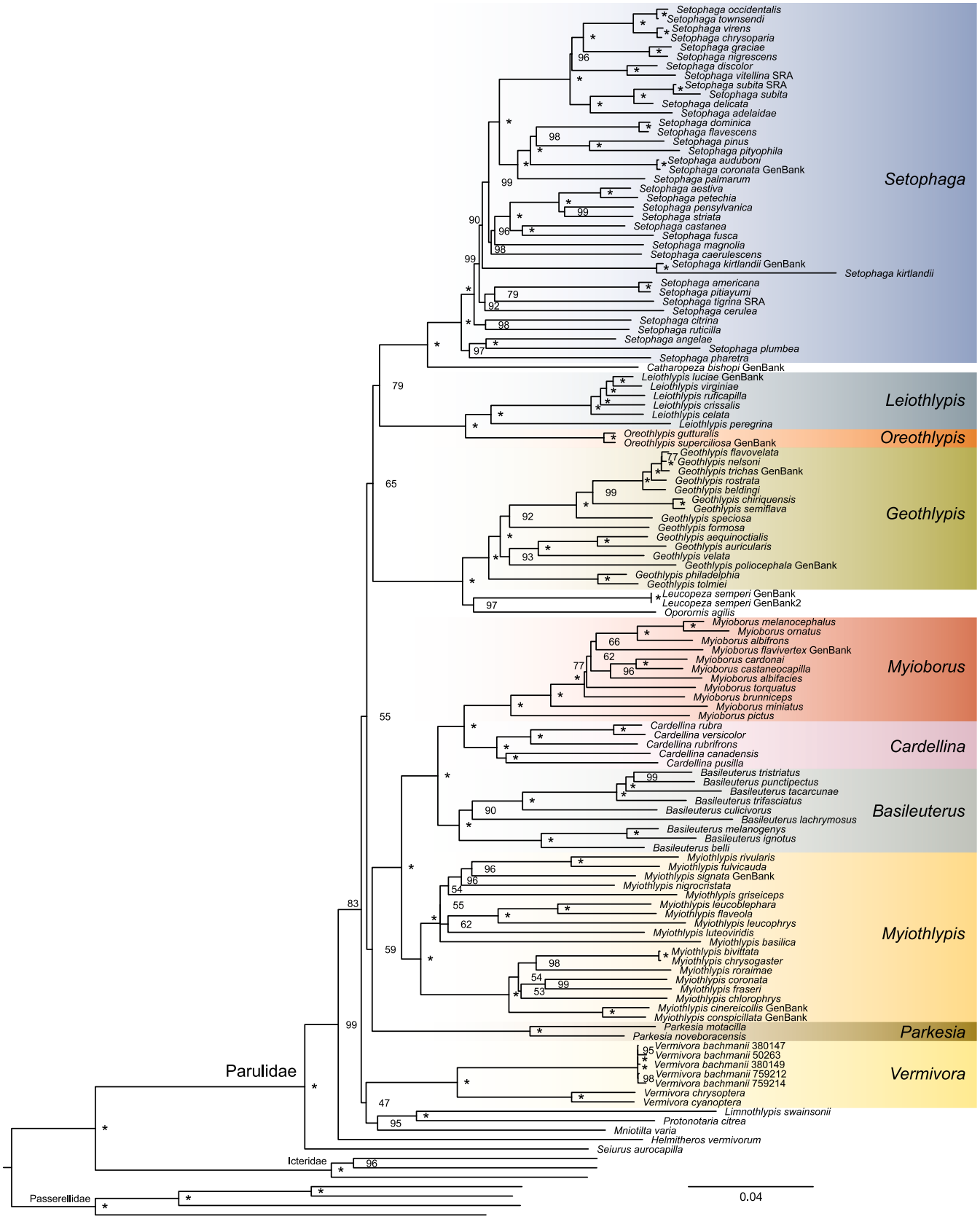

Fig. 7. Mitogenome tree of 122 parulid taxa. Values at nodes are ultrafast bootstrap support values (* indicates full support).

problems and limitations that arose in initial comprehensive analyses.

4.1. Data quality considerations

Many bird species are represented in museums only by historical specimens from which DNA recovery can be poor. Both UCE loci recovery and reduced UCE loci length are negatively affected by specimen age (McCormack et al., 2016) which ultimately reduces their utility for phylogenetics. Library preparation kits, enrichment protocols, and bead cleanup protocols (e.g., Oswald et al., 2023) are available to recover highly fragmented and degraded DNA from samples. However, these kits and protocols can be resource intensive (e.g., time, money, etc.) or unavailable options at library preparation and sequencing facilities. While deeper sequencing of degraded samples aids in the recovery of more UCE loci (McCormack et al., 2016), it also increases costs. For UCE samples that have not undergone remedial pre-sequencing processing, computational approaches (e.g., locus trimming and filtering; Supplementary Figure S1) may help with poorly sequenced samples, which otherwise often appear in phylogenies as long branches (e.g., Braun et al., 2024) and are sometimes placed in spurious phylogenetic positions due to increased sequencing errors and assembly issues. Here we demonstrate that the Parulidae UCE dataset composed of 101 species with < 70% gaps & ambiguities recovered a robust phylogeny for this family, where the long branches that reflect missing data did not appear to affect relationships. With this robust set of relationships, the subsequent inclusion of bad samples also yielded the expected results.

The above strategy can be applied to other phylogenomics studies which might include both good quality samples and historical specimens. The first and most fundamental step would be to understand the overall quality of all the samples. Our cutoffs of selecting good, bad and ugly are based on the quality of our samples, thus can be arbitrary. Other studies should evaluate their own datasets and determine the appropriate cutoffs. To create a reliable backbone tree, we used both concatenation ML tree and ASTRAL coalescent species tree. This was done due to topological conflicts between the two trees and that Parulidae has gone through rapid radiation therefore incomplete lineage sorting can be exacerbated in some of the short internal branches. Using different analytical approaches in the initial exploration can be necessary, especially when the focal group is known to have complex evolutionary histories such as radiation and hybridization.

4.2. Intergeneric relationships in Parulidae based on high UCE loci recovery

Despite topological conflicts observed for analyses using different analytical methods, information contents and data types, our study provides a robust estimate of the phylogenetic relationships for Parulidae, especially the trees based on the good taxonset with relatively high sequence recovery (Figs. 2 & 3). Our trees also largely agreed on the generic circumscription and intergeneric relationships shown in Lovette et al. (2010), but there are several topological differences between studies. We compared the topologies in our phylogenies to the tree estimated using all data in Lovette et al. (2010, their Figure 5, hereinafter "Lovette tree") in detail below.

Clade I (*Myioborus*, *Cardellina*, *Basileuterus*, *Myiothlypis* and *Setophaga*) – All UCE trees (concatenation and ASTRAL; Fig. 4) supported Clade I with strong support, as was also identified in the Lovette tree. However, our UCE trees placed *B. lachrymosus* (represented by 4,815 UCE loci with only 8% missing data) sister to the clade uniting *Myioborus*, *Cardellina,* and the remaining *Basileuterus* species whereas the Lovette tree placed *B. lachrymosus* as sister to other *Basileuterus*. Both our mitogenome tree and the mitochondrial tree of Lovette et al. (2010) estimated a different position of *Setophaga*. The mitogenome tree did not recover Clade I, but it strongly supported a monophyletic *Basileuterus* (Fig. 7), placing *B. lachrymosus* sister to the clade comprising the "three-striped" warblers (*B. tristriatus* etc.), and similar to the mitochondrial tree in Lovette et al. (2010). The position of *B. lachrymosus* in the Lovette tree was likely driven by mitochondrial data, and mito-nuclear discordance sufficiently explains topological differences between UCE trees and the mitogenome trees. *B. lachrymosus*, the Fan-tailed Warbler, has been placed in the monotypic genus *Euthlypis* (e.g., Curson et al., 1994; Sibley and Monroe, 1990) owing to its unique morphology, distinctive side-to-side tail bobbing behavior, and foraging ecology (e.g., ant-following (Curson, 2020a) and commensal foraging (Komar and Hanks, 2002; Schaefer and Fagan, 2006)). Given the strong support across UCE analyses placing it sister to *Basileuterus* + *Cardellina* + *Myioborus*, we recommend resurrecting *Euthlypis* Cabanis, 1850.

Clade II (*Vermivora*, *Limnothlypis* and *Protonotaria*) – Clade II was recovered in all UCE trees, but not in the mitogenome tree or the Lovette tree. We observed topological incongruence for three low-diversity lineages, i.e., *Mniotilta*, *Parkesia,* and *Helmitheros*, in all of our analyses. The Lovette tree placed *Mniotilta* within Clade II, consistent with our mitogenome tree (Fig. 4; BS = 95, note that the node uniting everything only received 47). The ASTRAL tree, instead, placed *Mniotilta* sister to Clade II (quartet support = 0.93). However, concatenated UCE trees all grouped *Mniotilta* and Clade I together, with strong (BS = 100) to poor support (BS = 61). It has been suggested that *Mniotilta* may be closely related with *Setophaga*, given that viable hybrids were found between *Mniotilta varia* and *Setophaga* species, e.g., *S. coronata* (Vallender et al., 2009) and *S. cerulea* (Parkes, 1978), although intergeneric hybridization across the family is well-documented in Parulidae (e.g., https://avianhybrids.wordpress.com/parulidae/). Therefore, phylogenetic affinities of *Mniotilta* remain unsolved.

Clade III (*Geothlypis* and *Oporornis*) and Clade IV (*Leiothlypis* and *Oreothlypis*) – Clade III and Clade IV were grouped in all concatenated UCE trees and the Lovette tree, but not in our mitogenome or ASTRAL trees. For Clade III, the close relationships among *Geothlypis* and *Oporornis* species were found in previous phylogenetic studies (Escalante et al., 2009; Lovette et al., 2010). Lovette et al. (2010) found that *Geothlypis* and *Oporornis*, as formerly recognized, were paraphyletic, thus merged all *Geothlypis*, *Leucopeza,* and *Oporornis* into *Geothlypis*. Current taxonomy (e.g., Clements et al., 2023; Gill et al., 2023) only recognized a single species for *Oporornis*, the Connecticut Warbler *O. agilis*. Without inclusion of *Leucopeza*, *O. agilis* was consistently sister to *Geothlypis* in all of our trees with relatively shallow genomic divergence. However, it is unclear whether *O. agilis* should be merged into *Geothlypis* given that the position of *Leucopeza* remains unsupported (see below). For Clade IV, Lovette et al. (2010) merged all six *Leiothlypis* species into *Oreothlypis*, since they formed a well-supported clade. Our results are consistent with the current generic circumscription in that *Leiothlypis* and *Oreothlypis* formed two reciprocally monophyletic clades. However, given that the internal branch between *L. peregrina* and other *Leiothlypis* was very similar in length to the internal branch connecting *Oreothlypis* spp. and the most recent common ancestor (MRCA) of Clade IV, in both genetic divergence (e.g., Fig. 2) and coalescent (Fig. 3) units, it is reasonable to simplify the taxonomy and subsume *Leiothlypis* Sangster, 2008 in *Oreothlypis* Ridgway, 1884, although continued treatment as two separate genera is also valid.

Seiurus*, *Helmitheros* and *Parkesia – All UCE concatenated trees were consistent in recovering these three genera as the earliest diverging lineages, generally with strong support. The placement of *Seiurus* as the sister to all the other Parulidae was also supported by the ASTRAL tree and mitogenome tree, as well as the Lovette tree and previous phylogenetic work (e.g., Klein et al., 2004; Lovette and Bermingham, 2002). For *Helmitheros* and *Parkesia*, the ASTRAL tree showed a slightly different topology with *Helmitheros* sister to other parulids excluding *Parkesia and Seiurus*, although with a low quartet support of 0.66. The mitogenome tree placed *Parkesia* outside *Myiothlypis* plus *Myioborus*, *Cardellina,* and *Basileuterus* (BS = 59), whereas the Lovette tree placed *Parkesia* outside *Vermivora* plus *Mniotilta*, *Limnothlypis,* and *Protonotaria*,

but also without strong support.

4.3. Intrageneric relationships in Parulidae

For genera that comprise more than three species, we compared the intrageneric relationships produced by our analyses, the Lovette tree, and other previous work below.

Cardellina – The most consistent relationships were found for *Cardellina* across all of our trees with strong support, regardless of the data type and analytical method used. Both our UCE and mitogenome trees strongly supported *C. canadensis* and *C. pusilla* forming a clade sister to the rest of *Cardellina*. However, the Lovette tree separated the two species with *C. pusilla* being sister to the three red-colored *Cardellina* species though without strong support. Evidence from our trees and plumage coloration supported a sister relationship between *C. canadensis* and *C. pusilla*.

Myioborus – All of our trees (except the ASTRAL tree), the Lovette tree, and a previous study using three mitochondrial regions (Pérez-Emán, 2005) found that the two black and red redstarts, *Myioborus pictus* and *M. miniatus*, were the two earliest diverging lineages in *Myioborus*. However, the ASTRAL tree placed *M. ornatus* sister to the other *Myioborus*, likely due to a high level of missing data (65%) compared to the other well-sampled *Myioborus* species (on average ~7%). Although we sampled over 3,800 loci for *M. ornatus*, large amounts of missing data in each alignment may bias individual gene tree estimation and further lead to erroneous results when using gene tree summary methods (Hosner et al., 2016) such as ASTRAL used here. Our concatenated UCE trees placed *M. ornatus* sister to *M. albifrons*, and *M. melanocephalus* with strong support, whereas our mitogenome tree and the Lovette tree placed *M. albifrons* sister to *M. melanocephalus* and *M. ornatus* which was similar to Pérez-Emán (2005). Given the exceptionally short internal branches connecting these taxa (e.g., Fig. 2), it is worth revisiting this problem with an examination of factors such as incomplete lineage sorting and hybridization. For one poor-quality sample, *M. flavivertex*, our concatenated UCE tree (Supplementary Figure S4) found it being sister (BS = 90) to the well supported clade uniting *M. albifacies*, *M. cardonai* and *M. castaneocapilla*, in agreement with Pérez-Emán (2005). However, the mitogenome tree and the Lovette tree placed *M. flavivertex* outside the clade of *M. ornatus*, *M. albifrons* and *M. melanocephalus* with poor support (BS = 66; Fig. 7), possibly indicating cytonuclear discordance for these relationships. We had one unsampled species for this genus, the endangered *M. pariae* (Paria Whitestart), which was sister to the clade uniting *M. albifacies*, *M. cardonai* and *M. castaneocapilla* in the Lovette tree.

Basileuterus – *Basileuterus* was the genus with the least sampling in both our study and Lovette et al. (2010); our sampling lacked three species and Lovette et al. (2010) lacked five. For taxa that were sampled (excluding *B. lachrymosus*, see above), our trees and the Lovette tree recovered two groups: one containing species that all have red on the head such as the Golden-browed Warbler *B. belli*, and one containing the more drab-looking "three-striped" warblers (e.g., the *B. tristriatus* species complex, see also Gutiérrez-Pinto et al., 2012). The fair-quality sample *B. ignotus* sample was sister to the similar-looking *B. melanogenys* in both UCE and mitogenome trees. Topologies within the "three-striped" group differed across analyses. The concatenated UCE tree inferred from the good taxonset and the mitogenome tree grouped *B. tristriatus* and *B. punctipectus* as sister species, similar to Gutiérrez-Pinto et al. (2012) based only on ND2 sequences. The ASTRAL tree and other concatenated UCE trees supported *B. tristriatus* and *B. trifasciatus* as sister species. The Lovette tree grouped *B. melanotis* and *B. trifasciatus* together. Despite some topological differences, our concatenated UCE and mitogenome trees as well as the tree from Gutiérrez-Pinto et al. (2012) all supported *B. culicivorus* being basal to the rest of the "three-striped" group.

We had three unsampled *Basileuterus* species: *B. rufifrons* (Rufous-capped Warbler), *B. delattrii* (Chestnut-capped Warbler), and *B. melanotis* (Black-eared Warbler, a.k.a. Costa Rican Warbler). The Lovette tree sampled two of these and placed them as expected based on plumage similarity: *B. delattrii* (labeled as *B. rufifrons* in the Lovette tree but it was collected from Panama) was placed together with other *Basileuterus* warblers that have red on the head, and *B. melanotis* (labeled as *B. tristriatus* in the Lovette tree but it was collected from Panama) with other "three-striped" warblers. The taxonomic status of *B. rufifrons* and *B. delattrii* have long been debated and they were sometimes treated as conspecifics (Curson and García, 2020). Based on the taxonomic history of the unsampled *B. rufifrons*, it may be most closely related to *B. delattrii* but with some divergence in plumage coloration, vocalization and morphometric measurements (Demko et al., 2020).

Myiothlypis – Our study sampled all 18 currently recognized species for the genus *Myiothlypis*, including three bad and four ugly samples. The concatenated UCE and ASTRAL trees inferred from the good taxonset were largely congruent except for the position of *M. signata*, a sample with relatively high missing data (56%) and an exaggerated branch length. As discussed above, concatenation may yield less biased topologies for a taxon with rampant missing data than gene tree reconciliation. Supporting this, we found that *M. signata* was sister to *M. nigrocristata* in all concatenated UCE trees and the Lovette tree but not in the ASTRAL tree. The good taxonset trees (Figs. 2 & 3) separated *Myiothlypis* into two well-supported major groups: clade i) six species – *M. flaveola*, *M. leucoblephara*, *M. rivularis*, *M. nigrocristata*, *M. signata* and *M. luteoviridis*; and clade ii) five species – *M. bivittata*, *M. chrysogaster*, *M. roraimae*, *M. coronata* and *M. fraseri*. These two clades were also present in our mitogenome tree and the Lovette tree though with slight in-group rearrangements. After adding the seven fair- and poor-quality *Myiothlypis* samples to the concatenation analysis, all but two clustered together with long terminal branches (Supplementary Figure S4); *M. fulvicauda* was sister to *M. rivularis*, and *M. cinereicollis* erroneously fell sister to the rest of Parulidae. However, in both our mitogenome tree and the Lovette tree, *M. leucophrys* fell into clade i, and *M. cinereicollis* was sister to *M. conspicillata* within clade ii; *M. fulvicauda* was consistently sister to *M. rivularis*, which is expected given their shared habitat preference for forest rivers and streams (Curson 2020b; Curson et al., 2022). For those that were sampled in the mitogenome tree but not in the Lovette tree, *M. griseiceps* and *M. basilica* were nested in clade i (not as sister species), and *M. chlorophrys* in clade ii. Although the support values for some of the internal nodes were low in the mitogenome tree, clade i and clade ii each had full support. Therefore, these fair and poor quality samples (except for *M. cinereicollis*) were likely to be placed in the correct subgroup, but the cluster of *M. griseiceps* and *M. basilica* in the good + bad taxonset tree (Fig. 6) and the cluster of the five samples in the good + bad + ugly taxonset tree (Supplementary Figure S4) could be explained by long branch attraction, similarly to the UCE tree based on the unfiltered dataset (Fig. 1).

Setophaga – *Setophaga* is the most speciose genus in Parulidae, with 37 species recognized in IOC 13.1 (36 sampled in our study). Relationships within *Setophaga* were somewhat variable across analyses (Supplementary Figure S7). The concatenated UCE trees based on good + bad and good + bad + ugly taxonsets yielded the exact same topologies within *Setophaga*; they were generally congruent with the concatenated good taxonset tree except for some rearrangement of *S. caerulescens* and *S. kirtlandii*. The ASTRAL tree largely agreed with the concatenated UCE trees, but it did not recover the clade comprising the three island endemic species (*S. pharetra*, *S. angelae* and *S. plumbea*). However, this clade was also present in the mitogenome tree and the Lovette tree. Additionally, *S. delicata* was likely placed in the wrong position of the ASTRAL tree due to its high proportion of missing data (66%), whereas the other analyses including the Lovette tree all grouped *S. delicata* and *S. subita* as sisters. These two formed a clade with *S. adelaidae*, representing a second independent island radiation in the Caribbean. Some other interesting topological incongruences were also noted across analyses, for example, our UCE trees strongly supported a sister relationship between the widespread *S. palmarum* and the *S. coronata* species complex, but this was not supported by the

mitogenome tree or the Lovette tree. *S. cerulea* was found to be sister to the two parulas (*S. americana* and *S. pitiayumi*) in our concatenated UCE trees and the Lovette tree, but not in the mitogenome or ASTRAL trees. Although *S. citrina* and *S. ruticilla* differ strikingly in plumage coloration, they formed a strongly supported clade in all of our trees, but not in the Lovette tree. These species each employ similar foraging behaviors leveraging their long, strikingly patterned tails (green with white spots in *S. citrina*, black with red spots in *S. ruticilla*) to flush and startle insects. Within Parulidae, this foraging strategy has evolved convergently in *Myioborus* and *Basileuterus lachrymosus*.

The unsampled *S. goldmani* is part of the *S. coronata* species complex and has often been treated as a subspecies of *S. coronata* (e.g., Clements et al., 2023). It is the most morphologically, geographically, and genetically distinctive population in the species complex (Brelsford et al., 2011; Toews et al., 2016). Therefore, we expect *S. goldmani* to cluster together with *S. coronata sensu stricto* and *S. auduboni*, and likely to be a sister taxon.

Leiothlypis – Our concatenated and ASTRAL UCE trees identified congruent relationships within *Leiothlypis*. *L. peregrina* consistently resolved sister to the remaining *Leiothlypis*, which was also recovered in the mitogenome tree and the Lovette tree. However, the position of *L. ruficapilla*, Nashville Warbler, varied between UCE and mitochondrial analyses, as well as between our trees and the Lovette tree. All UCE trees placed *L. ruficapilla* sister to *L. luciae*, *L. virginiae* and *L. crissalis*, whereas the mitogenome tree placed *L. ruficapilla* sister to *L. luciae* and *L. virginiae*. Lovette et al. (2010) found *L. ruficapilla* sister to *L. virginiae*. Interestingly, a previous phylogenetic study (Klein et al., 2004) sampled the mitochondrial cytochrome *b* sequence for four individuals of *L. ruficapilla* and found the two western individuals collected in California (*L. r. ridgwayi*) sister to *L. luciae* and *L. virginiae* rather than the two eastern individuals (one known to be collected in Louisiana; *L. r. ruficapilla*). The same pattern was also observed for the sample used in the Lovette tree (Washington State) and our sample (UF49581) that was collected off the US East Coast. The two subspecies of *L. ruficapilla* breed in disjunct locations and differ in plumage coloration, vocalizations, tail-bobbing behavior (present only in *L. r. ridgwayi*), and morphometrics (Lowther and Williams, 2020). Collectively, these results and observations suggest deep divergences between *L. ruficapilla* populations, which could be recognized as separate species.

Geothlypis – Our UCE analyses yielded consistent relationships within *Geothlypis*, recovering three major groups that generally agreed with the Lovette tree: 1) a group of yellowthroats that predominantly occur in the North America including the widespread Common Yellowthroat *G. trichas*; 2) *G. formosa*, the Kentucky Warbler; and 3) a group of yellowthroats that occur in the Central and South America, plus the two "gray-hooded" warblers *G. philadelphia* and *G. tolmiei*. Our mitogenome tree placed *G. philadelphia* and *G. tolmiei* sister to other *Geothlypis*, thus recovering four major lineages, similar to results in Escalante et al. (2009) inferred from three mitochondrial regions. Adding the fair-quality *G. auricularis* sample into the concatenated analyses, we found that the UCE trees (Fig. 6, Supplementary Figure S4) supported a sister relationship between *G. auricularis* and *G. aequinoctialis* (BS = 100) although *G. auricularis* featured an exceptionally long terminal branch due to its high level of missing and/or poor-quality data. Neither *G. auricularis* nor *G. aequinoctialis* were sampled in the Lovette tree (they sampled *G. [aequinoctialis] velata*), but their sister relationship was also supported by our mitogenome tree (BS = 100).

4.4. Phylogenetics of extinct and near-extinct parulids

The sister relationship between *Oporornis agilis* and likely extinct *Leucopeza semperi* were supported by the mitogenome tree (BS = 97) and the Lovette tree, however both trees were based on the same ND2 sequences of *L. semperi*, thus lacking the power to independently resolve its phylogenetic position. The concatenated UCE tree (Supplementary Figure S5) placed *L. semperi* sister to the clade uniting *Myioborus*, *Cardellina*, *Basileuterus*, and *Myiothlypis* with moderate support (BS = 86), but this was based on 79 UCE loci. Given such limited data and strong cytonuclear discordance, the placement of *Leucopeza* remained unclear.

Vermivora bachmanii is another extinct species in Parulidae. Because UCEs were extracted from old museum skins (see Smith et al., 2021), the five samples of *V. bachmanii* all had a high proportion of missing data. Despite that, our analyses consistently supported a sister relationship between *V. bachmanii* and the clade containing its two extant congeners, *V. chrysoptera* and *V. cyanoptera*, which are famous for their extensive hybridization (e.g., Confer et al., 2020; Gill et al., 2020; Shapiro et al., 2004). This sister relationship was also supported by the Lovette tree based on a single ND2 sequence of *V. bachmanii* and by a recent genomic study (Wood et al., 2023).

Catharopeza bishopi is an endangered species endemic to St. Vincent of Lesser Antilles. Our concatenation analysis (Supplementary Figure S6) found that it was nested within *Setophaga* (BS = 79), sister to a diverse clade to the exclusion of the other island-dwelling, dark gray and white plumaged warblers *S. pharetra* (Jamaica) and *S. angelae* (Puerto Rico) (and *S. plumbea* [Guadeloupe and Dominica of Lesser Antilles], Fig. 6), although this was only based on 296 UCE loci. The mitogenome tree, which used the same six mitochondrial regions as in Lovette et al. (2010) for *C. bishopi*, strongly supported (BS = 100) a sister relationship between *Catharopeza* and *Setophaga*, in agreement with the Lovette tree. In comparison, Baiz et al. (2021), which focused on color evolution in *Setophaga*, placed *Catharopeza* within the island-dwelling clade and sister to *S. plumbea*, though the clade received low support (Baiz et al., 2021). Although the position of *Catharopeza* could not be fully resolved, it is likely to be most closely related with *Setophaga*. Nevertheless, our results further support the hypothesis that these island endemics may be a part of the historically widespread island radiation in the Caribbean (Ricklefs and Cox, 1972).

5. Conclusion

Achieving accurate, well-resolved phylogenies can be a challenging goal, particularly with heterogeneous sample quality, as was available for Parulidae. Thus, it is important to develop approaches that allow researchers to leverage any data they have to best identify relationships likely to represent the true species tree. Here, we found that iteratively adding data from lower-quality samples mitigated some obvious problems in phylogenetic inference. This allowed us to infer a strong, conservative backbone hypothesis of relationships with confidence (the good taxonset tree). Of the 10 fair-quality samples that we added in for the good + bad taxonset, the estimated positions of nine samples were coherent based on previous literature. The remaining species, *Myiothlypis griseiceps*, has not been included in previous molecular studies and its placement could not be validated with other data. Thus, the good + bad taxonset tree should be a good estimate of the true underlying species tree. Inclusion of poor-quality samples (for the good + bad + ugly taxonset) and the placements of *Leucopeza semperi* and *Catharopeza bishopi* are more speculative, but we have presented initial hypotheses for their relationships to be for validation by future studies. Overall, while every dataset may differ, it can be beneficial to consider and explore different analytical strategies to achieve high-quality trees of life. Our Parulidae phylogenies provide a solid backbone for understanding the trait and molecular evolution, biogeographic patterns, diversification and changes in biodiversity over time of this charismatic bird group.

CRediT authorship contribution statement

Min Zhao: Writing – original draft, Validation, Investigation, Formal analysis. **Jessica A. Oswald:** Writing – original draft, Project administration, Methodology, Investigation. **Julie M. Allen:** Writing – review & editing, Resources, Methodology, Formal analysis. **Hannah L. Owens:**

Writing – review & editing, Investigation, Funding acquisition, Conceptualization. **Peter A. Hosner:** Writing – review & editing, Methodology, Formal analysis. **Robert P. Guralnick:** Writing – review & editing, Supervision, Funding acquisition, Conceptualization. **Edward L. Braun:** Writing – review & editing, Methodology, Funding acquisition. **Rebecca T. Kimball:** Writing – original draft, Supervision, Project administration, Methodology, Investigation, Funding acquisition, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was supported by funding from the United States National Science Foundation (DEB 1655683 to RTK and ELB; NSF DEB-2033905 to RPG; NSF DEB-2034316 to JAO, JMA), and support from a UF Biodiversity Institute seed grant to HLO and RPG. PAH and HLO acknowledge funding from the Villum Foundation (grant no. 25925). We thank the museums that provided tissues that were critical for the success of this project: American Museum of Natural History (AMNH), Field Museum of Natural History (FMNH), Florida Museum of Natural History (UF), University of Kansas Biodiversity Institute (KU), Louisiana State University Museum of Natural History (LSUMZ), Museum of Southwestern Biology (MSB), Royal Ontario Museum (ROM), Texas A&M University Biodiversity Research and Teaching Collections (TCWC), National Museum of Natural History, Smithsonian Institution (USNM), University of Washington Burke Museum (UWBM). We thank Gordon Burleigh for assistance in collapsing tree clades. We thank Hanchen Huang for help in tree visualization. We are grateful to Hanyang Ye and Dan Liang for providing the warbler photos that were used in the graphical abstract. Two warbler photos are from Wikimedia Commons. Photo credit can be found in Supplementary Table S2. We thank two anonymous reviewers whose comments improved this manuscript.

Disclaimer

The findings and conclusions in this article are those of the author(s) and do not necessarily represent the views of the U.S. Fish and Wildlife Service.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ympev.2024.108235.

Data availability

All data used in this study (alignments, models, and tree files) are available on figshare (DOI: 10.6084/m9.figshare.25504699).

References

Abascal, F., Zardoya, R., Telford, M.J., 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38, W7–W. https://doi.org/10.1093/nar/gkq291.

Allio, R., Schomaker-Bastos, A., Romiguier, J., Prosdocimi, F., Nabholz, B., Delsuc, F., 2020. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. Mol. Ecol. Resour 20, 892–905. https://doi.org/10.1111/1755-0998.13160.

Baiz, M.D., Wood, A.W., Brelsford, A., Lovette, I.J., Toews, D.P.L., 2021. Pigmentation genes show evidence of repeated divergence and multiple bouts of introgression in *Setophaga* warblers. Curr. Biol. 31, 643–649.e3. https://doi.org/10.1016/j.cub.2020.10.094.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V. M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., Pyshkin, A.V., Sirotkin, A.V., Vyahhi, N., Tesler, G., Alekseyev, M.A., Pevzner, P.A., 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19, 455–477. https://doi.org/10.1089/cmb.2012.0021.

Barker, F.K., Burns, K.J., Klicka, J., Lanyon, S.M., Lovette, I.J., 2013. Going to extremes: contrasting rates of diversification in a recent radiation of new world passerine birds. Syst. Biol. 62, 298–320. https://doi.org/10.1093/sysbio/sys094.

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. https://doi.org/10.1642/AUK-14-110.1.

Barrera-Guzmán, A.O., Milá, B., Sánchez-González, L.A., Navarro-Sigüenza, A.G., 2012. Speciation in an avian complex endemic to the mountains of Middle America (*Ergaticus*, Aves: Parulidae). Mol. Phylogenet. Evol. 62, 907–920. https://doi.org/10.1016/j.ympev.2011.11.020.

Bergsten, J., 2005. A review of long-branch attraction. Cladistics 21, 163–193. https://doi.org/10.1111/j.1096-0031.2005.00059.x.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Braun, E.L., Oliveros, C.H., White Carreiro, N.D., Zhao, M., Glenn, T.C., Brumfield, R.T., Braun, M.J., Kimball, R.T., Faircloth, B.C., 2024. Testing the mettle of METAL: A comparison of phylogenomic methods using a challenging but well-resolved phylogeny. BioRxiv. https://doi.org/10.1101/2024.02.28.582627.

Brelsford, A., Mila, B., Irwin, D.E., 2011. Hybrid origin of Audubon's warbler. Molecular Ecology 20, 2380–2389.

Carter, J.K., Kimball, R.T., Funk, E.R., Kane, N.C., Schield, D.R., Spellman, G.M., Safran, R.J., 2023. Estimating phylogenies from genomes: A beginners review of commonly used genomic data in vertebrate phylogenomics. J. Hered. 114, 1–13. https://doi.org/10.1093/jhered/esac061.

Chen, D., Braun, E.L., Forthman, M., Kimball, R.T., Zhang, Z., 2018. A simple strategy for recovering ultraconserved elements, exons, and introns from low coverage shotgun sequencing of museum specimens: Placement of the partridge genus *Tropicoperdix* within the Galliformes. Mol. Phylogenet. Evol. 129, 304–314. https://doi.org/10.1016/j.ympev.2018.09.005.

Chesser, R.T., Banks, R.C., Barker, F.K., Cicero, C., Dunn, J.L., Kratter, A.W., Lovette, I.J., Rasmussen, P.C., Remsen, J.V., Rising, J.D., Stotz, D.F., Winker, K., 2011. Fifty-Second Supplement to the American Ornithologists' Union Check-list of North American Birds. Auk 128, 600–613. https://doi.org/10.1525/auk.2011.128.3.600.

Chesser, R.T., Burns, K.J., Cicero, C., Dunn, J.L., Kratter, A.W., Lovette, I.J., Rasmussen, P.C., Remsen, J.V., Rising, J.D., Stotz, D.F., Winker, K., 2016. Fifty-seventh Supplement to the American Ornithologists' Union Check-list of North American Birds. Auk 133, 544–560. https://doi.org/10.1642/AUK-16-77.1.

Chesser, R.T., Burns, K.J., Cicero, C., Dunn, J.L., Kratter, A.W., Lovette, I.J., Rasmussen, P.C., Remsen, J.V., Stotz, D.F., Winger, B.M., Winker, K., 2018. Fifty-ninth Supplement to the American Ornithological Society's Check-list of North American Birds. Auk 135, 798–813. https://doi.org/10.1642/AUK-18-62.1.

Chesser, R.T., Burns, K.J., Cicero, C., Dunn, J.L., Kratter, A.W., Lovette, I.J., Rasmussen, P.C., Remsen, J.V., Stotz, D.F., Winker, K., 2019. Sixtieth Supplement to the American Ornithological Society's Check-list of North American Birds. Auk 136. https://doi.org/10.1093/auk/ukz042.

Clements, J.F., Rasmussen, P.C., Schulenberg, T.S., Iliff, M.J., Fredericks, T.A., Gerbracht, J.A., Lepage, D., Spencer, A., Billerman, S.M., Sullivan, B.L., Wood, C.L., 2023. The eBird/Clements checklist of Birds of the World: v2023. https://www.birds.cornell.edu/clementschecklist/download/ (accessed 4.11.23).

Confer, J.L., Hartman, P., Roth, A., 2020. Golden-winged Warbler (*Vermivora chrysoptera*), in: Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.gowwar.01.

Curson, J., García, N.C., 2020. Rufous-capped Warbler (*Basileuterus rufifrons*), in: Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.rucwar.01.1.

Curson, J., J. del Hoyo, N. Collar, G. M. Kirwan, and P. F. D. Boesman, 2022. Riverbank Warbler (*Myiothlypis rivularis*), in: Keeney, B.K. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.rivwar1.01.

Curson, J., Quinn, D., Beadle, D., 1994. New World Warblers. Christopher Helm Publishers, London.

Curson, J., 2020a. Fan-tailed Warbler (*Basileuterus lachrymosus*), in: Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.fatwar.01.

Curson, J., 2020b. Buff-rumped Warbler (*Myiothlypis fulvicauda*), in: del Hoyo J., Elliott A., Sargatal J., Christie D. A., and de Juana E. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.burwar1.01.

Demko, A.D., Sosa-López, J.R., Simpson, R.K., Doucet, S.M., Mennill, D.J., 2020. Divergence in plumage, voice, and morphology indicates speciation in Rufous-capped Warblers (*Basileuterus rufifrons*). The Auk 137, ukaa029.

Escalante, P., Márquez-Valdelamar, L., de la Torre, P., Laclette, J.P., Klicka, J., 2009. Evolutionary history of a prominent North American warbler clade: the *Oporornis-Geothlypis* complex. Mol. Phylogenet. Evol. 53, 668–678. https://doi.org/10.1016/j.ympev.2009.07.014.

Faircloth, B.C., 2013. Illumiprocessor - software for Illumina read quality filtering. Brant Faircloth. https://doi.org/10.6079/j9ill.

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646.

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. https://doi.org/10.1093/sysbio/sys004.

Gilbert, P.S., Wu, J., Simon, M.W., Sinsheimer, J.S., Alfaro, M.E., 2018. Filtering nucleotide sites by phylogenetic signal to noise ratio increases confidence in the Neoaves phylogeny generated from ultraconserved elements. Mol. Phylogenet. Evol. 126, 116–128. https://doi.org/10.1016/j.ympev.2018.03.033.

Gill, F.B., Canterbury, R.A., Confer, J.L., 2020. Blue-winged Warbler (*Vermivora cyanoptera*), in: Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (Eds.), Birds of the World. Cornell Lab of Ornithology. Doi: 10.2173/bow.buwwar.01.

Gill, F., Donsker, D., Rasmussen, P., 2023. IOC World Bird List (v 13.1). https://www.worldbirdnames.org/new/ioc-lists/master-list-2/ (accessed 4.11.23).

Gutiérrez-Pinto, N., Cuervo, A.M., Miranda, J., Pérez-Emán, J.L., Brumfield, R.T., Cadena, C.D., 2012. Non-monophyly and deep genetic differentiation across low-elevation barriers in a Neotropical montane bird (*Basileuterus tristriatus*; Aves: Parulidae). Mol. Phylogenet. Evol. 64, 156–165. https://doi.org/10.1016/j.ympev.2012.03.011.

Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., Vinh, L.S., 2018. Ufboot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35, 518–522. https://doi.org/10.1093/molbev/msx281.

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. 33, 1110–1125. https://doi.org/10.1093/molbev/msv347.

Hosner, P.A., Tobias, J.A., Braun, E.L., Kimball, R.T., 2017. How do seemingly non-vagile clades accomplish trans-marine dispersal? Trait and dispersal evolution in the landfowl (Aves: Galliformes). Proc. Biol. Sci. 284. https://doi.org/10.1098/rspb.2017.0210.

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285.

Katoh, K., Standley, D.M., 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 772–780. https://doi.org/10.1093/molbev/mst010.

Kimball, R.T., Wang, N., Heimer-McGinn, V., Ferguson, C., Braun, E.L., 2013. Identifying localized biases in large datasets: a case study using the avian tree of life. Mol. Phylogenet. Evol. 69, 1021–1032. https://doi.org/10.1016/j.ympev.2013.05.029.

Klein, N.K., Burns, K.J., Hackett, S.J., Griffiths, C.S., 2004. Molecular phylogenetic relationships among the wood warblers (Parulidae) and historical biogeography in the Caribbean basin. Journal of Caribbean Ornithology 17, 3–17.

Komar, O., Hanks, C.K., 2002. Fan-tailed Warbler foraging with nine-banded armadillos. The Wilson Bulletin 114, 526–528.

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923.

Leroy, H., Bowie, R.C.K., Rubáčová, L., Matysioková, B., Remeš, V., 2024. A late burst of colour evolution in a radiation of songbirds (Passeriformes: Parulidae) suggests secondary contact drives signal divergence. J. Evol. Biol. https://doi.org/10.1093/jeb/voae023.

Lovette, I.J., Bermingham, E., 2002. What is a Wood-Warbler? Molecular Characterization of a Monophyletic Parulidae. Auk 119, 695–714. https://doi.org/10.1093/auk/119.3.695.

Lovette, I.J., Pérez-Emán, J.L., Sullivan, J.P., Banks, R.C., Fiorentino, I., Córdoba-Córdoba, S., Echeverry-Galvis, M., Barker, F.K., Burns, K.J., Klicka, J., Lanyon, S.M., Bermingham, E., 2010. A comprehensive multilocus phylogeny for the wood-warblers and a revised classification of the Parulidae (Aves). Mol. Phylogenet. Evol. 57, 753–770. https://doi.org/10.1016/j.ympev.2010.07.018.

Lowther, P.E., Williams, J.M., 2020. Nashville Warbler (*Leiothlypis ruficapilla*). In: Billerman, S.M., Keeney, B.K., Rodewald, P.G., Schulenberg, T.S. (Eds.), Birds of the World. Cornell Lab of Ornithology. https://doi.org/10.2173/bow.naswar.01.

MacArthur, R.H., 1958. Population ecology of some warblers of northeastern coniferous forests. Ecology 39, 599–619. https://doi.org/10.2307/1931600.

Marçais, G., Yorke, J.A., Zimin, A., 2015. Quorum: an error corrector for illumina reads. PLoS ONE 10, e0130821.

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.E., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol. Ecol. Resour 16, 1189–1203. https://doi.org/10.1111/1755-0998.12466.

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015.

Mitchell, L.R., Benedict, L., Cavar, J., Najar, N., Logue, D.M., 2019. The evolution of vocal duets and migration in New World warblers (Parulidae). Auk 136. https://doi.org/10.1093/auk/ukz003.

Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32, 268–274. https://doi.org/10.1093/molbev/msu300.

Oliveros, C.H., Field, D.J., Ksepka, D.T., Barker, F.K., Aleixo, A., Andersen, M.J., Alström, P., Benz, B.W., Braun, E.L., Braun, M.J., Bravo, G.A., Brumfield, R.T., Chesser, R.T., Claramunt, S., Cracraft, J., Cuervo, A.M., Derryberry, E.P., Glenn, T.C., Harvey, M.G., Hosner, P.A., Faircloth, B.C., 2019. Earth history and the passerine superradiation. Proc Natl Acad Sci USA 116, 7916–7925. https://doi.org/10.1073/pnas.1813206116.

Oswald, J.A., Harvey, M.G., Remsen, R.C., Foxworth, D.U., Cardiff, S.W., Dittmann, D.L., Megna, L.C., Carling, M.D., Brumfield, R.T., 2016. Willet be one species or two? A genomic view of the evolutionary history of *Tringa semipalmata*. Auk 133, 593–614. https://doi.org/10.1642/AUK-15-232.1.

Oswald, J.A., Smith, B.T., Allen, J.M., Guralnick, R.P., Steadman, D.W., LeFebvre, M.J., 2023. Changes in parrot diversity after human arrival to the Caribbean. Proc Natl Acad Sci USA 120. https://doi.org/10.1073/pnas.2301128120 e2301128120.

Parkes, K.C., 1978. Still another Parulid intergeneric hybrid (*Mniotilta* × *Dendroica*) and its taxonomic and evolutionary implications. The Auk 95, 682–690.

Pérez-Emán, J.L., 2005. Molecular phylogenetics and biogeography of the Neotropical redstarts (*Myioborus*; Aves, Parulinae). Mol. Phylogenet. Evol. 37, 511–528. https://doi.org/10.1016/j.ympev.2005.04.013.

Ricklefs, R.E., Cox, G.W., 1972. Taxon cycles in the West Indian avifauna. The American Naturalist 106, 195–219.

Salter, J.F., Hosner, P.A., Tsai, W.L.E., McCormack, J.E., Braun, E.L., Kimball, R.T., Brumfield, R.T., Faircloth, B.C., 2022. Historical specimens and the limits of subspecies phylogenomics in the New World quails (Odontophoridae). Mol. Phylogenet. Evol. 175, 107559. https://doi.org/10.1016/j.ympev.2022.107559.

Sayyari, E., Whitfield, J.B., Mirarab, S., 2017. Fragmentary gene sequences negatively impact gene tree and species tree reconstruction. Mol. Biol. Evol. 34, 3279–3291. https://doi.org/10.1093/molbev/msx261.

Sayyari, E., Whitfield, J.B., Mirarab, S., 2018. DiscoVista: Interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. 122, 110–115. https://doi.org/10.1016/j.ympev.2018.01.019.

Schaefer, R.R., Fagan, J.F., 2006. Commensal foraging by a fan-tailed warbler (*Euthlypis lachrymosa*) with a nine-banded armadillo (*Dasypus novemcinctus*) in southwestern Mexico. The Southwestern Naturalist 560–562.

Schmieder, R., Edwards, R., 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864. https://doi.org/10.1093/bioinformatics/btr026.

Shapiro, L.H., Canterbury, R.A., Stover, D.M., Fleischer, R.C., 2004. Reciprocal introgression between Golden-winged Warblers (*Vermivora chrysoptera*) and Blue-winged Warblers (*V. pinus*) in eastern North America. The Auk 121, 1019–1030.

Shen, X.-X., Hittinger, C.T., Rokas, A., 2017. Contentious relationships in phylogenomic studies can be driven by a handful of genes. Nat. Ecol. Evol. 1, 126. https://doi.org/10.1038/s41559-017-0126.

Sibley, C.G., Monroe, B.L., 1990. Distribution and taxonomy of birds of the world. Yale University Press, New Haven.

Simmons, M.P., 2014. A confounding effect of missing data on character conflict in maximum likelihood and Bayesian MCMC phylogenetic analyses. Mol. Phylogenet. Evol. 80, 267–280. https://doi.org/10.1016/j.ympev.2014.08.021.

Simpson, R.K., Mistakidis, A.F., Doucet, S.M., 2020. Natural and sexual selection shape the evolution of colour and conspicuousness in North American wood-warblers (Parulidae). Biological Journal of the Linnean Society 130, 89–100. https://doi.org/10.1093/biolinnean/blaa015.

Simpson, R.K., Wilson, D.R., Mistakidis, A.F., Mennill, D.J., Doucet, S.M., 2021. Sympatry drives colour and song evolution in wood-warblers (Parulidae). Proc. Biol. Sci. 288, 20202804. https://doi.org/10.1098/rspb.2020.2804.

Smith, S.A., Dunn, C.W., 2008. Phyutility: A phyloinformatics tool for trees, alignments and molecular data. Bioinformatics 24, 715–716. https://doi.org/10.1093/bioinformatics/btm619.

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Syst. Biol. 63, 83–95. https://doi.org/10.1093/sysbio/syt061.

Smith, B.T., Mauck, W.M., Benz, B.W., Andersen, M.J., 2020. Uneven Missing Data Skew Phylogenomic Relationships within the Lories and Lorikeets. Genome Biol. Evol. 12, 1131–1147. https://doi.org/10.1093/gbe/evaa113.

Smith, B.T., Gehara, M., Harvey, M.G., 2021. The demography of extinction in eastern North American birds. Proc. Biol. Sci. 288, 20201945. https://doi.org/10.1098/rspb.2020.1945.

Sukumaran, J., Holder, M.T., 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26, 1569–1571. https://doi.org/10.1093/bioinformatics/btq228.

Sun, K., Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2014. The evolution of peafowl and other taxa with ocelli (eyespots): a phylogenomic approach. Proc. Biol. Sci. 281. https://doi.org/10.1098/rspb.2014.0823.

Toews, D.P., Brelsford, A., Grossen, C., Milá, B., Irwin, D.E., 2016. Genomic variation across the Yellow-rumped Warbler species complex. The Auk: Ornithological Advances 133, 698–717.

Tsai, W.L.E., Schedl, M.E., Maley, J.M., McCormack, J.E., 2020. More than skin and bones: Comparing extraction methods and alternative sources of DNA from avian museum specimens. Mol. Ecol. Resour 20, 1220–1227. https://doi.org/10.1111/1755-0998.13077.

Vallender, R., Gagnon, J.-P., Lovette, I., 2009. An intergeneric wood-warbler hybrid (*Mniotilta varia* × *Dendroica coronata*) and use of multilocus DNA analyses to diagnose avian hybrid origins. The Wilson Journal of Ornithology 121, 298–305. https://doi.org/10.1676/08-050.1.

Vilaça, S.T., Santos, F.R., 2010. Biogeographic history of the species complex *Basileuterus culicivorus* (Aves, Parulidae) in the Neotropics. Mol. Phylogenet. Evol. 57, 585–597. https://doi.org/10.1016/j.ympev.2010.07.010.

Wang, N., Hosner, P.A., Liang, B., Braun, E.L., Kimball, R.T., 2017. Historical relationships of three enigmatic phasianid genera (Aves: Galliformes) inferred using

phylogenomic and mitogenomic data. Mol. Phylogenet. Evol. 109, 217–225. https://doi.org/10.1016/j.ympev.2017.01.006.
Winger, B.M., Lovette, I.J., Winkler, D.W., 2012. Ancestry and evolution of seasonal migration in the Parulidae. Proc. Biol. Sci. 279, 610–618. https://doi.org/10.1098/rspb.2011.1045.
Wood, A.W., Szpiech, Z.A., Lovette, I.J., Smith, B.T., Toews, D.P., 2023. Genomes of the extinct Bachman's warbler show high divergence and no evidence of admixture with other extant *Vermivora* warblers. Current Biology. 33, 2823–2829. https://doi.org/10.1016/j.cub.2023.05.058.
Zhang, C., Mirarab, S., 2022. Weighting by Gene Tree Uncertainty Improves Accuracy of Quartet-based Species Trees. Mol. Biol. Evol. 39. https://doi.org/10.1093/molbev/msac215.
Zhang, C., Rabiee, M., Sayyari, E., Mirarab, S., 2018. ASTRAL-III: Polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics 19, 153. https://doi.org/10.1186/s12859-018-2129-y.
Zhao, M., Kurtis, S.M., White, N.D., Moncrieff, A.E., Leite, R.N., Brumfield, R.T., Braun, E.L., Kimball, R.T., 2023. Exploring conflicts in whole genome phylogenetics: A case study within manakins (Aves: Pipridae). Syst. Biol. 72, 161–178. https://doi.org/10.1093/sysbio/syac062.