UNIVERSITY OF HELSINKI

https://helda.helsinki.fi

Phylogenomics of manakins (Aves: Pipridae) using alternative locus filtering strategies based on informativeness

Leite, Rafael N.; Kimball, Rebecca T.; Braun, Edward L.; Derryberry, Elizabeth P.; Hosner, Peter A. ...

2021-02

Academic Press Inc.

http://hdl.handle.net/10138/338055

Leite, R N, Kimball, R T, Braun, E L, Derryberry, E P, Hosner, P A, Derryberry, G E, Anciaes, M, McKay, J S, Aleixo, A, Ribas, C C, Brumfield, R T & Cracraft, J 2021, 'Phylogenomics of manakins (Aves: Pipridae) using alternative locus filtering strategies based on informativeness', Molecular Phylogenetics and Evolution, vol. 155, 107013. https://doi.org/10.1016/j.ympev.2020.107013

Downloaded from Helda, University of Helsinki institutional repository. https://helda.helsinki.fi
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.
Please cite the original version.

Phylogenomics of manakins (Aves: Pipridae) using alternative locus filtering strategies based on informativeness

Rafael N. Leite [a, *, 1], Rebecca T. Kimball [b], Edward L. Braun [b], Elizabeth P. Derryberry [c], Peter A. Hosner [b], Graham E. Derryberry [d], Marina Anciães [e], Jessica McKay [f], Diogo Meyer [a], Alexandre Aleixo [g], Camila C. Ribas [e], Robb T. Brumfield [d], Joel Cracraft [f]

[a] Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, SP, Brazil
[b] Department of Biology, University of Florida, Gainesville, FL, USA
[c] Department of Ecology and Evolutionary Biology, Tulane University, New Orleans, LA, USA
[d] Museum of Natural Science, Louisiana State University, Baton Rouge, LA, USA
[e] Coordenação de Biodiversidade, Instituto Nacional de Pesquisas da Amazônia, Manaus, AM, Brazil
[f] Department of Ornithology, American Museum of Natural History, New York, NY, USA
[g] Department of Zoology, Museu Paraense Emílio Goeldi, Belém, PA, Brazil

* Corresponding author.
E-mail addresses: rnleite@gmail.com (R.N. Leite), rkimball@ufl.edu (R.T. Kimball), ebraun68@ufl.edu (E.L. Braun), ederrybe@tulane.edu (E.P. Derryberry), idioptilon@gmail.com (P.A. Hosner), g.derryberry@gmail.com (G.E. Derryberry), marina.anciaes@gmail.com (M. Anciães), jessicaalexmckay@gmail.com (J. McKay), diogo@ib.usp.br (D. Meyer), alexandre.aleixo@helsinki.fi (A. Aleixo), camilaribas@gmail.com (C.C. Ribas), robb@lsu.edu (R.T. Brumfield), jlc@amnh.org (J. Cracraft).
[1] *Present address:* Programa de Pós-Graduação em Ecologia, Instituto Nacional de Pesquisas da Amazônia, Avenida André Araújo 2936, Petrópolis, Manaus, AM, 69.067-375, Brazil.

Declaration of interest: None.

Highlights

- Target capture provides thousands of loci to reconstruct the phylogeny of manakins.
- Sequence data from ultraconserved elements resolve basal relationships.
- Recovery of sequences is more efficient with UCE probes than with exon probes.
- Locus filtering based on informative sites has an impact on species tree estimation.
- Support for recalcitrant nodes can be examined using alternative filtering schemes.

Abstract

Target capture sequencing have been used increasingly to generate several unlinked loci for different types of markers that are useful for molecular systematics. Phylogenomic studies have employed concatenation and multispecies coalescent approaches to estimate phylogenetic trees from data sets that are usually more heterogeneous and contain less phylogenetic information per each gene. Locus filtering may improve species tree estimation due to the removal of loci with little phylogenetic signal under these circumstances. We addressed some of the challenges to empirical phylogenomic data sets by analyzing thousands of loci obtained from sequence capture of ultraconserved elements as well as exons and their flanking regions for the avian family Pipridae (manakins). Manakins are a group of small suboscine passerine birds, which early phylogenetic hypotheses were based on behavioral and morphological traits. In addition, previous molecular studies were unable to resolve several relationships among key taxa within this family. We examined three different methods of phylogenetic estimation and the impact of locus filtering strategies based on the number of parsimony-informative sites considering all taxa and specific clades in the alignments. Reconstruction of deep-level relationships using the UCE data yielded trees where most nodes are in agreement and have high confidence, regardless of analytical method. Moreover, filtering "clade-specific" genes can produce distinct topologies and thus support for alternative results under the same estimation method should be interpreted with caution. Yet, in spite of some continuing uncertainties, the phylogenetic hypothesis advanced for the Pipridae can provide a firmer comparative context for future ecomorphological and behavioral studies.

Key words: ultraconserved elements, exon probes, informative loci, concatenation, multispecies coalescent, Tyrannides

1. Introduction

The field of phylogenetics has significantly advanced due to multilocus inferences using next-generation sequencing (NGS) technologies (McCormack and Faircloth, 2013). Massive-parallel sequencing can be employed via target capture protocols (Mamanova et al., 2010) to generate hundreds or thousands of unlinked loci (McCormack et al., 2013). This increase in the number of independent markers suitable for phylogenetic studies has revolutionized molecular systematics and the tree of life (e.g., Jarvis et al., 2014; Prum et al., 2015). A commonly used class of nuclear markers is ultraconserved elements, or UCEs (Faircloth et al., 2012). UCEs constitute highly conserved orthologous segments found across the genome of distinct vertebrates (Bejerano et al., 2004), characterized by flanking regions with more variable sites that can be used to investigate historical relationships at deep and shallow taxonomic levels (Faircloth et al., 2012). While most studies focus solely on UCEs, some studies have also targeted exons (and their flanking introns and untranslated regions) along with UCEs (Smith et al., 2014).

Many phylogenomic studies have used the concatenation approach, in which sequences of all genes are combined for each taxon and analyzed as single sequences in a supermatrix with all taxa. Analyses of concatenated data are computationally efficient (e.g., RAxML; Stamatakis, 2014) and provide intuitive measures of branch lengths that are typically expressed in substitutions per site. This has led to a considerable reliance on concatenated phylogenies when large-scale NGS data are analyzed. However, the concatenation approach has received criticism (see Edwards et al., 2016 for review) because it assumes all genes share the same underlying history (i.e., tree topology and branch lengths). Data collected using NGS methods have revealed substantial heterogeneity among different genes (Edwards, 2009), which may exhibit discordant gene trees (e.g., Jarvis et al., 2014; Salichos and Rokas, 2013). The incongruence among estimated gene trees can reflect a number of factors, including true conflicts among gene trees (e.g., deep coalescence and incomplete lineage sorting (ILS), horizontal gene transfer or introgression, and gene duplication and loss; Maddison, 1997) and errors in the estimation of gene trees (Gatesy and Springer, 2014; Patel et al., 2013). This has important implications for phylogenetic analyses because the most probable gene trees will differ from the species tree when the internal branches of a phylogeny are short relative to effective population size, in a part of parameter space called the anomaly zone (Degnan and Rosenberg, 2006). Under this circumstance, maximum likelihood (ML) bootstrap analyses of concatenated data can yield high support for an incorrect topology (Kubatko and Degnan, 2007).

Methods of phylogenetic inference that model ILS given the multispecies coalescent (MSC) may be consistent estimators of the species tree. Simulations have demonstrated

that Bayesian implementations of MSC methods can recover the true phylogeny under conditions of high ILS (Edwards et al., 2007; Heled and Drummond, 2010; Liu and Pearl, 2007). Nevertheless, those probabilistic models require intense computation because the species tree and gene trees need to be co-estimated, so they cannot be applied to phylogenomic data sets (Mirarab et al., 2016; Rannala and Yang, 2017). This has prompted the development of simpler algorithms that model ILS but make use of summary statistics in a two-step procedure to reduce the computational burden of estimating species trees from genomic-scale data (Liu et al., 2015). However, there is a trade off between model simplification and computation efficiency such that more loci are often needed to achieve reliable species tree estimates (Liu et al., 2009; Roch and Warnow, 2015).

The performance of MSC summary methods relative to concatenation has been the subject of intense debate (Gatesy and Springer, 2014; Meiklejohn et al., 2016; Patel et al., 2013; Song et al., 2012; Springer and Gatesy, 2016). Nevertheless, simulation and empirical studies find, by and large, similar relationships for the majority of nodes under both frameworks (Chen et al., 2015; Pyron et al., 2014; Tonini et al., 2015), although concatenated phylogenies generally show higher nodal support. In fact, a criticism of concatenation is that support values are inflated due to model violations (Liu et al., 2015; Roch and Steel, 2015). On the other hand, statistical properties of summary methods combining gene trees reconstructed beforehand for use in heuristic species tree estimation require that input gene trees are known without error (Mirarab et al., 2016; Roch and Warnow, 2015). However, the susceptibility of summary methods to gene tree estimation errors make their practice for estimating robust species tree questionable using empirical datasets (Molloy and Warnow, 2018; Patel et al., 2013; Roch and Warnow, 2015).

Inaccurate estimates of gene trees typically reflect cases in which the gene sequences are too short or the loci evolve too slowly to accumulate informative changes (Chou et al., 2015; Meiklejohn et al., 2016; Xi et al., 2015). Estimated gene trees also have higher probability of presenting polytomies when the species tree is in the anomaly zone simply given the random process of mutation (Huang and Knowles, 2009). Indeed, low phylogenetic signal from individual loci is a common feature in genomic-scale analyses of rapid radiations using biological data such as UCEs (Bayzid et al., 2015; Hosner et al., 2016; Meiklejohn et al., 2016; Molloy and Warnow, 2018), which is manifested in gene trees with low bootstrap support (Molloy and Warnow, 2018) and unresolved relationships (Huang et al., 2010). The inherent difficulties of estimating gene trees from empirical data sets (Roch and Warnow, 2015) has resulted in the development of MSC methods that circumvent gene tree estimation by extracting phylogenetic signal directly from site patterns in a sequence data matrix (Chifman and Kubatko, 2014; Chou et al., 2015). In addition, a variety of locus filtering strategies that remove loci with low

phylogenetic information were proposed with the potential to enhance the accuracy of species tree methods (Chen et al., 2015; Molloy and Warnow, 2018). Notwithstanding observed gains in overall data quality (e.g., reduced gene tree estimation error), the exclusion of several genes may also have an impact on estimating the species tree with good confidence (Mirarab et al., 2016; Molloy and Warnow, 2018). Yet, locus filtering can be useful for understanding sources of errors in species tree estimates and investigating difficult nodes more in depth with specific loci (Chen et al., 2015; Salichos and Rokas, 2013).

In order to address some of the challenges to empirical phylogenomic data sets, we examined the impact of different analytical approaches on phylogeny estimation of the avian family Pipridae (manakins) using UCE and exon probes obtained from sequence capture to generate thousands of loci. Specifically, we employed standard concatenation in RAxML (Stamatakis, 2014) and two MSC methods, the gene tree reconciliation method ASTRAL-II (Mirarab and Warnow, 2015) and the site-pattern method SVDQuartets (Chifman and Kubatko, 2014). Reconstruction of deep-level relationships within the Pipridae using UCE data yielded trees where most nodes are in agreement and have high confidence, regardless of analytical method. We devote special attention to recalcitrant relationships using locus filtering strategies to better understand the impact of distinct species tree methods on topology and nodal support. We show that filtering "clade-specific" genes can produce distinct topologies and that support for alternative results under the same estimation method should be interpreted with caution.

Manakins (Pipridae) are a group of small suboscine passerine birds characterized by a number of unique behaviors and morphological features (Kirwan and Green, 2012). This family is a well-supported clade including 53 named species that have been divided into 17 genera (Gill and Donsker, 2018). Piprids have their greatest diversity in Neotropical humid forests, but some taxa occur in dry woodlands and along riparian forests as well (Anciães and Peterson, 2009; Kirwan and Green, 2012). Early phylogenetic hypotheses were based on syringeal morphology, lek-display behavior and sexual plumage traits (Prum, 1990, 1992, 1994, 1997). Many species have strong sexually dimorphic plumage and elaborate lekking courtship rituals, which make this a very charismatic group of the tree of life. Previous molecular studies on manakins and their allies addressed the systematic relationships among representative genera (McKay et al., 2010; Tello et al., 2009) and defined new taxonomic ranks for three major clades of the Pipridae. More recently, Ohlson et al. (2013a) used additional taxon sampling to reexamine classification schemes for the family, ultimately recognizing a new genus. However, these former assessments used a limited number of markers and relationships among a number of key taxa remain unclear. Moreover, none of the previous phylogenetic studies employed MSC methods, making it impossible to evaluate the impact of these methods on estimates

of the manakin phylogeny. Our goal is to advance a robust phylogenetic hypothesis that can be used for future investigations focused on the evolution of traits, like the intriguing social behaviors of these fascinating neotropical birds.

2. Materials and Methods

2.1. Taxon sampling

We obtained samples for a total of 51 taxa within the family Pipridae (Tello et al., 2009), including almost all currently recognized species except for *Neopelma aurifrons* and the newly described species *Machaeropterus eckelberryi*. We also sampled three additional taxa (*Pyroderus scutatus*, *Onychorhynchus coronatus*, *Pachyramphus minor*) as representative genera of closely-related families (Cotingidae, Tyrannidae and Tityridae, respectively). Most samples came from freshly-preserved tissue or blood, but we also successfully sequenced two samples from museum specimens. Voucher numbers and institutions are listed in the Supplementary Material (Appendix A).

2.2. Library preparation, target enrichment and sequencing

We extracted genomic DNA from samples using Qiagen DNeasy Blood & Tissue Kits (Qiagen, Valencia, CA, USA). Sequence data were obtained by RAPiD Genomics (Gainesville, FL, USA) following methods detailed in Faircloth et al. (2012) with minor modifications. Briefly, the sequence-capture workflow involved preparation of Illumina TruSeq libraries using the manufacturer's protocols (Illumina Inc., San Diego, CA, USA) and primers with custom index tags for multiplexing. We enriched each library using a set of 4,715 custom probes (MYbaits, MYcroarray, Ann Arbor, MI, USA) targeting 49 exons plus 2,320 UCE loci with 100-nt paired-end reads sequenced on an Illumina HiSeq 2500.

2.3. Bioinformatic preprocessing

After massive parallel sequencing, we de-multiplexed the raw fastq data and removed adapter contamination and low-quality bases from reads using Illumiprocessor (Faircloth, 2013) as a parallel wrapper for Trimmomatic (Bolger et al., 2014). We processed the cleaned read data following standard bioinformatic pipelines implemented in Phyluce (Faircloth, 2016). We assembled the contigs using Trinity r2013-02-25 (Grabherr et al., 2011), then extracted exons and UCEs from those contigs matching enriched loci, and discarded as putative duplicates same contigs matching probes designed for multiple loci

or multiple contigs matching probes for same locus. We created a data matrix configuration including all taxa and the incomplete list of loci in each taxon to generate a data set that allowed missing data for some taxa per locus. We performed sequence alignments in parallel across all loci using MAFFT (Katoh and Standley, 2013) with the default edge-trimming settings of Phyluce.

2.4. Sequence data and locus filtering

We performed two sets of phylogenetic analyses to address the systematic relationships within Pipridae. Firstly, sequence matrices used in standard concatenated analyses consisted of UCE loci containing at least 75% and 95% of all taxa in each locus, and we also obtained alignments for all exon loci. The exon loci, which also included flanking non-coding sequences, were analyzed separately to provide an independent estimate of the manakin phylogeny using a different data type. Secondly, UCEs had more recovery than the exon loci and were used for comparisons among concatenation and multispecies coalescent methods. We generated matrices for all UCE loci that included the taxon *Pyroderus scutatus* to ensure that sequence alignments contained at least this outgroup taxon while also retaining other non-Pipridae taxa used in downstream analyses (see below). We applied inclusive and clade-specific filtering schemes to explore potential impacts of these different criteria on phylogenetic inferences using UCEs.

Loci with little phylogenetic information may compromise coalescent-based species tree algorithms relying on reconciliation of estimated gene trees (Meiklejohn et al., 2016; Xi et al., 2015). To avoid potential biases due to uninformative loci in empirical studies with UCEs, previous studies (Hosner et al., 2016; Meiklejohn et al., 2016) estimated a series of species trees using different gene subsets based on the number of parsimony-informative sites computed for each locus. They showed that distinct species tree methods were mostly congruent only when the most informative loci were analyzed. Thus, we assessed whether this recommendation was also valid in this example or whether it was unique to their system. We assembled data sets within four groups of UCEs based on the number of parsimony-informative sites in each locus, using: 1) all UCE loci, including those without any informative site; 2) only parsimony-informative loci, each having at least one parsimony-informative site; 3) the 50% most parsimony-informative loci; and 4) the 25% most parsimony-informative loci. In addition, we focused on *Chiroxiphia*/*Antilophia* as an empirical example of a problematic clade, showing moderately to weakly supported nodes, to further understand the utility of this simple metric of informativeness for phylogenomic inferences of recalcitrant relationships. We applied an additional taxon-oriented filtering to extract only *Chiroxiphia*/*Antilophia* taxa from the selected UCE matrices to new alignments, which

were used to calculate the number of parsimony-informative sites in each locus exclusively for these taxa. We then again generated full data sets with four different combinations of UCEs as described above, but loci were assembled following the clade-specific phylogenetic informativeness contained in *Chiroxiphia*/*Antilophia* taxa. We used Phyluce (Faircloth, 2016) to filter alignments and compute parsimony-informative sites.

2.5. Phylogenomic analyses

We concatenated UCE loci of 75% and 95% complete matrices and all exon loci into separate alignments, which were used as input to select the best partitioning scheme for each data set in PartitionFinder 2 (Lanfear et al., 2017). The exon loci recovered by sequence capture included coding regions targeted by the probes as well as flanking intronic and untranslated region (UTR) sequences. We defined separate data blocks within each exon locus based on the three codon positions for coding regions and on intron or UTR for the associated non-coding regions. UCE data blocks were defined by locus. We applied the relaxed hierarchical clustering algorithm (Lanfear et al., 2014) using default weights and percentage of schemes analyzed, with the maximum number of subsets set to 100, and estimated a maximum parsimony starting tree and unlinked branch lengths in RAxML (Stamatakis, 2014) using the general time reversible (GTR) model with gamma distribution of rate heterogeneity (+G). We used Bayesian information criterion (BIC) for model selection among three available options under these settings in PartitionFinder 2: GTR; GTR+G; or GTR+G+I, with a proportion of invariable sites.

We conducted ML concatenation inferences for unpartitioned and partitioned data sets comprised of the 75% and 95% complete UCE and all exon loci using RAxML under the GTR+G model, with *Pyroderus scutatus* as the outgroup and 20 initial random trees. We assessed nodal support via the *autoMRE* option to generate bootstrap replicates until convergence is reached and to draw bipartitions onto the best-scoring ML tree. Concatenation analyses using the filtered UCE data, where locus inclusion was based on informativeness, were conducted only for unpartitioned alignments using the same settings as above.

We also estimated ML gene trees and 100 bootstrapped gene tree replicates for each locus under these settings in RAxML. These estimated gene trees were used as input for the MSC gene tree reconciliation program ASTRAL-II (Mirarab and Warnow, 2015). We assessed branch support in two different ways. First, we conducted 100 bootstrap replicates resampling by locus and by site (Seo, 2008), and computed a greedy consensus tree from bootstrapped species trees. Second, we used the local posterior probabilities of branch support based on quartet frequencies (Sayyari and Mirarab, 2016).

We also evaluated a multispecies coalescent approach that rather takes input directly from the sequence data (SVDQuartets). SVDQuartets (Chifman and Kubatko, 2014, 2015) computes singular value decomposition scores to infer relationships among quartets of taxa and then estimates the species tree by assembling the collection of quartet splits. We partitioned the UCE data by locus and used *Pyroderus scutatus* as the outgroup. SVDQuartets analyses were implemented in PAUP* (Swofford, 2017) using 100,000 random quartets and we computed a 50% majority-rule consensus tree from 100 bootstrap replicates as measure of uncertainty.

3. Results

3.1. Sequence data

After we trimmed the raw data for adapter contamination and low-quality bases we obtained an average of 5,671,732 sequence reads with an average length of 97.4 base pairs (bp) (Appendix 2). The cleaned reads were assembled into an average of 11,438 contigs with an average length of 492 bp and an average sequencing coverage of 31× (Appendix 3). The UCE data sets of 75% and 95% completeness contained an average of 52 and 53 taxa (out of 54 taxa) per locus, respectively, and their partition schemes included 14 and 12 subsets. The number of taxa where sequence data were recovered using the exon probes was smaller than the number of taxa where sequence data were recovered using UCE probes, averaging 28 taxa per locus and ranging from 10 to 43 taxa. Table 1 shows descriptive summaries of the different data sets used in standard concatenation analyses. The “exon” data set comprised 73% of the 49 loci targeted by probes and contained the coding regions targeted by probes along with flanking non-coding regions; for simplicity we refer to these regions as “exon loci” since their recovery reflects the use of exon probes. We note that the PSMA2 locus was not assembled as a single contig, instead it was recovered as two non-contiguous segments. The aligned exon data set included 15,668 coding sites (45.6%), 16,256 intron sites (47.3%), and 2,227 UTR sites (6.5%). Partition schemes selected for the exon data set included five subsets. The UCE data used for comparison among concatenation and MSC methods of phylogenetic inference contained an average of 52 taxa per locus; see Table 2 for summary descriptions of the different UCE data sets filtered based on the number of parsimony-informative sites for alignments of inclusive and clade-specific taxa.

Table 1. Summary statistics for UCE and exon data sets used in standard concatenated analyses.

Data set	**Total number of loci**	**Average locus length**	**Total number of parsimony-informative sites**	**Average number of parsimony-informative sites**
UCE 75% complete loci	2,237	639	63,741	28
UCE 95% complete loci	1,796	653	52,642	29
Exon loci	36	955	1,875	52

Table 2. Summary statistics comparing UCE data sets under different inclusive and clade-specific filtering schemes. UCEs were filtered based on the number of parsimony-informative sites for all taxa and for clade-specific taxa. Numbers above and below lines within each scheme correspond to values calculated for the entire alignments (above) or alignments including only *Chiroxiphia*/*Antilophia* taxa (below), respectively.

Filtering scheme	Total number of loci	Average locus length	Total number of parsimony-informative sites	Average number of parsimony-informative sites	Average proportion of informative sites per difference
All loci	2,071	639.6	59,980	29.0	0.40
			3,966	1.9	0.27
All informative loci	2,062	640.1	59,980	29.1	0.40
			3,966	1.9	0.28
Chiroxiphia/*Antilophia* All informative loci	1,516	644.1	50,774	33.5	0.42
			3,966	2.6	0.38
50% most informative loci	1,055	655.2	45,652	43.3	0.45
			2,978	2.8	0.31
25% most informative loci	520	665.1	28,532	54.9	0.47
			1,888	3.6	0.33
Chiroxiphia/*Antilophia* 25% most informative loci	600	655.7	26,298	43.8	0.44
			2,691	4.5	0.45

3.2. Phylogenomics

3.2.1. Standard concatenation

The topologies from maximum likelihood concatenation of UCEs with 75% and 95% completeness using both unpartitioned and partitioned analyses were completely congruent among all four inferences. Most nodes on the phylogeny had 100% bootstrap support; however, a few relationships did not receive full support in the ML analyses and their bootstrap values varied according to the amount of taxon completeness and data partitioning (Fig. 1). These clades had slightly higher bootstrap support in the unpartitioned analysis (the difference in bootstrap values was ≤6%) than in the partitioned analysis of the more complete data set, with a maximum of 5% of missing taxa per locus. When up to 25% of missing taxa per locus was allowed in the data set, nodes without full support for *Lepidothrix, Pipra* and *Ceratopipra* had the same or slightly better bootstrap values (≤2% bootstrap difference) in the unpartitioned analysis, but more change (≥10% bootstrap difference) is observed for *Chiroxiphia*/*Antilophia*. Partitioned ML analyses produced overall higher nodal support values using the 75% complete data set with more missing data in the concatenated alignment, which amounts to ~20% additional loci despite a more incomplete taxon representation compared to the 95% complete matrices. In contrast, unpartitioned concatenated analyses had an average of 10% less bootstrap support for the conflicting relationships of *Chiroxiphia*/*Antilophia* as missing data increase in the 75% complete matrices.

The exon data analyzed using unpartitioned and partitioned maximum likelihood concatenation recovered some interrelationships with moderate to high nodal support (i.e., ≥70% bootstrap support), but several nodes had low bootstrap support, particularly among many of the genera (Fig. 2). This is likely due to the limited recovery of data from the exon probes (see above) and the more limited size of the exon data set. Moreover, unpartitioned versus partitioned inferences differed in the placement of *Chiroxiphia caudata* and *C. pareola* as well as of *Lepidothrix isidorei* and *L. coeruleocapilla*. The former pair was recovered as sister-taxa in the unpartitioned inference, whereas the two species of *Lepidothrix* were recovered as sisters in the partitioned inference. Nevertheless, for the few nodes that showed high support values (i.e., ≥95% bootstrap support) in both analyses of the exon data, the only conflicting relationship with the UCE results (including those topologies estimated under different filtering schemes, except when the 25% most informative loci were used with ASTRAL) was in the *Lepidothrix iris, L. nattereri, L. vilasboasi* clade (see Discussion).

3.2.2. Coalescent-based species trees and locus filtering

Estimates of the species tree obtained using MSC methods were largely congruent with concatenation results. Nodes with 100% bootstrap in the ML concatenated trees had strong support in the ASTRAL and SVDQuartets species trees. At the same time, those nodes with lower support in concatenated analyses varied in topology and/or were poorly supported in coalescent trees.

The ASTRAL best quartet-based species trees differed from the ASTRAL consensus trees in the placement of several taxa (Figs. 3a, S1 and S2), specifically when more loci were included in the data set, with a larger number of informative sites in total but lower information content per locus (Table 2). For instance, basal relationships among the clades of *Heterocercus*/*Manacus*, *Pipra* and *Machaeropterus* had high posterior probability from quartet frequencies but virtually no support from bootstrap resampling for the data set with at least one informative site per locus (i.e., only informative loci). This is due to the incongruent position of *Machaeropterus regulus* in the consensus versus best quartet-based tree. Analyses of the data set containing the 25% most informative loci ameliorated this conflict (Fig. 3a). Despite some topological differences among taxa within the genera *Neopelma*, *Chiroxiphia*, *Lepidothrix* and *Heterocercus*, the most basal relationships are highly supported in the ASTRAL best trees, regardless of the filtering scheme used. Posterior probabilities tend to decrease for the data sets with less loci and informative sites in total.

The species trees inferred using SVDQuartets were more similar to the topology of concatenated trees than to ASTRAL trees, and the basal nodes of the SVDQuartets trees likewise had overall strong support (Fig. 3b and S3). Yet, the phylogenetic relationships of *Neopelma*, *Chiroxiphia* and *Pipra* species contain areas of disagreement among the data sets with different filtering schemes.

3.2.3. Topology of the Chiroxiphia/Antilophia clade

Topologies within the *Chiroxiphia*/*Antilophia* clade were especially variable so we conducted a more detailed analysis of this part of the manakin tree. Concatenated analyses of the unpartitioned data sets did not display topological differences until the data were filtered to include the 25% most informative loci for that specific clade (Fig. 4). This clade-specific filtering enabled the identification of more informative loci for the *Chiroxiphia*/*Antilophia* clade itself when compared to the number of informative sites calculated for all taxa across the whole alignment (Table 2). ASTRAL consensus trees consistently recovered *C. caudata* in a basal split with the remainder of the taxa, followed by *C. boliviana* except when we used the 25% most informative loci for the clade-

specific data set (Fig. S1). On the other hand, the best trees from ASTRAL and the SVDQuartets trees recovered more variable relationships within the *Chiroxiphia*/*Antilophia* clade depending on the locus filtering scheme used.

The topologies were identical between concatenation and SVDQuartets methods when using the 25% most informative loci, but the inclusive and clade-specific filtering schemes recovered different relationships. In the inclusive filtering there is a basal split between *C. boliviana* and the remainder of the taxa, whereas in the clade-specific filtering *C. boliviana* is sister to *C. caudata* plus the two species of *Antilophia*. The ASTRAL best tree in this latter case instead recovered *C. caudata* as sister to *C. boliviana* plus the *Antilophia* species (Fig. 4), but that relationship had little support from posterior probabilities and virtually no bootstrap support; the ASTRAL consensus tree placed *C. caudata* in a basal split with the remainder of the taxa (Fig. S1). Overall, the topologies obtained from the 25% most informative loci filtered for the specific clade were more similar among the different methods than those trees recovered from data sets with less informative loci and inclusive filtering.

In general, a decrease in total nodal support (as measured by the average sum of bootstrap values or posterior probabilities for the clade) was observed for the relationships among *Chiroxiphia*/*Antilophia* taxa inferred from data sets with fewer loci, but more information per locus and per site difference (Table 1). For instance, compare the results of the concatenated and ASTRAL analyses using the 25% most informative loci to the results including all those loci with at least one informative site per locus (horizontal and diagonal arrows; Fig. 5). This tendency was also seen in the comparison of the inclusive with the clade-specific filtering schemes, although the magnitude of change in support difference was overall smaller (vertical arrows; Fig. 5). However, we detected some disparities in this general pattern in relation to the coalescent-based estimates of the ASTRAL best and SVDQuartets trees using the 25% most informative loci combined with the clade-specific filtering (red arrows; Fig. 5).

It is important to note that filtering the *Chiroxiphia*/*Antilophia* clade for an informativeness of at least 25% per locus resulted in 80 more loci than the inclusive filtering with the same amount of informativeness per locus (Table 1). Therefore, the greater support for the ASTRAL best tree under the clade-specific filtering scheme with the 25% most informative loci may in part reflect the more availability of information from estimated gene trees used as input in this analysis, and this may also be the case for the SVDQuartets results under these same criteria. Nevertheless, we noticed some major disparities in support difference for SVDQuartets between the clade-specific data set with the 25% most informative loci and the clade-specific or inclusive data sets containing any loci with informative sites (Fig. 5), despite the latter two having 2.5 and 3.4 times more

loci than the former, respectively. In addition, there were only minor differences in the other SVDQuartets comparisons among data sets having more information content versus those data sets with more loci. Finally, bootstrap proportions of the nodes containing recalcitrant relationships for the clade-specific data set with the 25% most informative loci were all above half, whereas in the other SVDQuartets analyses at least one node was equal to or less than 50% bootstrap support (Fig. 4).

3.2.4. Phylogenetic relationships

Despite some discordance within genera among different analyses, we found solid resolution for higher-level relationships in the family Pipridae. More specifically, the nominal subfamilies Neopelminae and Piprinae were recovered as clades A and B, respectively (see Fig. 1), with high support in all inferences using the exon and UCE loci. In addition, the UCE data consistently recovered sub-clades B1 and B2 within the Piprinae. Although the exon data did not recover these exact same sub-clades because *Xenopipo* was placed as a sister-group of sub-clade B1, such a relationship cannot be asserted with confidence given the limited resolution of analyses using the exon loci (Fig. 2). On the other hand, the position of *Xenopipo* was strongly supported on the UCE phylogeny, branching off at the base of sub-clade B2 (Figs. 1 and 3). Although generic support was less pronounced in analyses of the exon loci, the majority of the genera were monophyletic with well-resolved interrelationships in the UCE trees; the only exceptions were within *Neopelma*/*Tyranneutes* and *Chiroxiphia*/*Antilophia*, both pairs of which were paraphyletic with high nodal support using concatenation (Fig. 1) and MSC methods (Fig. 3).

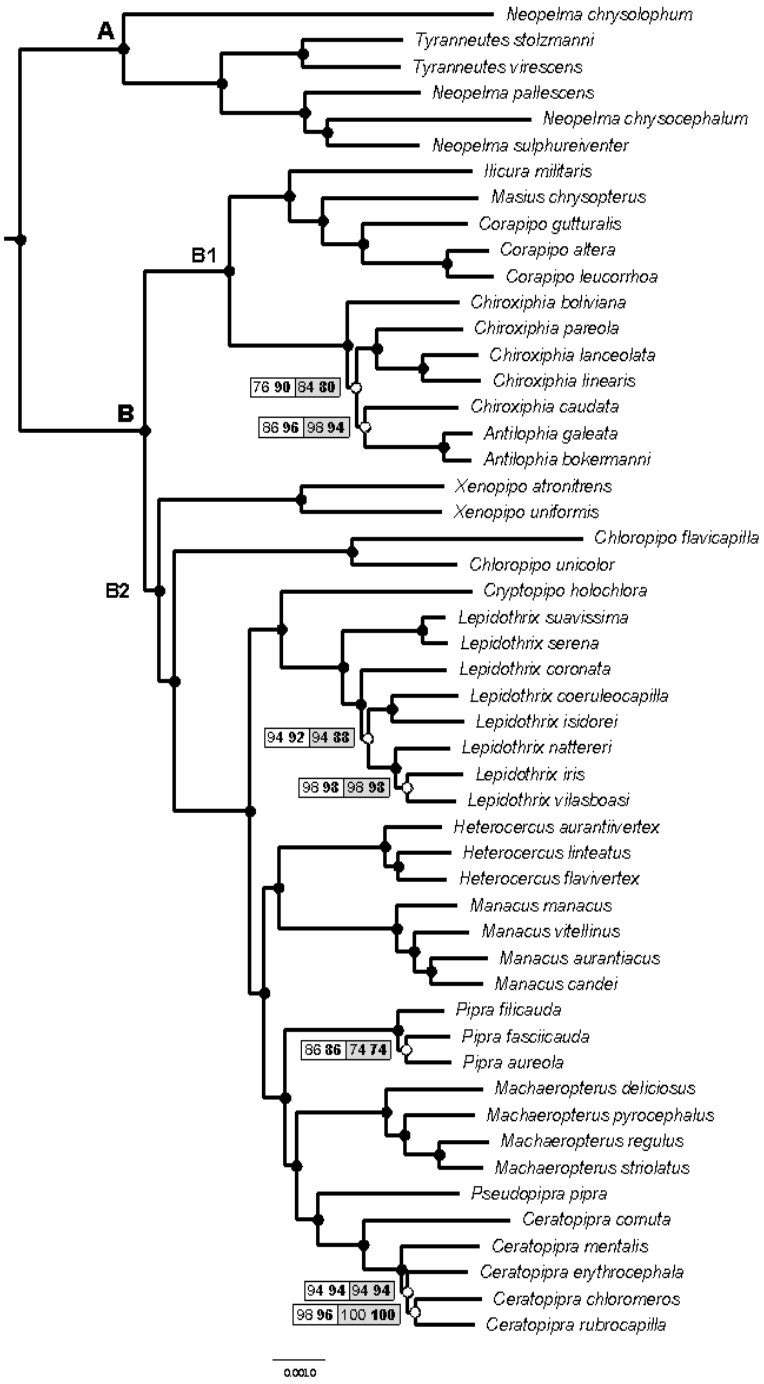


Figure 1. Maximum likelihood concatenated tree based on UCE loci. Numbers inside boxes correspond to nodal support values for unpartitioned and partitioned (in bold) inferences of the 75% (white) and 95% (gray) complete data sets; dark circles indicate 100% bootstrap support in all analyses. Subfamily ranks for Neopelminae (A) and Piprinae (B) follow the South American Classification Committee SACC591 (Remsen et al., 2018), and the proposed tribes Ilicurini (B1) and Piprini (B2) are based in a classification scheme modified from Ohlson et al. (2013a).

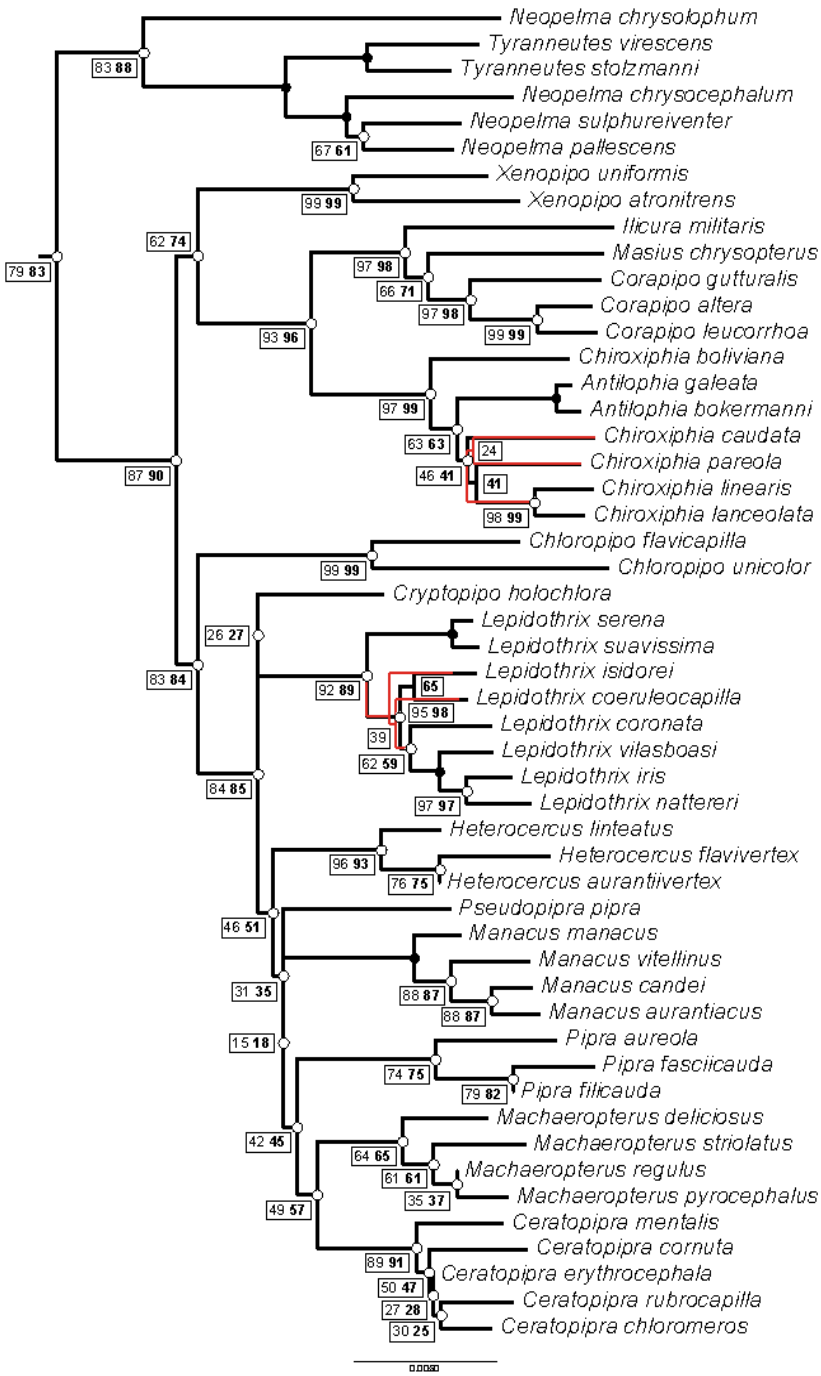


Figure 2. Maximum likelihood concatenated tree based on 36 exon loci. Numbers inside boxes correspond to nodal support values for unpartitioned and partitioned (in bold) inferences; dark circles indicate 100% bootstrap support in both analyses. Topological discordances between partition schemes are overlaid, with respective support values in separate boxes and the unpartitioned estimate depicted in red.

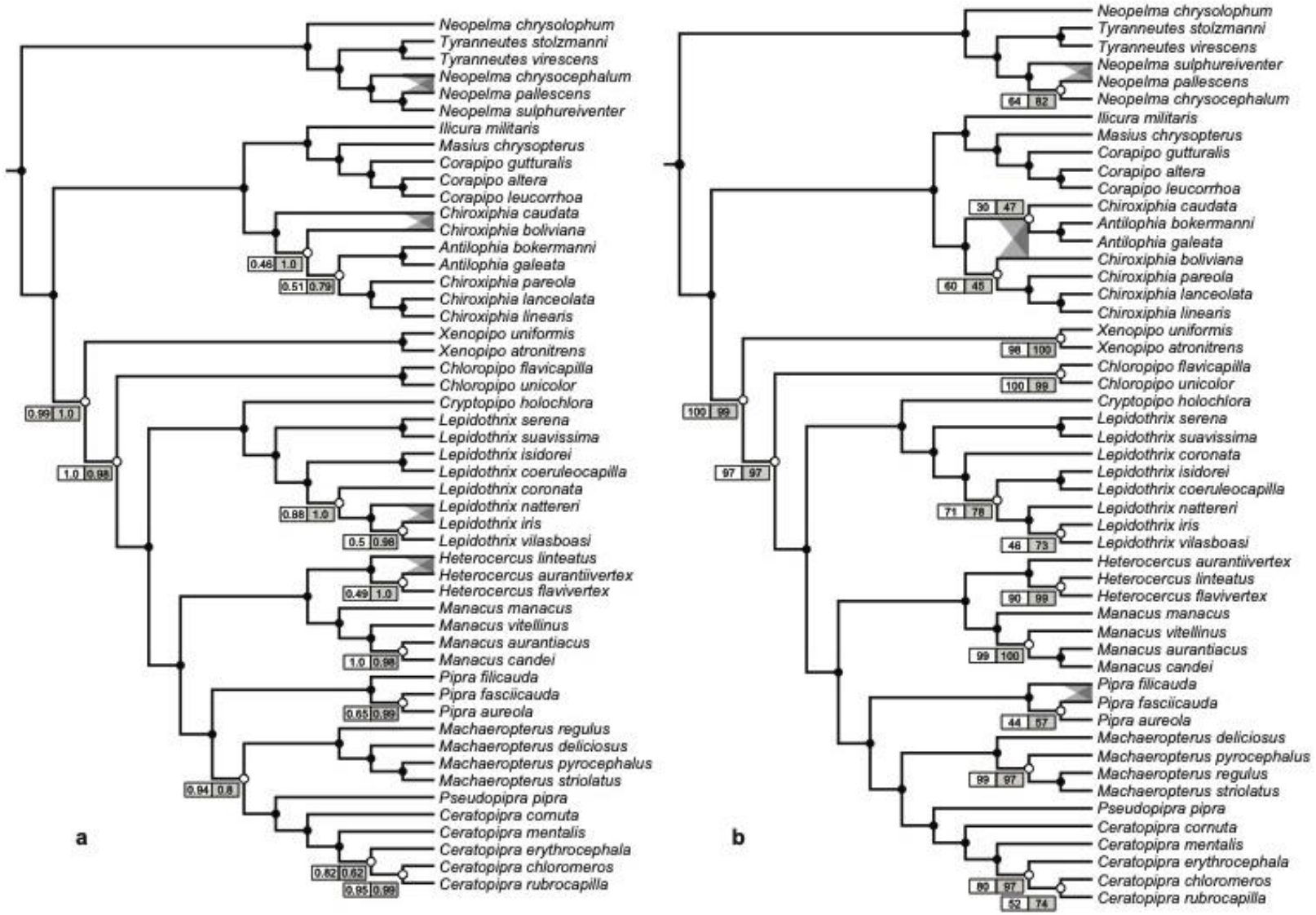

Neopelma chrysolophum
Tyranneutes stolzmanni
Tyranneutes virescens
Neopelma chrysocephalum
Neopelma pallescens
Neopelma sulphureiventer
Ilicura militaris
Masius chrysopterus
Corapipo gutturalis
Corapipo altera
Corapipo leucorrhoa
Chiroxiphia caudata
Chiroxiphia boliviana
Antilophia bokermanni
Antilophia galeata
Chiroxiphia pareola
Chiroxiphia lanceolata
Chiroxiphia linearis
Xenopipo uniformis
Xenopipo atronitrens
Chloropipo flavicapilla
Chloropipo unicolor
Cryptopipo holochlora
Lepidothrix serena
Lepidothrix suavissima
Lepidothrix isidorei
Lepidothrix coeruleocapilla
Lepidothrix coronata
Lepidothrix nattereri
Lepidothrix iris
Lepidothrix vilasboasi
Heterocercus linteatus
Heterocercus aurantiivertex
Heterocercus flavivertex
Manacus manacus
Manacus vitellinus
Manacus aurantiacus
Manacus candei
Pipra filicauda
Pipra fasciicauda
Pipra aureola
Machaeropterus regulus
Machaeropterus deliciosus
Machaeropterus pyrocephalus
Machaeropterus striolatus
Pseudopipra pipra
Ceratopipra cornuta
Ceratopipra mentalis
Ceratopipra erythrocephala
Ceratopipra chloromeros
Ceratopipra rubrocapilla
a
b

Figure 3. Species trees based on (a) ASTRAL best tree and (b) SVDQuartets tree estimates. Numbers inside boxes are nodal support values of the posterior probabilities from quartet frequencies (ASTRAL) and bootstrap replicates (SVDQuartets) inferred using UCE loci with at least one informative site (gray) and the 25% most informative UCE loci (white); dark circles indicate full support in both analyses. The depicted topologies were obtained using all alignments with at least one informative site per locus (i.e., all informative loci); gray zones represent taxa with conflicting relationships between the data set of 25% most informative UCE loci.

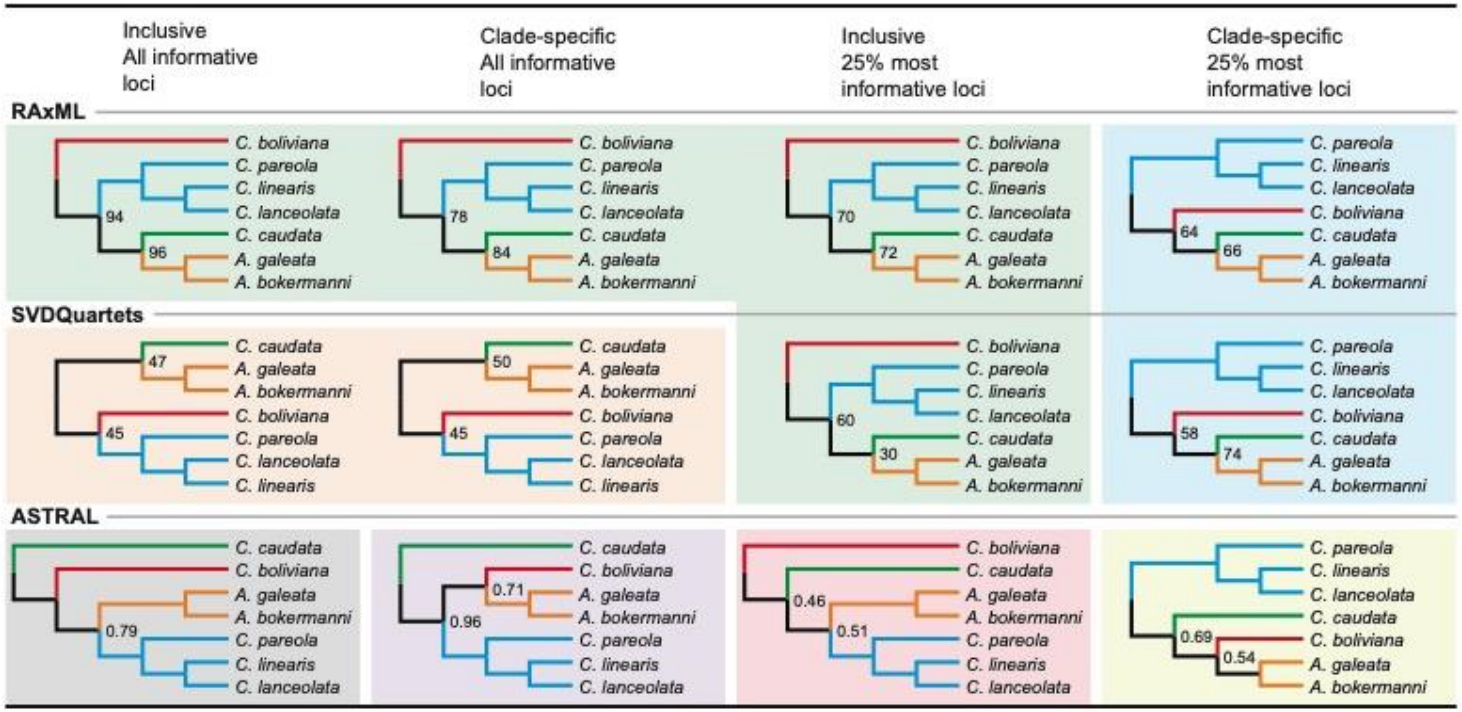


Figure 4. UCE data sets filtered by the number of parsimony-informative sites per locus. Nodal values represent bootstrap support (RAxML and SVDQuartets) and posterior probabilities from quartet frequencies (ASTRAL); fully supported nodes are omitted. Colored shades indicate alternative topologies. In some cases, subtrees differ only in the placement of the root (see section 4.2 in the Discussion).

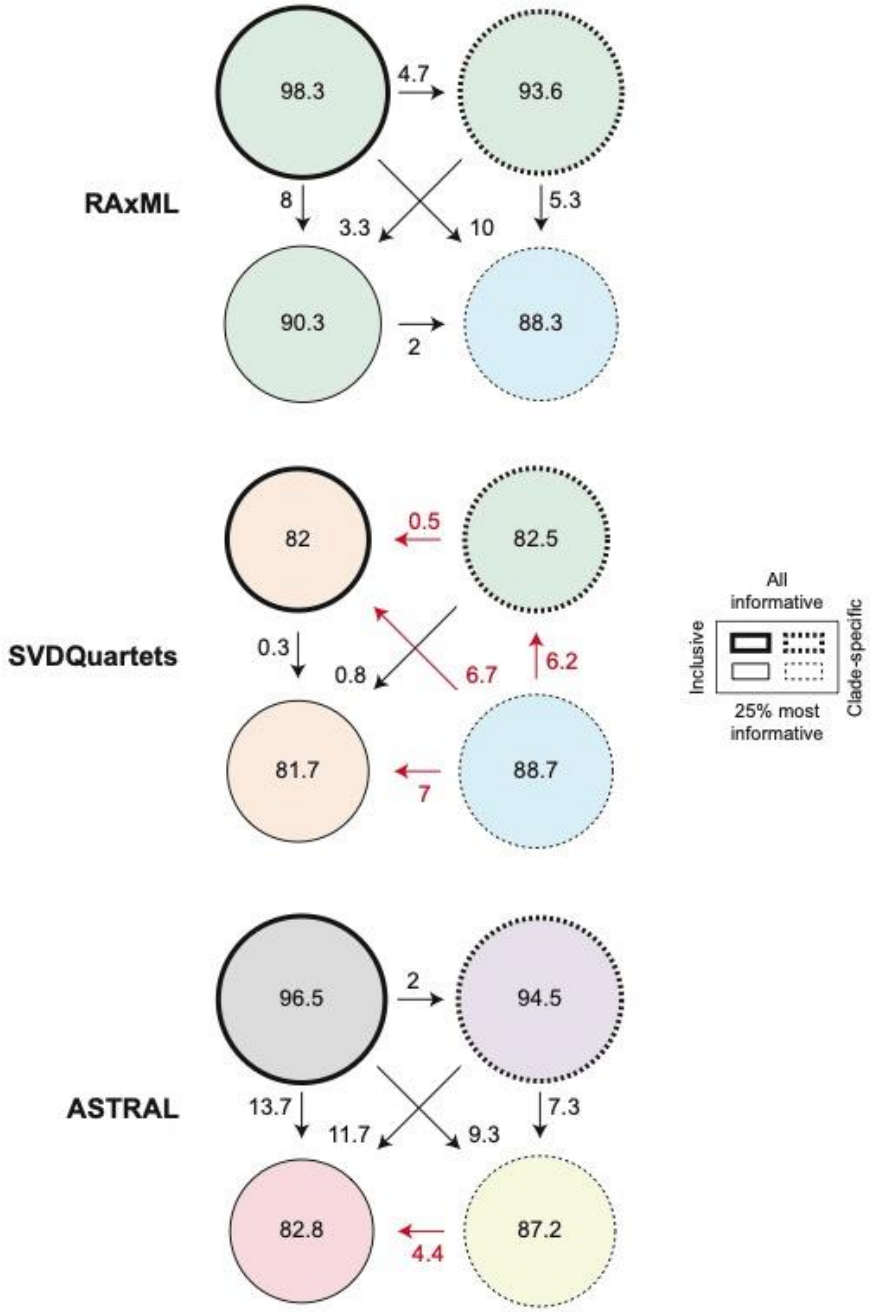


Figure 5. Schematic representation of changes in total nodal support (as measured by the sum of bootstrap values or posterior probabilities) for the clade containing *Chiroxiphia*/*Antilophia* taxa. In this topology with seven species (also the case of a four-taxon tree), one fully supported clade would correspond to 600. The four different locus filtering schemes analyzed for this clade are represented by the circles for the inclusive (top continuous) and clade-specific (bottom dotted) data sets using those UCEs with at least one informative site per locus (left bold) and the 25% most informative UCE loci (right).

4. Discussion

4.1. Performance of sequence capture in manakins

Overall, we obtained excellent recovery of UCE loci and those loci were sufficient to produce an estimate of manakin phylogeny with high support for most nodes regardless of the analytical method used. Although most avian phylogenies using UCEs have focused on probe sets that exclusively contain UCEs (e.g., Moyle et al., 2016; Sun et al., 2014), we used a subset of the "Tetrapods-UCE-5Kv1" probe set (Sun et al., 2014) and added exon probes designed to amplify commonly-used introns and coding regions in avian systematics (e.g., Kimball et al., 2009; Reddy et al., 2017). The exon probes had the potential advantage of obtaining loci that may have also been sampled in previous studies, potentially facilitating the inclusion of the manakin data in large-scale supermatrix phylogeny efforts (e.g., Burleigh et al., 2015). However, the phylogeny using the exon data was poorly resolved, due in part to the poor recovery of these regions, with only 73% of loci being recovered. There were also more missing data within the exon loci, since the best probe appeared in only 84% of taxa and many loci were recovered in fewer than half of the taxa sampled. Thus, the UCEs seemed to be a much more efficient use of probes, and although individually less variable than the loci recovered using the exon probes, collectively provided much more data.

There were essentially no conflicts between the exon and UCE trees that exhibited a high degree (≥95%) of bootstrap support. The one exception was the clade that comprised *Lepidothrix iris, L. nattereri* and *L. vilasboasi.* Barrera-Guzmán et al. (2018) showed that *L. vilasboasi* is likely a hybrid species derived from the *L. iris* and *L. nattereri* lineages, and that a larger proportion (~2/3) of the hybrid species genome is more closely related to the *L. iris* putative parental lineage. Analyses of the UCEs recovered a sister relationship between *L. vilasboasi* and *L. iris* using either concatenation or coalescent-based approaches, with the only exception of ASTRAL for the most reduced data set (Figs. 2, 3 and S1). The exon tree presented a different relationship, instead finding support for a sister relationship between the two putative parental species (*L. iris* and *L. nattereri*). These results are consistent with the hypothesis of a hybrid origin in that phylogenetic methods that model species relationships on a bifurcating tree (rather than modeling hybridization explicitly in a species network) are expected to support the topology congruent with the origin of the majority of the genome when a large number of loci is analyzed. In contrast, sampling error might prevail when only a limited number of loci is used, thus rendering the topology unstable.

This conflict is likely to reflect the limited number of loci necessary for a reliable species tree from estimated gene trees as required by ASTRAL, and for the exonic data may also

reflect the limited number of taxa since there were only 22 exon loci that were sampled for all three species. Moreover, short sequence alignments are known to impact gene tree reconciliation methods due to higher gene tree estimation error (Roch and Warnow, 2015). We did not assess the utility of the exon data set with coalescent-based methods because of the low recovery of loci and poor resolution of the manakin phylogeny; however, the average locus length for UCEs did not vary significantly among the different data sets (Table 1). Thus, the relative performance of summary methods such as ASTRAL for a typical UCE data set will depend not only on accurate gene trees (Mirarab et al., 2016) but also on sufficient genes, and although exon probes may have desirable properties, collectively these results emphasize the need to sample large numbers of loci when there are difficult nodes in a phylogeny.

4.2. Locus filtering in phylogenomics

There has been substantial debate regarding the value of excluding subsets of the genome in phylogenomic analyses. Molloy and Warnow (2018) highlighted recent empirical papers that examined the impact of locus filtering based on various proxies for gene tree estimation error and noted that the recommendations based on those empirical studies were at least somewhat contradictory. These contradictions likely result from a combination of the different parameters used in each empirical study, the different ways that gene tree estimation error is accounted for, and specific features of the empirical data sets. Using simulations, Molloy and Warnow (2018) found that the benefits of locus filtering were limited to non-existent. More specifically, they showed that removing loci based on proxies for gene-tree estimation error did not improve concatenation using RAxML or SVDQuartets, but it did improve gene tree reconciliation methods (e.g., ASTRAL) when levels of ILS were low to moderate. On the other hand, the accuracy of species trees changed considerably with higher levels of ILS, and RAxML and SVDQuartets outperformed gene tree reconciliation methods when gene tree estimation error was high.

Based on those simulations, filtering loci with greater distances between the estimated and true gene trees could improve species tree estimation for some methods depending on the ILS conditions. Recommendations rely on the assumption that errors in analyses of the simulated data are similar to those in empirical studies; however, if the estimated gene trees for empirical data are, on average, further than expected from the true gene trees based on simulations, then filtering is likely to be beneficial. The interplay between ILS and gene tree estimation error will determine the performance of the different species tree methods, thus the central question ultimately is where do the data lie in parameter space. Concatenation and SVDQuartets share the feature of estimating the species tree directly from site patterns in the aligned sequences, therefore both consider the

mutational process as a source of variability, with a potential advantage of SVDQuartets in incorporating additional variance of the coalescent process (Chifman and Kubatko, 2014). In contrast, summary methods such as ASTRAL model the MSC but rely on estimated gene trees, thus much of the information associated with the variation in individual loci (e.g., branch lengths and substitution model) is ignored. In cases where individual loci typically have little phylogenetic information, like UCEs, it may be difficult to estimate the topology, branch lengths and substitution model for each locus, so gene tree estimation error is expected to be relatively high (e.g., Meiklejohn et al., 2016), which can aggravate the difficulties of estimating the species tree from reconciled gene trees.

In spite of these differences among methods, many relationships in the manakin phylogeny were highly robust. The differences only become obvious if we focus on difficult nodes that did not receive full support (i.e., 100% bootstrap support or a posterior probability of 1.0). RAxML and SVDQuartets yielded a single topology for the relationships within *Lepidothrix,* whereas ASTRAL yielded two other topologies that also differed from each other depending on the number of informative loci included. As described above, *L. vilasboasi* appears to be a hybrid species (Barrera-Guzmán et al., 2018). Consequently, this violates the models used for data analyses, since hybrid speciation violates both the "single-tree model" implicit in standard concatenation and the MSC model that underlies ASTRAL and SVDQuartets. Hybrid speciation has seldom been documented in birds, although it remains possible that the apparent paucity of hybrid bird species is simply a function of the limited data used to examine species histories. The increased sampling of loci from phylogenomic data may reveal additional cases of hybrid speciation and this may power the reevaluation of many nodes in the bird tree.

The situation for the *Chiroxiphia*/*Antilophia* clade, where we compared three methods and four treatments of the data, was slightly more complex. Given that two subclades within the *Chiroxiphia*/*Antilophia* clade are strongly supported, this larger clade can be viewed of as a rooted four-taxon tree; the four lineages correspond to the two *Antilophia* species, *C. boliviana*, *C. caudata,* and the *C. pareola, C. linearis*, and *C. lanceolata* subclade. Thus, there are only 15 candidate topologies. Three distinct topologies emerged from concatenation and SVDQuartets analyses, whereas the topologies from ASTRAL analyses exhibited more variation (Fig. 4). The trees generated by RAxML and SVDQuartets differed only in the placement of the root (and the alternative placement of the root in SVDQuartets was poorly supported; Fig. 4). In contrast, all four treatments of the data analyzed with ASTRAL resulted in a tree with distinct topology (Fig. 4). Even more significantly, two distinct unrooted topologies emerged from ASTRAL analyses and both of those topologies differed from the unrooted topologies recovered in

concatenation and SVDQuartets analyses. These results emphasize the uncertainty associated with analyses of the *Chiroxiphia*/*Antilophia* clade since there are only three plausible unrooted trees. The high degree of instability associated with ASTRAL analyses might be expected if both ILS and gene tree estimation error are high within this clade.

Like some prior studies (Hosner et al., 2016; Meiklejohn et al., 2016), we focused on the number of parsimony informative sites as a proxy for locus informativeness. However, the best metric for locus informativeness is far from clear. Chen et al. (2015) suggested that different loci might perform better at different depths in the tree and suggested that one should focus on "question-specific" genes, which they defined as genes that resolved a specific node. Briefly, Chen et al. (2015) identified specific challenging nodes and proposed three possible resolutions of those nodes; any genes that did not yield one of those resolutions were rejected from their "question-specific" set. The Chen et al. (2015) approach is useful if there really are only three plausible resolutions of a challenging node (i.e., the node is a soft polytomy involving three taxa). However, there are many cases in the tree of life where many more than three resolutions are possible (e.g., the 15 plausible resolutions for the *Chiroxiphia*/*Antilophia* clade). Thus, although the Chen et al. (2015) method is an interesting approach it is unclear how it could be extended to more challenging problems.

Herein, we proposed an alternative way to filter loci: we focused on those genes with large numbers of informative sites in a problematic clade (the *Chiroxiphia*/*Antilophia* clade in this study). We called these "clade-specific" loci. An ideal locus filtering approach would result in distinct species tree methods to converge on the same estimate of phylogeny. Locus filtering may have improved the overall accuracy of estimated species trees by reducing the proportion of inconsistent genes and the amount gene tree estimation errors. Different MSC methods like ASTRAL and SVDQuartets would be expected to converge on the same phylogeny if ILS was prominent. In contrast, if the true species tree is in the part of parameter space where concatenation is inconsistent, one would expect concatenated trees to differ from the species tree. That was not what we observed in this study. Concatenation and SVDQuartets resulted in the same tree for the *Chiroxiphia*/*Antilophia* clade when clade-specific genes were used (Fig. 4), but that tree differed from the ASTRAL tree for the same loci, possibly because gene tree errors have a more significant impact on species tree estimates. It seems likely that the topology for the *Chiroxiphia*/*Antilophia* clade simply requires more data to resolve and no locus filtering or improved analyses will yield a satisfying result.

All of these results emphasize the caution with which systematists should approach analyses of NGS sequence data when challenging nodes are examined. This is not

surprising considering that whole-genome analyses were unable to resolve recalcitrant nodes at the base of Neoaves (Jarvis et al., 2014), and may depend upon specific types of loci that are analyzed (Jarvis et al., 2014; Reddy et al., 2017). Despite these challenges, phylogenomic analyses often yield trees in which most nodes are both well supported and insensitive to analytical methodology (e.g., Hosner et al., 2016; Moyle et al., 2016). That was certainly true in this study; the backbone for manakin relationships was strongly supported and estimates of phylogeny based on concatenation, ASTRAL, and SVDQuartets were congruent with just the few exceptions noted above.

4.3. Systematic considerations

The first phylogenetic hypothesis for Pipridae was based on the analysis of syringeal morphology and behavioral characters (Prum, 1990). Subsequently, Prum (1992) used morphology of the syrinxes to propose taxonomic changes and recognition of four tribes within the family, although he did not consider the "tyrant" manakins *Neopelma* and *Tyranneutes* to be members of Pipridae because they shared characters with other tyrannoids that were interpreted as synapomorphies. Subsequent molecular phylogenetic studies using mitochondrial and/or nuclear markers clarified several aspects of the higher-level relationships among the Tyrannides (Barber et al., 2007; Chesser, 2004; McKay et al., 2010; Ohlson et al., 2008; Ohlson et al., 2013a; Ohlson et al., 2013b; Tello et al., 2009), in which the Pipridae comprised a well-supported monophyletic group that included the two genera of tyrant manakins. In agreement with those molecular phylogenies, we found that the sexually monomorphic genera *Neopelma* and *Tyranneutes* form a clade that is sister to all other manakin genera with typical plumage dimorphism (the "core" manakins).

Previous molecular studies that assessed the validity of the four tribes erected by Prum (1992) found support for only one of them (McKay et al., 2010; Tello et al., 2009). These studies were only able to corroborate Ilicurini, suggesting the non-monophyly of Prum's Manacini and Piprini; his fourth tribe, Machaeropterini, comprises a single genus (*Machaeropterus*). Molecular phylogenetic studies also found support for two subgroups within the clade formed by the "core" manakins. Tello et al. (2009) used two protein-coding genes (RAG1 and RAG2) and ranked those three clades as the subfamilies Neopelminae, Piprinae and Ilicurinae. However, the relationship of *Xenopipo atronitens* and *X. uniformis* as sister to the Ilicurinae could not be strongly confirmed, partially because of incomplete taxon sampling. Ohlson et al. (2013b) analyzed a somewhat similar taxon set using those two RAG genes and three additional intron loci, and found that *Xenopipo atronitens* and *Chloropipo unicolor* were successive sister-taxa to a clade nested within the "core" manakins, albeit those relationships were not firmly resolved.

Our UCE results are consistent with that finding and have strong nodal support (Figs. 1 and 3).

Previously, the most complete taxon sampling of the manakin phylogeny was that of Ohlson et al. (2013a), who specifically addressed the monophyly of *Pipra* and *Chloropipo*. They proposed a rearrangement of the Piprinae into two tribes, Ilicurini and Piprini, with *Xenopipo* being sister to the remaining Piprini. *Chloropipo* was found to be non-monophyletic, with *C. unicolor* plus *C. flavicapilla* being sister to their Ilicurini with very poor support, whereas *C. holochlora* was placed in a new monotypic genus (*Cryptopipo*) that was sister to *Lepidothrix*. Our UCE results, on the other hand, found *C. unicolor* plus *C. flavicapilla* to be sister to all other Piprini except *Xenopipo* with strong support (Figs. 1 and 3). In addition, we found substantial differences from Ohlson et al. 2013a) in species relationships within various genera, including *Manacus*, *Machaeropterus*, and *Ceratopipra* (= *Pipra*).

This paper provides the most well-supported tree to date for the Pipridae, with important taxonomic implications that can be further addressed in future comparative studies (e.g., including morphological and behavioral data). We found that the enigmatic genera *Xenopipo* and *Chloropipo* were sequential sisters to the remainder of the taxa within the Piprini, thus representing a new rearrangement of this tribe (sensu Ohlson et al., 2013a). Although Rêgo et al.'s (2007) mitochondrial study confirmed the validity of *Pseudopipra* (currently *Dixiphia*) as well as two separate clades corresponding to the former genus *Pipra* (currently *Pipra* and *Ceratopipra*), their relationships among other genera remained unclear. Moreover, subsequent molecular studies that added a few more loci still could not resolve deeper interrelationships within the "core" manakins (Ohlson et al., 2013a; Ohlson et al., 2013b). Nevertheless, we were able to resolve the intergeneric relationships among *Pipra*, *Ceratopipra*, *Pseudopipra* and *Machaeropterus* and among *Ilicura*, *Masius* and *Corapipo*, as well as clarify the sister relationship between the genera *Heterocercus* and *Manacus*. Our results also substantially contribute to clarifying multiple species interrelationships.

It also became evident from our results that the available taxonomy awaits revision of two paraphyletic genera. In the case of *Neopelma*/*Tyranneutes*, concatenation and coalescent-based inferences produced the same topology when the 25% most informative loci were used. A recent paper using only five (three mitochondrial and two nuclear) genes recovered the same topology for these taxa that comprise an old history of divergence (Capurucho et al., 2018). However, for more recent scenarios of diversification as in the case of *Chiroxiphia*/*Antilophia* (Silva et al., 2018), more loci with adequate information content will be required to resolve their history. Yet, in spite of

some continuing uncertainties, the phylogenetic hypothesis advanced herein can provide a firmer comparative context for future ecomorphological and behavioral studies.

5. Acknowledgments

This work was funded by research grants from NSF (DEB 1146248 and 1241066) to J.C. and FAPESP (2012/50260-6) to Lucia G. Lohmann; R.N.L. was supported by FAPESP (2016/08356-7) and CNPq (457267/2014-3; 150128/2018-5). R.T.K. and E.L.B received support from NSF (DEB 1655683). This work was also facilitated by the Manakin Genomics RCN with support from NSF (DEB 1457541) to Bette Loiselle, Emily DuVal, Christopher Balakrishnan, Michael Braun and W. Alice Boyle. We are thankful to scientific collections and museums (American Museum of Natural History, Field Museum of Natural History, Florida Museum of Natural History, Instituto Nacional de Pesquisas da Amazônia, Louisiana State Museum of Natural Science, Museu Paraense Emílio Goeldi) and to the collectors and curatorial personnel who provided samples for this study. R.N.L. thanks Dr. Fernanda Werneck for providing logistic support.

6. References

Anciães, M., Peterson, A.T., 2009. Ecological niches and their evolution among Neotropical manakins (Aves: Pipridae). Journal of Avian Biology 40, 591-604.

Barber, B.R., Rice, N.H., Klicka, J., 2007. Systematics and evolution in the Tityrinae (Passeriformes: Tyrannoidea). The Auk 124, 1317-1329.

Barrera-Guzmán, A.O., Aleixo, A., Shawkey, M.D., Weir, J.T., 2018. Hybrid speciation leads to novel male secondary sexual ornamentation of an Amazonian bird. Proceedings of the National Academy of Sciences 115, E218.

Bayzid, M.S., Mirarab, S., Boussau, B., Warnow, T., 2015. Weighted Statistical Binning: Enabling Statistically Consistent Genome-Scale Phylogenetic Analyses. PLOS ONE 10, e0129183.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved Elements in the Human Genome. Science 304, 1321.

Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114-2120.

Burleigh, J.G., Kimball, R.T., Braun, E.L., 2015. Building the avian tree of life using a large-scale, sparse supermatrix. Mol Phylogenet Evol 84, 53-63.

Capurucho, J.M.G., Ashley, M.V., Ribas, C.C., Bates, J.M., 2018. Connecting Amazonian, Cerrado, and Atlantic forest histories: Paraphyly, old divergences, and modern population dynamics in tyrant-manakins (Neopelma/Tyranneutes, Aves: Pipridae). Mol Phylogenet Evol 127, 696-705.

Chen, M.-Y., Liang, D., Zhang, P., 2015. Selecting Question-Specific Genes to Reduce Incongruence in Phylogenomics: A Case Study of Jawed Vertebrate Backbone Phylogeny. Syst Biol 64, 1104-1120.

Chesser, R.T., 2004. Molecular systematics of New World suboscine birds. Mol Phylogenet Evol 32, 11-24.

Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics 30, 3317-3324.

Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. Journal of Theoretical Biology 374, 35-47.

Chou, J., Gupta, A., Yaduvanshi, S., Davidson, R., Nute, M., Mirarab, S., Warnow, T., 2015. A comparative study of SVDquartets and other coalescent-based species tree estimation methods. BMC Genomics 16, S2.

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of species trees with their most likely gene trees. Plos Genetics 2, 762-768.

Edwards, S.V., 2009. Is a new and general theory of molecular systematics emerging? Evolution 63, 1-19.

Edwards, S.V., Liu, L., Pearl, D.K., 2007. High-resolution species trees without concatenation. Proceedings of the National Academy of Sciences 104, 5936-5941.

Edwards, S.V., Xi, Z., Janke, A., Faircloth, B.C., McCormack, J.E., Glenn, T.C., Zhong, B., Wu, S., Lemmon, E.M., Lemmon, A.R., Leaché, A.D., Liu, L., Davis, C.C., 2016. Implementing and testing the multispecies coalescent model: A valuable paradigm for phylogenomics. Mol Phylogenet Evol 94, Part A, 447-462.

Faircloth, B.C., 2013. Illumiprocessor: a Trimmomatic wrapper for parallel adapter and quality trimming.

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786-788.

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.

Gatesy, J., Springer, M.S., 2014. Phylogenetic analysis at deep timescales: Unreliable gene trees, bypassed hidden support, and the coalescence/concatalescence conundrum. Mol Phylogenet Evol 80, 231-266.

Gill, F., Donsker, D., 2018. IOC World Bird List (version 8.2).

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

Heled, J., Drummond, A.J., 2010. Bayesian Inference of Species Trees from Multilocus Data. Mol Biol Evol 27, 570-580.

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.

Huang, H., He, Q., Kubatko, L.S., Knowles, L.L., 2010. Sources of Error Inherent in Species-Tree Estimation: Impact of Mutational and Coalescent Effects on Accuracy and Implications for Choosing among Different Methods. Syst Biol 59, 573-583.

Huang, H., Knowles, L.L., 2009. What Is the Danger of the Anomaly Zone for Empirical Phylogenetics? Syst Biol 58, 527-536.

Jarvis, E.D., Mirarab, S., Aberer, A.J., Li, B., Houde, P., Li, C., Ho, S.Y.W., Faircloth, B.C., Nabholz, B., Howard, J.T., Suh, A., Weber, C.C., da Fonseca, R.R., Li, J., Zhang, F., Li, H., Zhou, L., Narula, N., Liu, L., Ganapathy, G., Boussau, B., Bayzid, M.S., Zavidovych, V., Subramanian, S., Gabaldón, T., Capella-Gutiérrez, S., Huerta-Cepas, J., Rekepalli, B., Munch, K., Schierup, M., Lindow, B., Warren, W.C., Ray, D., Green, R.E., Bruford, M.W., Zhan, X., Dixon, A., Li, S., Li, N., Huang, Y., Derryberry, E.P., Bertelsen, M.F., Sheldon, F.H., Brumfield, R.T., Mello, C.V., Lovell, P.V., Wirthlin, M., Schneider, M.P.C., Prosdocimi, F., Samaniego, J.A., Velazquez, A.M.V., Alfaro-Núñez, A., Campos, P.F., Petersen, B., Sicheritz-Ponten, T., Pas, A., Bailey, T., Scofield, P., Bunce, M., Lambert, D.M., Zhou, Q., Perelman, P., Driskell, A.C., Shapiro, B., Xiong, Z., Zeng, Y., Liu, S., Li, Z., Liu, B., Wu, K., Xiao, J., Yinqi, X., Zheng, Q., Zhang, Y., Yang, H., Wang, J., Smeds, L., Rheindt, F.E., Braun, M., Fjeldsa, J., Orlando, L., Barker, F.K., Jønsson, K.A., Johnson, W., Koepfli, K.-P., O'Brien, S., Haussler, D., Ryder, O.A., Rahbek, C., Willerslev, E., Graves, G.R., Glenn, T.C., McCormack, J., Burt, D., Ellegren, H., Alström, P., Edwards, S.V., Stamatakis, A., Mindell, D.P., Cracraft, J., Braun, E.L., Warnow, T., Jun, W., Gilbert, M.T.P., Zhang, G., 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science 346, 1320-1331.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Mol Biol Evol 30, 772-780.

Kimball, R.T., Braun, E.L., Barker, F.K., Bowie, R.C.K., Braun, M.J., Chojnowski, J.L., Hackett, S.J., Han, K.-L., Harshman, J., Heimer-Torres, V., Holznagel, W., Huddleston, C.J., Marks, B.D., Miglia, K.J., Moore, W.S., Reddy, S., Sheldon, F.H., Smith, J.V., Witt, C.C., Yuri, T., 2009. A well-tested set of primers to amplify regions spread across the avian genome. Mol Phylogenet Evol 50, 654-660.

Kirwan, G.M., Green, G., 2012. Cotingas and Manakins. Princeton University Press, Princeton, New Jersey.

Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. Syst Biol 56, 17-24.

Lanfear, R., Calcott, B., Kainer, D., Mayer, C., Stamatakis, A., 2014. Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol 14, 82.
Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., Calcott, B., 2017. PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for Molecular and Morphological Phylogenetic Analyses. Mol Biol Evol 34, 772-773.
Liu, L., Pearl, D.K., 2007. Species trees from gene trees: reconstructing Bayesian posterior distributions of a species phylogeny using estimated gene tree distributions. Syst Biol 56, 504-514.
Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data. Annals of the New York Academy of Sciences 1360, 36-53.
Liu, L., Yu, L., Pearl, D.K., Edwards, S.V., 2009. Estimating Species Phylogenies Using Coalescence Times among Sequences. Syst Biol 58, 468-477.
Maddison, W.P., 1997. Gene trees in species trees. Syst Biol 46, 523-536.
Mamanova, L., Coffey, A.J., Scott, C.E., Kozarewa, I., Turner, E.H., Kumar, A., Howard, E., Shendure, J., Turner, D.J., 2010. Target-enrichment strategies for next-generation sequencing. Nature Methods 7, 111.
McCormack, J.E., Faircloth, B.C., 2013. Next-generation phylogenetics takes root. Mol Ecol 22, 19-21.
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.
McKay, B.D., Barker, F.K., Mays Jr, H.L., Doucet, S.M., Hill, G.E., 2010. A molecular phylogenetic hypothesis for the manakins (Aves: Pipridae). Mol Phylogenet Evol 55, 733-737.
Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L., 2016. Analysis of a Rapid Evolutionary Radiation Using Ultraconserved Elements: Evidence for a Bias in Some Multispecies Coalescent Methods. Syst Biol 65, 612-627.
Mirarab, S., Bayzid, M.S., Warnow, T., 2016. Evaluating Summary Methods for Multilocus Species Tree Estimation in the Presence of Incomplete Lineage Sorting. Syst Biol 65, 366-380.
Mirarab, S., Warnow, T., 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. Bioinformatics 31, i44-i52.
Molloy, E.K., Warnow, T., 2018. To Include or Not to Include: The Impact of Gene Filtering on Species Tree Estimation Methods. Syst Biol 67, 285-303.
Moyle, R.G., Oliveros, C.H., Andersen, M.J., Hosner, P.A., Benz, B.W., Manthey, J.D., Travers, S.L., Brown, R.M., Faircloth, B.C., 2016. Tectonic collision and uplift of Wallacea triggered the global songbird radiation. Nature Communications 7, 12709.
Ohlson, J., Fjeldså, J., Ericson, P.G.P., 2008. Tyrant flycatchers coming out in the open: phylogeny and ecological radiation of Tyrannidae (Aves, Passeriformes). Zoologica Scripta 37, 315-335.
Ohlson, J.I., Fjeldså, J., Ericson, P.G.P., 2013a. Molecular phylogeny of the manakins (Aves: Passeriformes: Pipridae), with a new classification and the description of a new genus. Mol Phylogenet Evol 69, 796-804.
Ohlson, J.I., Irestedt, M., Ericson, P.G.P., Fjeldså, J.O.N., 2013b. Phylogeny and classification of the New World suboscines (Aves, Passeriformes). Zootaxa 3613, 1-35.
Patel, S., Kimball, R.T., Braun, E.L., 2013. Error in phylogenetic estimation for bushes in the tree of life. Journal of Phylogenetics and Evolutionary Biology 1, 110.
Prum, R.O., 1990. Phylogenetic Analysis of the Evolution of Display Behavior in the Neotropical Manakins (Aves: Pipridae). Ethology 84, 202-231.
Prum, R.O., 1992. Syringeal Morphology, Phylogeny, and Evolution of the Neotropical Manakins (Aves: Pipridae). Am. Mus. Novit. 3043, 65 pp.

Prum, R.O., 1994. Phylogenetic analysis of the evolution of alternative social behavior in the manakins (Aves: Pipridae). Evolution 48, 1657-1675.
Prum, R.O., 1997. Phylogenetic Tests of Alternative Intersexual Selection Mechanisms: Trait Macroevolution in a Polygynous Clade (Aves: Pipridae). The American Naturalist 149, 668-692.
Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature 526, 569.
Pyron, R.A., Hendry, C.R., Chou, V.M., Lemmon, E.M., Lemmon, A.R., Burbrink, F.T., 2014. Effectiveness of phylogenomic data and coalescent species-tree methods for resolving difficult nodes in the phylogeny of advanced snakes (Serpentes: Caenophidia). Mol Phylogenet Evol 81, 221-231.
Rannala, B., Yang, Z., 2017. Efficient Bayesian Species Tree Inference under the Multispecies Coalescent. Syst Biol 66, 823-842.
Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L., Harshman, J., Huddleston, C.J., Kingston, S., Marks, B.D., Miglia, K.J., Moore, W.S., Sheldon, F.H., Witt, C.C., Yuri, T., Braun, E.L., 2017. Why Do Phylogenomic Data Sets Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than Taxon Sampling. Syst Biol 66, 857-879.
Rêgo, P.S., Araripe, J., Marceliano, M.L.V., Sampaio, I., Schneider, H., 2007. Phylogenetic analyses of the genera *Pipra*, *Lepidothrix* and *Dixiphia* (Pipridae, Passeriformes) using partial cytochrome b and 16S mtDNA genes. Zoologica Scripta 36, 565-575.
Remsen, J.V., Jr, Areta, J.I., Cadena, C.D., Claramunt, S., Jaramillo, A., Pacheco, J.F., Robbins, M.B., Stiles, F.G., Stoz, D.F., Zimmer, K.J., 2018. A classification of the bird species of South America. Version: 8 May 2019. American Ornithologists' Union http://www.museum.lsu.edu/~Remsen/SACCBaseline.htm.
Roch, S., Steel, M., 2015. Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theoretical Population Biology 100, 56-62.
Roch, S., Warnow, T., 2015. On the robustness to gene tree estimation error (or lack thereof) of coalescent-based species tree methods. Syst Biol.
Salichos, L., Rokas, A., 2013. Inferring ancient divergences requires genes with strong phylogenetic signals. Nature 497, 327-331.
Sayyari, E., Mirarab, S., 2016. Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Mol Biol Evol 33, 1654-1668.
Seo, T.-K., 2008. Calculating Bootstrap Probabilities of Phylogeny Using Multilocus Sequence Data. Mol Biol Evol 25, 960-971.
Silva, S.M., Agne, C.E., Aleixo, A., Bonatto, S.L., 2018. Phylogeny and systematics of *Chiroxiphia* and *Antilophia* manakins (Aves, Pipridae). Mol Phylogenet Evol 127, 706-711.
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.
Song, S., Liu, L., Edwards, S.V., Wu, S., 2012. Resolving conflict in eutherian mammal phylogeny using phylogenomics and the multispecies coalescent model. Proceedings of the National Academy of Sciences 109, 14942-14947.
Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Mol Phylogenet Evol 94, Part A, 1-33.
Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312-1313.
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. Proceedings of the Royal Society B: Biological Sciences 281.

Swofford, D.L., 2017. PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods). Version 4.0. Sinauer Associates, Sunderland, Massachusetts.
Tello, J.G., Moyle, R.G., Marchese, D.J., Cracraft, J., 2009. Phylogeny and phylogenetic classification of the tyrant flycatchers, cotingas, manakins, and their allies (Aves: Tyrannides). Cladistics 25, 429-467.
Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., Ortí, G., 2015. Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. PLoS Currents 7, ecurrents.tol.34260cc27551a34527b34124ec34265f36334b34266be.
Xi, Z., Liu, L., Davis, C.C., 2015. Genes with minimal phylogenetic information are problematic for coalescent analyses when gene tree estimation is biased. Mol Phylogenet Evol 92, 63-71.

Supplementary Material

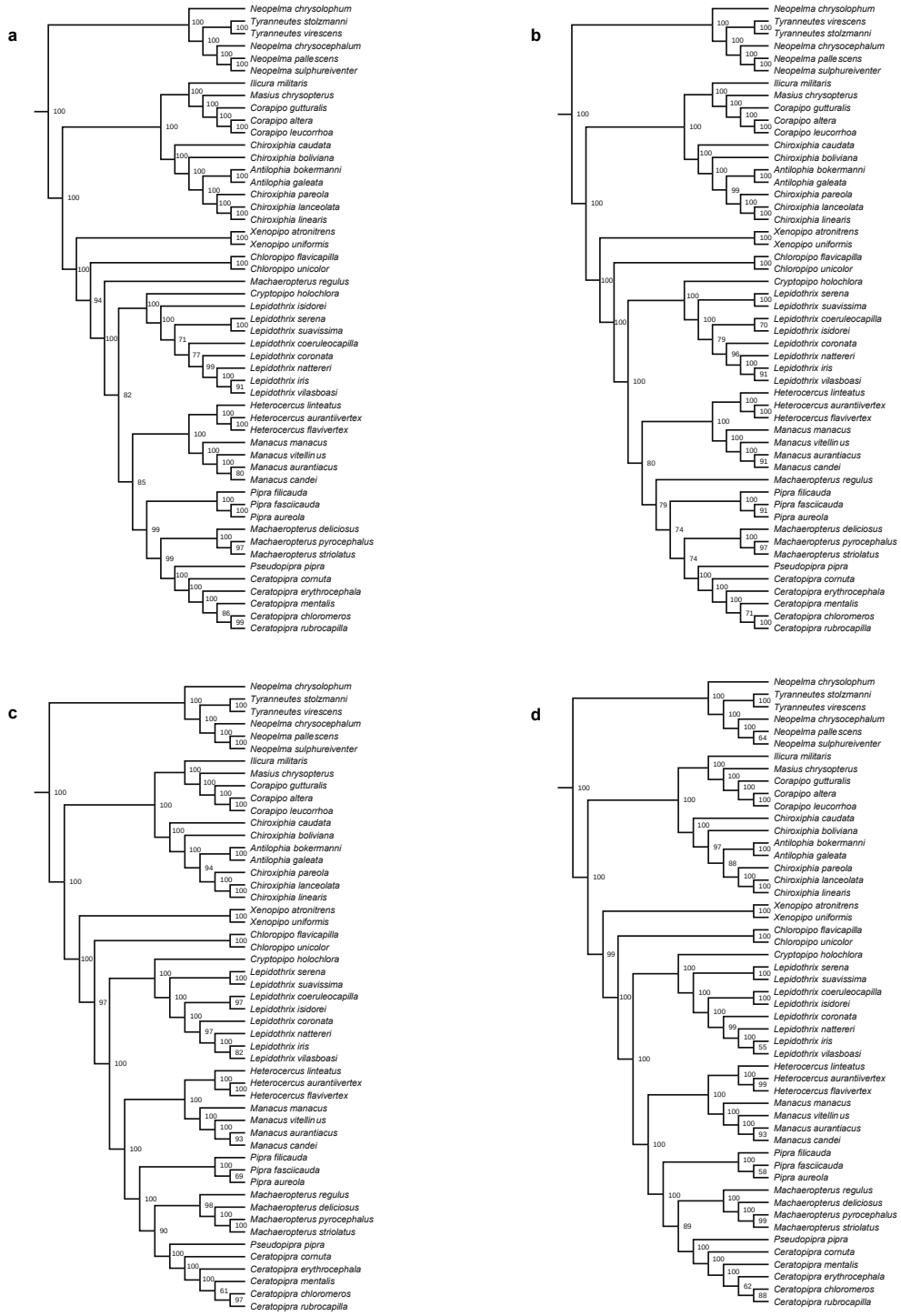


Figure S1. ASTRAL consensus trees estimated using (a) all loci with informative sites, (b) the 75% most informative loci, (c) the 50% most informative loci, and (d) the 25% most informative loci.

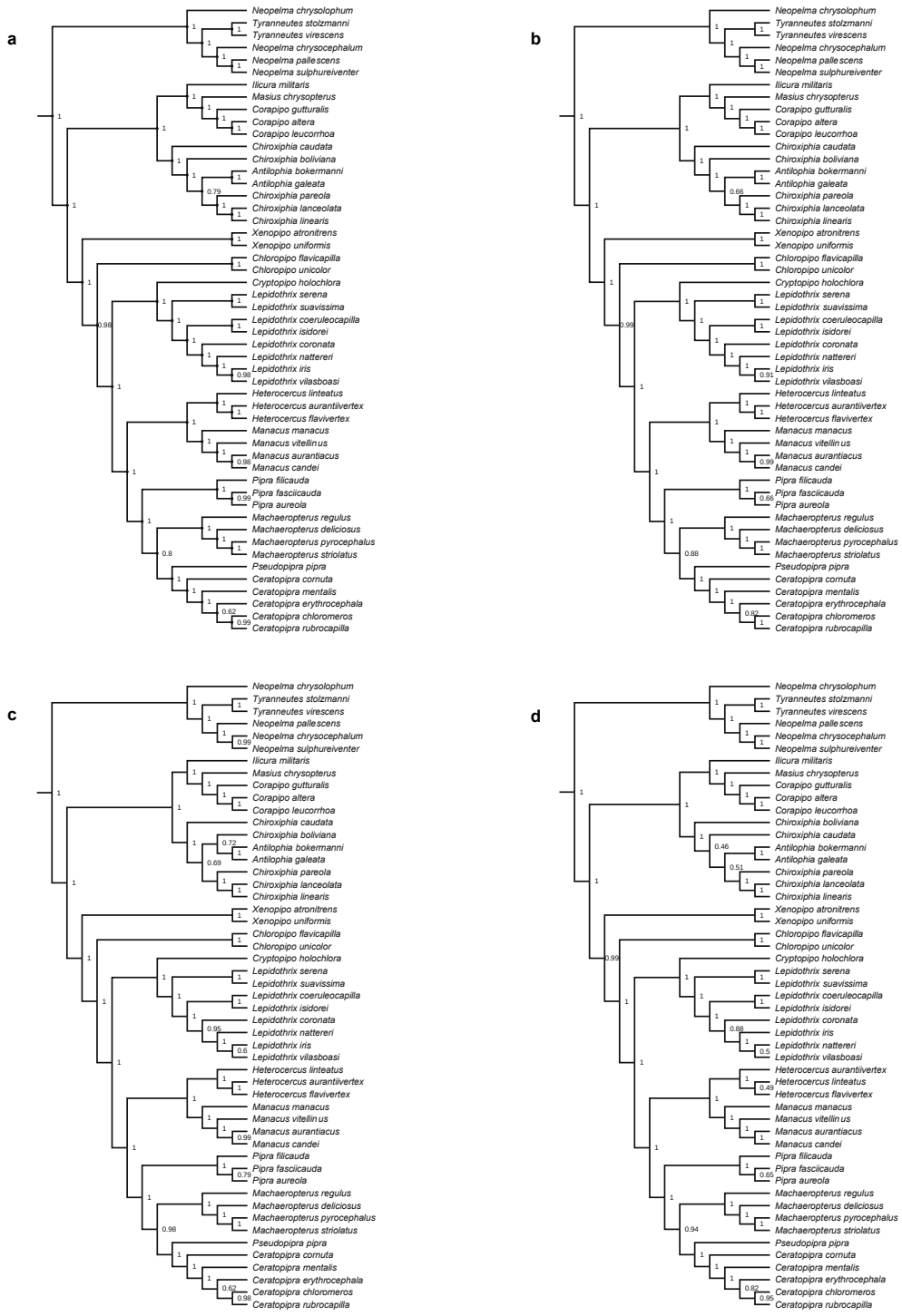


Figure S2. ASTRAL best trees estimated using (a) all loci with informative sites, (b) the 75% most informative loci, (c) the 50% most informative loci, and (d) the 25% most informative loci.

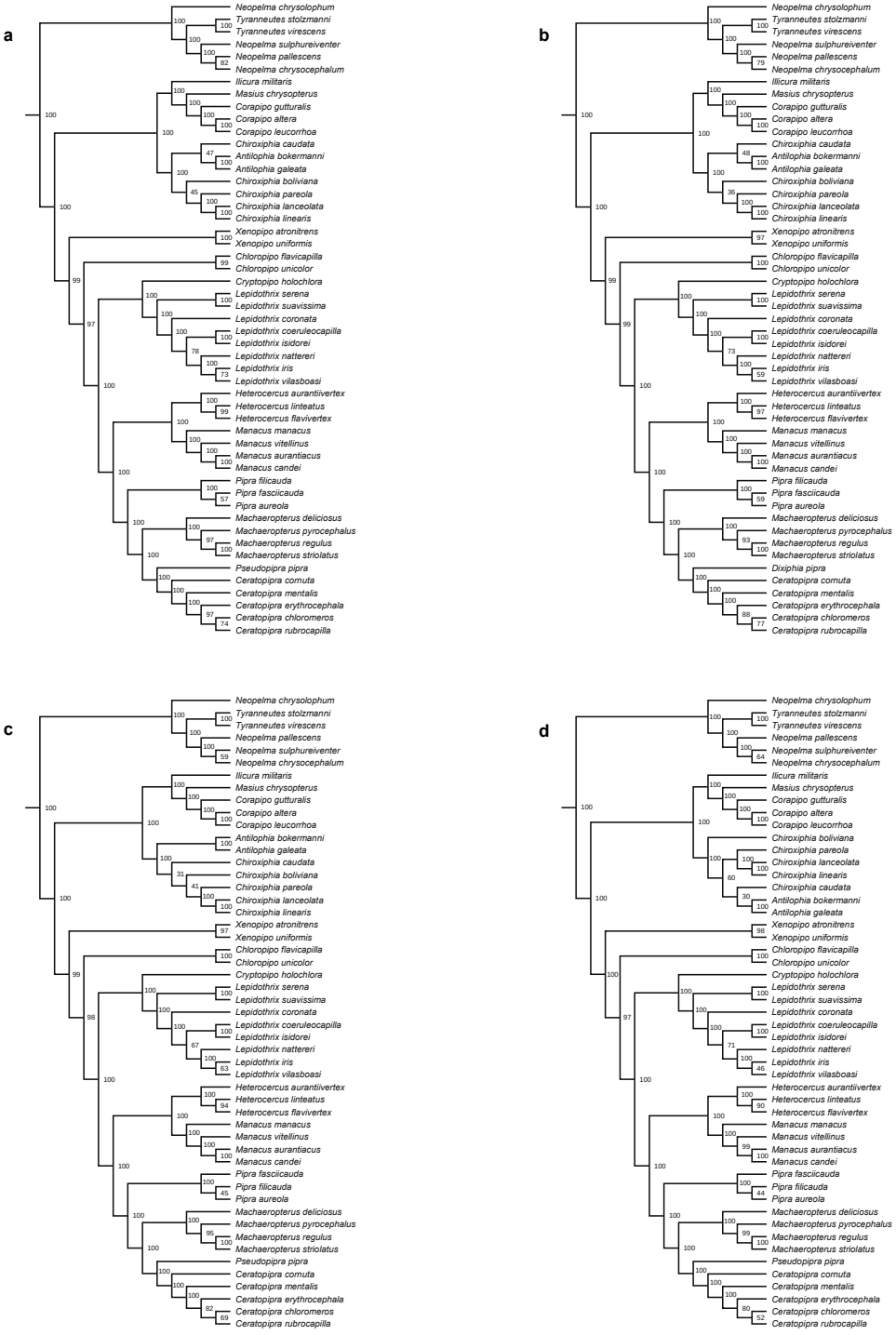


Figure S3. SVDQuartets trees estimated using (a) all loci with informative sites, (b) the 75% most informative loci, (c) the 50% most informative loci, and (d) the 25% most informative loci.

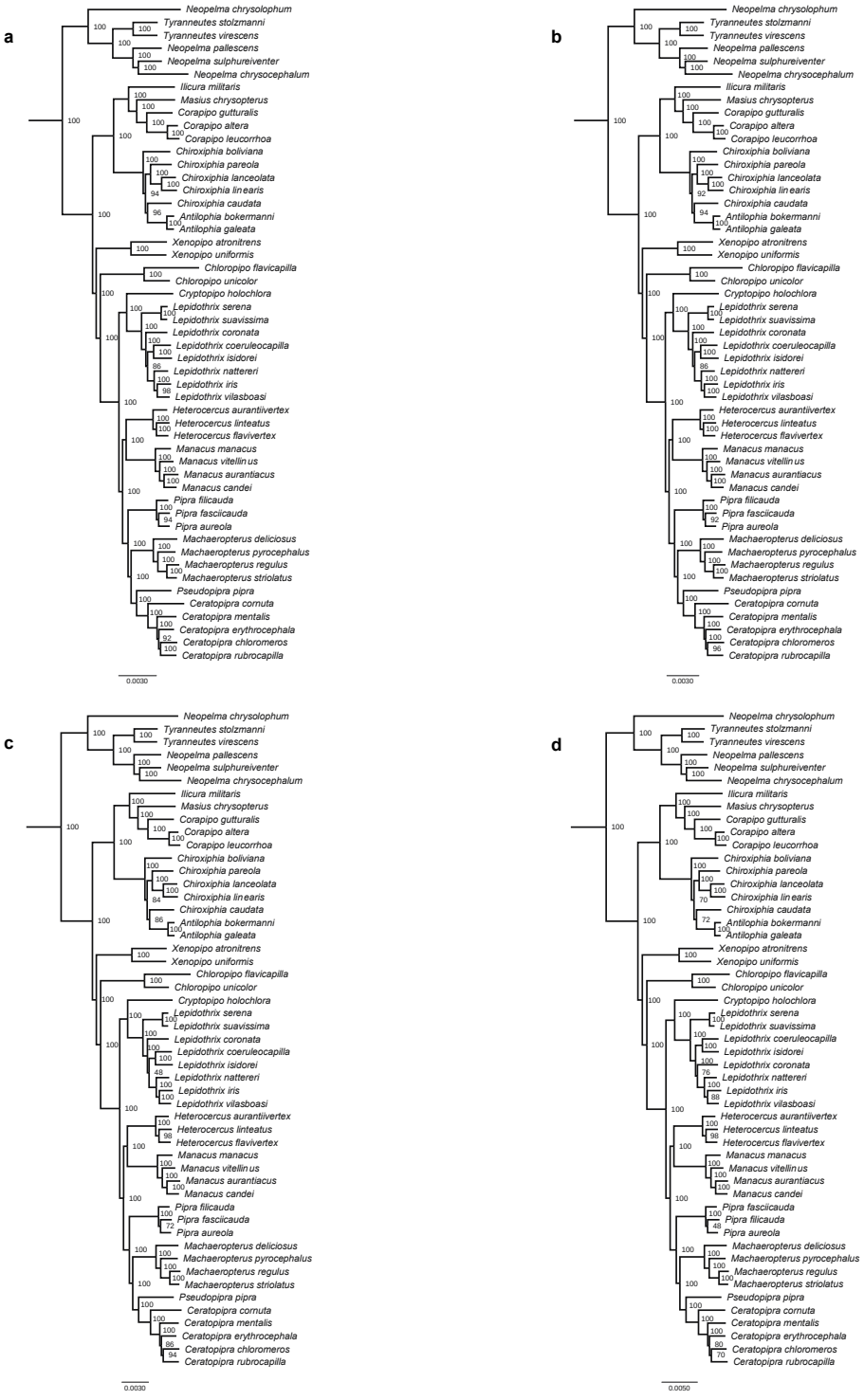


Figure S4. RAxML trees estimated using (a) all loci with informative sites, (b) the 75% most informative loci, (c) the 50% most informative loci, and (d) the 25% most informative loci.

species	voucher	institution
Antilophia bokermanni	Abo1	LSU
Antilophia galeata	L14634a	
Ceratopipra mentalis	L60940	
Chiroxiphia boliviana	L39150	LSU
Chiroxiphia caudata	A2458	
Chiroxiphia lanceolata	L46674	
Chiroxiphia linearis	W104044	
Chloropipo flavicapilla	H7851	
Corapipo altera	LSU28374	
Corapipo gutturalis	L48430	
Heterocercus aurantiivertex	LAlvarez	
Heterocercus flavivertex	A12395	
Heterocercus linteatus	L14715	
Ilicura militaris	GSB15	
Lepidothrix coeruleocapilla	L1985	
Lepidothrix coronata	MPEG62153	
Lepidothrix iris	FMNHWM280	
Lepidothrix isidorei	IC35523	
Lepidothrix nattereri	MPEG58232	
Lepidothrix serena	AMNHDOT12337	
Lepidothrix suavissima	AMNHDOT12049	
Lepidothrix vilasboasi	MPEG59270	
Machaeropterus deliciosus	L11828	
Machaeropterus pyrocephalus	L22641	
Machaeropterus striolatus	L42519	
Manacus aurantiacus	L71959	
Manacus candei	L71894	
Manacus vitellinus	L28569	
Masius chrysopterus	H7438	
Neopelma chrysocephalum	A13890	
Neopelma chrysolophum	F395453	
Neopelma pallescens	L14646	
Neopelma sulphureiventer	L15218	
Pipra filicauda	L20244	
Tyranneutes stolzmanni	L13803	
Tyranneutes virescens	L25483	
Xenopipo atronitrens	L42648	
Xenopipo uniformis	SB23771	
Ceratopipra chloromeros	AMNHDOT18183	
Ceratopipra cornuta	AMNHDOT4822	
Ceratopipra erythrocephala	INPAA18275	
Ceratopipra rubrocapilla	INPAA341	
Chiroxiphia pareola	INPAA238	

Chloropipo unicolor	FMNH474328
Corapipo leucorrhoa	ICN33757
Cryptopipo holochlora	LSUMZ11843
Pseudopipra pipra	FMNH474348
Machaeropterus regulus	AMNH468596
Manacus manacus	INPAA15127
Onychorhynchus coronatus	FMNH457461
Pachyramphus minor	INPAA5292
Pipra aureola	INPAA11923
Pipra fasciicauda	AMNHDOT18402
Pyroderus scutatus	FMNH474375

sample	reads	total bp
Antilophia_bokermanni_Abo1	7619933	742928581
Antilophia_galeata_L14634a	7237550	698641107
Ceratopipra_mentalis_L60940	2859133	278170359
Chiroxiphia_boliviana_L39150	2460332	238523023
Chiroxiphia_caudata_A2458	11458277	1100038859
Chiroxiphia_lanceolata_L46674	3475383	337410654
Chiroxiphia_linearis_W104044	4050114	391599956
Chloropipo_flavicapilla_H7851	9631279	848224513
Corapipo_altera_LSU28374	10065149	1397244908
Corapipo_gutturalis_L48430	11483556	1006320934
Heterocercus_aurantiivertex_LAlvarez	4853179	472275517
Heterocercus_flavivertex_A12395	4874132	593634937
Heterocercus_linteatus_L14715	4168311	365769531
Ilicura_militaris_GSB15	4777716	419184189
Lepidothrix_coeruleocapilla_L1985	5771568	562588888
Lepidothrix_coronata_MPEG62153	3407150	331161711
Lepidothrix_iris_FMNHWM280	5175954	503559556
Lepidothrix_isidorei_IC35523	19745926	1739336813
Lepidothrix_nattereri_MPEG58232	5870050	572431591
Lepidothrix_serena_AMNHDOT12337	7755711	753613492
Lepidothrix_suavissima_AMNHDOT12049	3927138	381661617
Lepidothrix_vilasboasi_MPEG59270	6225372	606594655
Machaeropterus_deliciosus_L11828	5763989	561164921
Machaeropterus_pyrocephalus_L22641	8533344	831417015
Machaeropterus_striolatus_L42519	5224136	507224526
Manacus_aurantiacus_L71959	4967010	479772449
Manacus_candei_L71894	9058108	875072436
Manacus_vitellinus_L28569	5243541	509832613
Masius_chrysopterus_H7438	17685287	1700231391
Neopelma_chrysocephalum_A13890	6501392	569226218
Neopelma_chrysolophum_F395453	10979893	966830366
Neopelma_pallescens_L14646	6589330	643858627
Neopelma_sulphureiventer_L15218	5302945	516188986
Pipra_filicauda_L20244	3062787	296218738
Tyranneutes_stolzmanni_L13803	3672094	358861703
Tyranneutes_virescens_L25483	4014624	352869806
Xenopipo_atronitrens_L42648	2752123	241638420
Xenopipo_uniformis_SB23771	3043140	286629064
Ceratopipra_chloromeros_AMNHDOT18183	1689498	168319148
Ceratopipra_cornuta_AMNHDOT4822	3198355	318236133
Ceratopipra_erythrocephala_INPAA18275	3866865	386270445
Ceratopipra_rubrocapilla_INPAA341	3943078	392898752
Chiroxiphia_pareola_INPAA238	4582380	456067894

Chloropipo_unicolor_FMNH474328	3021423	297008360
Corapipo_leucorrhoa_ICN33757	7172284	707760046
Cryptopipo_holochlora_LSUMZ11843	2092664	208766670
Pseudopipra_pipra_FMNH474348	2494300	248451541
Machaeropterus_regulus_AMNH468596	5831469	581263214
Manacus_manacus_INPAA15127	3152359	314484704
Onychorhynchus_coronatus_FMNH457461	2705906	268233355
Pachyramphus_minor_INPAA5292	3361262	334758030
Pipra_aureola_INPAA11923	4551975	454075173
Pipra_fasciicauda_AMNHDOT18402	3271008	326189528
Pyroderus_scutatus_FMNH474375	2052055	204338828
mean	**5671732**	
95% CI	**952298**	

mean length	95% CI length	min length	max length	median length
97,5	0,0035	40	100	100
96,5	0,0038	40	100	100
97,3	0,0053	40	100	100
96,9	0,0060	40	100	100
96,0	0,0036	40	100	100
97,1	0,0049	40	100	100
96,7	0,0049	40	100	100
88,1	0,0039	40	100	78
138,8	0,0074	40	151	151
87,6	0,0037	40	100	78
97,3	0,0040	40	100	100
121,8	0,0052	40	125	125
87,8	0,0060	40	100	78
87,7	0,0056	40	100	78
97,5	0,0036	40	100	100
97,2	0,0055	40	100	100
97,3	0,0044	40	100	100
88,1	0,0027	40	100	78
97,5	0,0040	40	100	100
97,2	0,0037	40	100	100
97,2	0,0051	40	100	100
97,4	0,0039	40	100	100
97,4	0,0037	40	100	100
97,4	0,0030	40	100	100
97,1	0,0040	40	100	100
96,6	0,0046	40	100	100
96,6	0,0034	40	100	100
97,2	0,0040	40	100	100
96,1	0,0029	40	100	100
87,6	0,0049	40	100	78
88,1	0,0037	40	100	78
97,7	0,0032	40	100	100
97,3	0,0038	40	100	100
96,7	0,0056	40	100	100
97,7	0,0043	40	100	100
87,9	0,0062	40	100	78
87,8	0,0075	40	100	78
94,2	0,0076	40	100	100
99,6	0,0056	40	101	101
99,5	0,0042	40	101	101
99,9	0,0033	40	101	101
99,6	0,0036	40	101	101
99,5	0,0035	40	101	101

98,3	0,0044	40	100	100
98,7	0,0025	40	100	100
99,8	0,0048	40	101	101
99,6	0,0046	40	101	101
99,7	0,0027	40	101	101
99,8	0,0039	40	101	101
99,1	0,0052	40	101	101
99,6	0,0040	40	101	101
99,8	0,0032	40	101	101
99,7	0,0039	40	101	101
99,6	0,0052	40	101	101
97,4				
2,1				

sample	contigs	total bp	mean length
Antilophia_bokermanni_Abo1	3157	2032914	643,9
Antilophia_galeata_L14634a	19320	7158431	370,5
Ceratopipra_chloromeros_L40196	6690	3088163	461,6
Ceratopipra_cornuta_L48294	3369	1714531	508,9
Ceratopipra_erythrocephala_INPAA18275	6997	3879436	554,4
Ceratopipra_mentalis_L60940	5693	2738075	481,0
Ceratopipra_rubrocapilla_INPAA341	7292	3374354	462,7
Chiroxiphia_boliviana_L39150	5330	2767866	519,3
Chiroxiphia_caudata_A2458	2914	1397154	479,5
Chiroxiphia_lanceolata_L46674	8004	3206396	400,6
Chiroxiphia_linearis_W104044	7166	3435968	479,5
Chiroxiphia_pareola_INPAA238	9458	3921117	414,6
Chloropipo_flavicapilla_H7851	3328	1677717	504,1
Chloropipo_unicolor_FMNH474328	5112	3011449	589,1
Corapipo_altera_LSU28374	185267	49501745	267,2
Corapipo_gutturalis_L48430	2957	1792073	606,0
Corapipo_leucorrhoa_ICN33757	7456	3409738	457,3
Cryptopipo_holochlora_LSUMZ11843	5421	3129077	577,2
Pseudopipra_pipra_FMNH474348	5903	3474383	588,6
Heterocercus_aurantiivertex_LAlvarez	10220	4259645	416,8
Heterocercus_flavivertex_A12395	41899	13656892	325,9
Heterocercus_linteatus_L14715	2422	1210025	499,6
Ilicura_militaris_GSB15	2447	1239742	506,6
Lepidothrix_coeruleocapilla_L1985	16199	5751728	355,1
Lepidothrix_coronata_MPEG62153	2900	1796151	619,4
Lepidothrix_iris_FMNHWM280	2764	1770099	640,4
Lepidothrix_isidorei_IC35523	2789	1385757	496,9
Lepidothrix_nattereri_MPEG58232	2790	1732300	620,9
Lepidothrix_serena_AMNHDOT12337	3127	2325256	743,6
Lepidothrix_suavissima_AMNHDOT12049	2704	2067455	764,6
Lepidothrix_vilasboasi_MPEG59270	3065	1884020	614,7
Machaeropterus_deliciosus_L11828	14036	5506205	392,3
Machaeropterus_pyrocephalus_L22641	28404	9395926	330,8
Machaeropterus_regulus_AMNH468596	9181	2858386	311,3
Machaeropterus_striolatus_L42519	12693	4738901	373,3
Manacus_aurantiacus_L71959	11170	4784617	428,3
Manacus_candei_L71894	34737	11526029	331,8
Manacus_manacus_INPAA15127	6497	3726880	573,6
Manacus_vitellinus_L28569	10777	4782203	443,7
Masius_chrysopterus_H7438	3919	1762446	449,7
Neopelma_chrysocephalum_A13890	4521	2041492	451,6
Neopelma_chrysolophum_F395453	2954	1480418	501,2
Neopelma_pallescens_L14646	16182	6061783	374,6

Neopelma_sulphureiventer_L15218	12772	5062260	396,4
Onychorhynchus_coronatus_FMNH457461	5861	3792931	647,1
Pachyramphus_minor_INPAA5292	6453	3289506	509,8
Pipra_aureola_INPAA11923	7586	4326254	570,3
Pipra_fasciicauda_AMNHDOT18402	6607	3788750	573,4
Pipra_filicauda_L20244	6703	3165036	472,2
Pyroderus_scutatus_FMNH474375	5466	3231390	591,2
Tyranneutes_stolzmanni_L13803	7515	3364691	447,7
Tyranneutes_virescens_L25483	3121	1338529	428,9
Xenopipo_atronitrens_L42648	2457	1079041	439,2
Xenopipo_uniformis_SB23771	3903	2167668	555,4
mean	**11438**		**492**
95% CI	**6753**		**29**

95% CI length	min length	max length	median legnth	contigs >1kb	coverage (x)
3,8	201	3300	713	31	149,1
2,4	201	17017	249	427	10
4,2	201	17025	356	126	21,2
3,5	201	3069	542	32	37,7
4,1	201	17083	576	142	25,8
4,6	201	17054	416	82	20,4
3,8	201	16002	332	137	19,9
5,4	201	16985	494,5	149	19,2
2,5	201	3028	504	6	52
3,3	201	16972	278	88	14,4
4,5	201	16975	313	192	15,1
3,8	201	17002	263,5	138	15,7
3,1	201	3115	535,5	5	39,4
6,1	201	17075	543	552	11,7
0,3	201	17054	240	462	3,8
3,6	201	3364	644	29	28,9
3,6	201	3595	277	348	13,4
5,2	201	9922	584	89	18,9
4,3	201	17106	619	92	19,1
3,6	201	17056	267	209	15,9
2,3	201	14474	222	495	8,1
2,6	201	3037	512	4	18,9
2,4	201	3023	518	4	23,3
2,3	201	17128	257	209	15,1
3,3	203	3208	669	21	59,9
3,5	201	3270	688	18	55,7
2,4	201	3055	516	4	39,4
3,3	201	3226	666	16	76,2
4,6	201	3418	831	176	158
4,5	201	3237	828	163	55,2
3,4	201	3217	673	19	146,6
2,8	201	17043	269	224	16
1,9	201	17106	246	301	11,2
2,6	201	9832	264	139	16
2,6	201	17084	259	159	14,8
3,9	201	16448	279	247	13,5
1,4	201	17017	250	442	9,1
4,8	201	15947	592	129	21,6
3,5	201	17044	291	269	17,6
2,6	201	3023	491	7	44
2,8	201	2195	400	15	25,2
2,9	201	3140	523,5	5	19,7
2,5	201	17070	263,5	194	16,4

3,0	201	17113	273	240	16,3
5,6	201	16814	671	223	20,7
3,2	201	4194	516	126	23,4
4,4	201	17027	582	215	25,5
4,1	201	10170	590	168	22,9
4,2	201	17062	354	129	14,3
6,6	201	14391	603	78	18
4,2	201	16993	312	143	18,2
2,8	201	1694	409	4	9,8
2,4	201	2839	449	4	14,7
5,8	201	17153	620	47	33,2

31
9