# Use of genomic data to resolve gene tree discordance in a Southeast Asian genus: Readdressing paraphyly in the spiderhunter (Nectariniidae, Arachnothera) phylogeny

Use of genomic data to resolve gene tree discordance in a Southeast Asian genus: Readdressing paraphyly in the spiderhunter (Nectariniidae, *Arachnothera*) phylogeny

By

Luke C. Campillo

Submitted to the graduate degree program in Ecology and Evolutionary Biology and the Graduate Faculty of the University of Kansas in partial fulfillment of the requirements for the degree of Master of Arts.

_____

Chairperson Dr. Robert Moyle

_____

Dr. Rich Glor

_____

Dr. Jamie Walters

Date Defended: 25 March 2016

The Thesis Committee for Luke C. Campillo

certifies that this is the approved version of the following thesis:

Use of genomic data to resolve gene tree discordance in a Southeast Asian genus: Readdressing paraphyly in the spiderhunter (Nectariniidae, *Arachnothera*) phylogeny

_____

Chairperson Dr. Robert Moyle

Date approved: 25 March 2016

**Abstract**

Reduced representation genomic sequencing methods provide a fast and relatively inexpensive way to gather sequence data from thousands of loci throughout the genome. These data can be used to test previous phylogenetic hypotheses produced from limited numbers of mitochondrial and nuclear loci that often revealed intriguing, but conflicting results. In this paper, we use phylogenomic data to revisit molecular phylogenetic work that while clarifying taxonomic relationships within spiderhunters (Aves: *Arachnothera*), also questioned monophyly of this distinctive genus. Previous phylogenetic analysis of two molecular markers (one mitochondrial and one nuclear locus) produced conflicting topologies, yet both strongly supported non-monophyly of *Arachnothera*. This present study assesses the nature of pervasive gene tree discordance in these birds and investigates phylogenetic relationships within spiderhunters. To accomplish this, we used target-capture of ultra conserved elements (UCEs) to produce a phylogenomic data matrix used to infer the evolutionary history of *Arachnothera*. Although we recovered many gene tree topologies, concatenated and species-tree methods of analysis converged on a phylogeny with strong support for monophyly of *Arachnothera*. The consistency in analytical results confers confidence that gene tree conflict has been resolved in this enigmatic genus.

**Acknowledgements**

I would first like to thank my parents, who have been a constant source of support throughout my life. I have induced a fair bit of anxiety for my parents over the years, but that never lessened their resolve to help me turn out the way I did. Because of my parents' love and support, I have been able to follow my dreams and make my own path in life. In addition to my parents, I want to thank my siblings, Sarah and Jess. Those two have been there for me through thick and thin. I think all three of us would agree that our parents are responsible for our cohesiveness, something that is rare among adult siblings. I know that no matter what, my brother and sister are in my corner. I can't thank them enough for that. I also want to thank Briana for putting up with me and supporting me throughout this process. Defending my thesis has challenged each of us in different ways, but we weathered the storm together.

In addition to my immediate family, I would like to give special thanks to my extended family. Many of my aunts, uncles and cousins have read my published work, and although they probably don't understand most of the technical science involved they are excited and proud of me nonetheless. I want to thank my friends, especially the Westmont boys and my two great pals from Drake University for embracing "the Birdman" and giving me a strong non-academic support group.

I want to thank Dr. Rob Moyle for taking me on as a grad student when everyone else rejected me. I hope that I have not let him down. As I progress through my career I hope to continue developing our personal and professional relationship. I want to thank all the past and present members of the Moyle Lab for their support and guidance throughout my time at KU. I especially want to thank Joe Manthey for being a great friend and great mentor and Mark Robbins for his sage advice and continual comedic relief.

I want to thank my extended support group in the BI, especially the herpers and the bug people. It has been a great experience learning from all their various expertise as we navigate grad school together. I would like to thank my committee members, Rich Glor and Jamie Walters for their guidance and support throughout this entire process.

And last but definitely not least, I would like to thank Dr. Muir Eaton. Muir is a great man who introduced me to ornithology and set me on this path. I cannot fully express the gratitude I feel towards Muir, but suffice it to say that I wouldn't be in this position today without his mentorship.

**Table of Contents**

## Introduction

Over the past several decades, analysis of DNA sequence data has clarified phylogenetic relationships in many groups of organisms (e.g. Wiens and Reeder, 1997; Bates et al., 1999). Despite the novel insights provided by sequence data for many taxonomic groups, this work has also revealed substantial limitations to inferring relationships among species. A major limitation to phylogenetic inference is gene tree discordance. Gene tree discordance is the discrepancy between individual gene tree topologies and the underlying species phylogeny. This phenomenon can occur for a number of reasons, such as deep coalescence or horizontal gene transfer (Maddison, 1997), and can have grave effects on phylogenetic estimation (see McGuire et al., 2007). In most cases of deep coalescence, adding loci to the analysis can yield a clearer idea of the species tree and identification of anomalous gene trees (e.g. Andersen et al., 2015; Brennan et al., 2016). However, some combinations of short internodes can produce a situation called the anomaly zone, in which the most common gene tree topologies do not match the species tree (Degnan and Rosenberg, 2006). Addition of loci in anomaly-zone situations can result in high support for an incorrect topology in some types of analysis (Philippe et al., 2011).

One particular type of gene tree discordance—between mitochondrial and nuclear markers—has been identified in many studies (reviewed by Toews and Brelsford, 2012). Often, these studies involved few loci produced with Sanger sequencing methods, and resolution of gene tree discordance was not possible with limited sampling. Moyle et al. (2011) documented discordance between mitochondrial and nuclear loci for a small group of Asian passerine birds, the spiderhunters (Aves: Nectariniidae: *Arachnothera*). Phylogenetic analysis of each marker supported paraphyly of *Arachnothera*, but each produced a different topology (Fig. 1). The mitochondrial markers strongly supported the inclusion of a monotypic genus, *Hypogramma*,

within *Arachnothera*. Analysis of a nuclear intron supported more extensive paraphyly, with strong support for a sister relationship between a clade of spiderhunters and several of the putative outgroup taxa from several genera of sunbirds (Nectarinidae). Analysis of the concatenated dataset produced a species tree estimate with a topology similar to the tree produced from the mitochondrial markers, highly supporting *Hypogramma* as sister to one of two well-supported clades within *Arachnothera*. These results were somewhat surprising because *Arachnothera* seemed a cohesive group, and monophyly had not previously been questioned. Traditional taxonomy considered the spiderhunters (Aves: *Arachnothera*) to comprise thirteen named species, however multiple taxonomic lists (i.e. IOC World Bird List v6.1; Clements Checklist v2015) subsumed *Hypogramma hypogrammicum* into *Arachnothera* following Moyle et al. (2011). Despite the accepted taxonomic changes, we will use traditional nomenclature *sensu stricto* to refer to the Purple-naped Sunbird as *Hypogramma hypogrammicum* and refer to the remaining species as *Arachnothera* or spiderhunters throughout.

*Arachnothera* is a relatively small genus (14 species) in the sunbird family, Nectariniidae (Clements et al., 2015). Sunbirds have a strictly old world distribution, and spiderhunters' range extends from mainland SE Asia through the Philippines and Indonesia. Interestingly, their range does not extend east of Wallace's line. Despite the name, spiderhunters feed on a variety of nectar and invertebrates and not exclusively spiders. Spiderhunters share traits with other sunbirds (e.g. long, decurved bill), however the clade is linked by unique physiological and behavioral traits. In respect to body size, spiderhunters are among the largest of all Nectariniidae. Spiderhunters are exceptional among sunbirds in their absence of iridescent feather patches and lack of overt sexual dimorphism (Cheke and Mann, 2001; del Hoyo, Elliot, & Christie, 2008). Spiderhunters have a tongue morphology that is distinctive among sunbirds, thus they can access

nectar from different flowers. Lastly, male and female spiderhunters share nest building and incubation responsibilities (Cheke and Mann, 2001).

Reduced representation genomic sequencing techniques (e.g., Restriction-site associated DNA sequencing (RAD-seq), Miller et al., 2007; target-capture of ultraconserved elements (UCEs; Faircloth et al., 2012) have developed as cost-effective methods to obtain sequence data from thousands of loci throughout the genome, thus increasing the inferential power to produce a genome-wide assessment of phylogenetic history. Analysis of hundreds or thousands of loci from throughout the genome allows investigation, and potential resolution, of gene tree discordance evident in Sanger sequencing results. Here we revisit the phylogenetic relationships of spiderhunters using DNA sequences of thousands of markers from throughout the avian genome. In this study, we performed target capture sequencing of UCEs (Faircloth et al., 2012) and their variable flanking regions to amass a genome wide dataset to estimate phylogenetic relationships in this genus. Following sequencing, we used multiple species tree reconstruction methods to produce species phylogenies from the reduced representation genomic dataset. The goals of this study were: 1. Examine heterogeneity across all gene tree topologies to investigate how a diversity of gene trees can affect species tree estimation; 2. Reconstruct a best estimate for the species phylogeny to resolve relationships within the genus *Arachnothera* and assess UCE gene trees to see if they reflect the discordance found in previous Sanger sequencing results.

## Methods

### *Sampling and Sample Preparation*

We extracted DNA from fresh tissue of twelve of fourteen described species of *Arachnothera* (including *Hypogramma hypogrammicum,* but excluding *A. affinis* and *A. juliae*), four additional

sunbirds, and one outgroup taxon (Table 1) using a QUIAGEN DNeasy blood and tissue kit. We quantified DNA concentrations with Qubit Flourometric Quantitation (Life Technologies) and standardized all samples to 10 ng/uL.

*Sequence Capture Protocol and Bioinformatics*

We used the Tetrapods-UCE-5Kv1 probe set (available at ultraconserved.org) to perform target-capture of 5060 UCE loci. We fragmented the genomic DNA (500-600bp) using a Covaris S220 focus-ultrasonicator, using settings optimized for other UCE projects using avian samples; sonicator settings were as follows: 175 W peak incident power, 2% duty factor, and 200 cycles per burst for 45 seconds. We then prepared sequencing libraries using Kapa Biosystems Library Prep Kits (KBLPK; Kit: KK8201) according to KBLPK instructions. Samples were subjected to end repair and A-tailing, and the iTruStub dual indexing system (available at baddna.org) was used to ligate compatible adapters to each sample. Indexed samples were amplified with a 10-cycle PCR and then pooled. We enriched libraries for the 5Kv1 probe set using a Mycroarray MyBaits Kit, followed by a brief PCR to amplify the part of each library enriched for UCEs. We tested libraries for quantity and quality using both quantitative PCR and the Agilent Tapestation at the University of Kansas Genome Sequencing Core Facility. Lastly, libraries were sequenced using 100bp paired-end reads on a partial lane of an Illumina HiSeq2500 at the KU Genome Sequencing Core Facility.

Raw reads were de-multiplexed using the Illumina program CASAVA ver. 1.8.2. We further cleaned the data using *illumiprocessor.py* (ver. 1) to trim adapter contamination and perform a quality check of the reads. Contigs for each species were assembled with Trinity (Grabherr et al., 2011). Subsequent bioinformatics of the UCE data used the PHYLUCE software

package of Python v2.7 scripts (Faircloth, 2015). We ran the *match_contigs_to_probes.py* script to match contigs to the UCE probe set and then aligned the UCE loci into a nexus file with the *seqcap_align_2.py* script, using MAAFT (Katoh and Standley, 2013) for alignment. A combination of custom Python and R (R Development Core Team, 2012) scripts were used to obtain summary statistics and convert the data to various files types for phylogenetic analyses.

*Phylogenetic Inference*

We used two phylogenetic methods to construct a species phylogeny with the complete concatenated matrix of 2,107 loci. First, we used RAxML (GTR + Gamma model of sequence evolution) to construct a maximum likelihood (ML) species tree. Support for the ML species phylogeny was assessed with 1000 rapid bootstrap replicates. We also used SVDquartets (Chifman and Kubatko, 2014) to estimate a species tree from the concatenated data. Quartet methods, including SVDquartets, exhaustively sample sets of four individuals from the data matrix, and then constructs a species tree from all sampled quartets. We implemented the SVDquartets species tree inference using the program PAUP* v4a146 (Swofford, 2015).

To infer species trees from gene trees, we used a two-step process. First, gene trees were estimated for each UCE locus and a multi-locus bootstrap was performed on the entire dataset. Next, the estimated gene trees were used as input for multiple species tree inference methods. These methods fall under the umbrella of the multispecies coalescent model, and will be referred to collectively as MSC methods (Liu et al., 2015). We used the PHYLUCE package (Faircloth, 2015) to estimate gene trees for each locus in RAxML (Stamatakis, 2014) using the *run_raxml_genetress.py* command. We performed 500 multi-locus bootstraps, in which all iterations resample loci within the dataset, as well as bases within each locus (Seo, 2008). With

RAxML gene trees as input, we used five species tree inference methods to estimate species relationships across all taxa: species trees from average ranks of coalescence (STAR; Liu et al., 2009), species tree estimation using average coalescence times (STEAC; Liu et al. 2009), neighbor joining species tree (NJst; Liu et al., 2011), accurate species tree algorithm (ASTRAL; Mirarab et al., 2014), and Maximum pseudo-likelihood of estimating species trees (MP-EST; Liu et al., 2010).

We used two different methods to compare topological congruence between individual gene trees and the topology of the species phylogeny. First, we constructed a 50% consensus tree and then selected the option to "Show frequencies of all observed bipartitions" in PAUP* v4a146 (Swofford, 2015). Given all topological combinations across 2,107 gene trees, the bipartition table recorded the number of gene trees that contained each unique clade. We also used the *CompareTree.pl* command in FastTree (Price et al., 2010) to calculate the proportion of gene trees that contained the same phylogenetic splits present in the species phylogeny.

**Results**

The mean number of UCE loci recovered for all samples (Table 1) was 4099.5, with a range of 3794 (*Arachnothera dilutior*) to 4342 (*Chloropsis soneratti*). The mean contig length across all samples was 851.6bp, with a range of 647.7bp (*Arachnothera clarae*) to 971.5bp (*Dicaeum hypoleucum*). The mean coverage for all samples was 37.8x, with a range of 27.6x (*Arachnothera clarae*) to 54.1x (*Chloropsis soneratti*). Maximum likelihood (ML) analysis of the concatenated matrix produced a strongly supported phylogenetic tree with all nodes receiving 100% bootstrap support. The ML tree topology was similar to that recovered by Moyle et al. (2011), with *Arachnothera* divided into the same "Clade A" and "Clade B" as before (see Fig.

1). In contrast to the Sanger based ML tree, our ML analysis (along with subsequent species phylogenies discussed below) recovered monophyly of *Arachnothera* with 100% bootstrap support. *Hypogramma hypogrammicum* was placed as the sister taxa to all spiderhunters, with other sunbird genera more distantly related still. The consensus tree produced from the SVDquartets analysis was identical in topology and support to the tree produced by RAxML. Additionally, both concatenation methods (RAxML & SVDquartets) produced trees identical in topology to those recovered by all MSC analyses performed.

Phylogenetic analysis of individual UCE loci produced many different gene tree topologies, including many that were similar to the trees produced by Moyle et al. (2011). The most frequent topology, meaning the unique topology shared by the most loci, was recovered only six times out of the 2,107 possible gene trees. This most frequent topology includes *Hypogramma* nested within *Arachnothera*, sister to "Clade A", matching the topology reported by Moyle et al. (2011). The majority of loci (n=2,038) were found to have their own gene tree topology (Table 2). Summarizing relationships across all gene trees revealed substantial discordance and support for conflicting topologies. For example, the bipartition representing monophyly of *Arachnothera* was recovered in only ~13.5% of all gene trees and the topologies consistent with the trees produced by Moyle et al. (2011) were recovered in 7.8% (Clade A + *Hypogramma*) and 6.4% (Clade B + Sunbird) of gene trees.

All MSC analyses produced highly supported species tree estimates with identical topologies. Only one analysis, STEAC, recovered bootstrap support values below 90% at any node. The topologies obtained from MSC species tree methods were identical to those from analysis of the concatenated data matrix (Fig 2).

**Discussion**

We recovered a wide variety of gene tree topologies from individual loci, many of which contradicted the species tree topology recovered herein and supported previous results (Moyle et al., 2011), yet phylogenetic analyses of thousands of loci from throughout the genome produced strong support for monophyly of *Arachnothera*. We found evidence for deeply coalescing lineages in some gene tree topologies, but given the number of uninformative loci and diversity of topologies recovered, those gene trees did not have a dramatic impact on the overall conclusion. Visual inspection of the bipartition table did not indicate horizontal gene transfer (HGT) among individuals, but to be safe we implemented a method (ASTRAL; Mirarab et al., 2014) that is statistically consistent despite increased levels of HGT. Despite disparate signal across individual gene trees, multiple analytical approaches (i.e. concatenated and coalescent species tree methods) produced an identical phylogenetic conclusion. Below we discuss details of gene tree discordance and taxonomic implications for the spiderhunters.

*Gene tree conflict*

Increased genetic sampling has been used to produce phylogenomic datasets to clarify problematic evolutionary relationships in many taxa, ranging from plants to ants to birds (e.g. Pease et al., 2016; Blaimer et al., 2015; Manthey et al., 2016). In this study, analysis of genomic data unequivocally supports monophyly of *Arachnothera*, with the monotypic genus *Hypogramma* as its sister taxon, and all other sunbirds more distantly related. Despite short internodes, we produced topologically equivalent estimates of the species phylogeny with all nodes receiving strong support (Fig. 2) across concatenated, or single-site, methods (RAxML and SVDquartets) and MSC methods (STAR, STEAC, NJST, ASTRAL & MPEST). The

consistency of our species tree estimates indicates that gene tree heterogeneity, specifically discordance due to deep coalescence or HGT, did not mislead our analyses.

Strong support for phylogenetic relationships in *Arachnothera* emerged from an incredible diversity of gene tree topologies. We used bipartition analyses to investigated gene tree heterogeneity and found that very few gene trees (~13.3%) supported the inferred species phylogeny (i.e., monophyly of *Arachnothera*; see Table 2). Although this number represents a small proportion of the total UCE loci examined, the emergent phylogenetic signal in multiple species tree analyses supports monophyly of *Arachnothera*. Additionally, the two main clades of *Arachnothera* (Clade A & Clade B; Fig. 1) that were recovered with strong support in all of our analyses and in Moyle et al. (2011) are represented in 45% and 76% of the gene trees, respectively. Further, the topologies from each locus in the Sanger sequencing study of Moyle et al. (2011) were recovered in approximately 8% (Clade A + *Hypogramma*) and 6% (Clade B + Sunbirds) of all gene trees.

MSC species tree reconstruction methods assume the bifurcating estimated gene trees to be true. As such, even gene trees lacking or possessing little phylogenetic signal are informing the species tree analysis (Appendix 3). To assess the impact uninformative loci may have on our results we culled the data matrix at two cutoffs, removing loci with less than 10 and 20 informative sites to produced reduced datasets of 1,712 and 1,291 loci, respectively. We then ran bipartition analyses to calculate the proportion of gene trees that recovered relationships of interest. We found removing the uninformative loci eliminated some noise in the dataset, but given the uniform increase in proportion of recovery for specific clades we can say that noise was not driving our phylogenetic inference. In the dataset that excluded loci with 20 or fewer uninformative loci, the recovery of *Arachnothera* monophyly increased by ~36% and the

proportion of recovery for Clade A and Clade B increased ~33% and ~24%, respectively (Appendix 4).

MSC methods have been designed to model deep coalescence, as such they are robust to varying amounts of incomplete lineage sorting (ILS; Liu et al., 2015). Conversely, HGT is not explicitly modeled in MSC summary methods. Due to the overlapping geographic range among many species within *Arachnothera* and the overlap in range between *Arachnothera* and *Hypogramma*, recent gene flow between species could have resulted in HGT and thus increased gene tree heterogeneity. To account for this, we visually inspected the bipartition table for evidence of gene flow across species. Given the rampant sympatry, if there was HGT we would expect a sister relationship between two distantly related species to be recovered in a high proportion of gene trees. We found no such pairing of individuals to indicate recent gene flow or horizontal gene transfer. Furthermore, both simulation and empirical investigations have shown two MSC methods we used (NJst & ASTRAL) to produce statistically consistent species trees in the presence of low levels of HGT, and one method (ASTRAL) to be particularly robust to high levels of HGT (Davidson et al., 2015). Given that MSC methods are designed to perform well despite deep coalescence, that we did not observe evidence to suggest HGT and that we implemented methods shown to be statistically consistent in the presence of varying levels of deep coalescence and HGT, we feel confident gene tree heterogeneity due to either did not distort our phylogenetic inference.

For taxa in the anomaly zone, using the "democratic method" of interpreting the most frequently recovered gene tree as the true species phylogeny would lead to a highly misleading species phylogeny (Degnan and Rosenberg, 2006). Based on overall gene tree topology, the most common gene tree recovered paraphyly of *Arachnothera*. In fact, many gene tree topologies that

included all individuals recovered patterns of paraphyly in *Arachnothera*, but this information did not clarify relationships at the nodes of interest (i.e. the relationships within *Arachnothera* species and relationships between *Arachnothera* and *Hypogramma* or *Arachnothera* and other sunbirds). Given the Sanger sequencing data from a few loci, finding many discordant gene trees in our dataset did not come as a surprise. Because of deep coalescence and the stochastic nature of substitutions, anomalous gene trees are not rare occurrences. With a complete matrix of 2,107 UCE loci, the number of gene trees that show "unconventional and even bizarre relationships" in comparison to the species phylogeny is unsurprisingly high (Springer and Gatsey, 2016). Although the most common gene tree topology with all individuals (n=6) supported paraphyly, we found relationships within gene trees supporting monophyly of *Arachnothera* to be much greater. The uniform results across concatenated and the MSC methods is confirmation we are not in the anomaly zone and the emergent phylogenetic signal, monophyly of *Arachnothera*, is recovered despite the presence of high gene tree heterogeneity.

Given the level of incongruence among gene trees in the dataset, randomly selecting a few UCE loci would be unlikely to produce the inferred species tree. Similarly, Sanger sequencing of a few known molecular markers would also be unlikely to obtain strong support for the species tree we recovered. Moyle et al. (2011) essentially made two draws (one mitochondrial and one nuclear) from a heterogeneous pool of gene trees, which yielded well supported topologies that likely do not represent the evolutionary history of this group. The number of taxonomic changes based on one or a few genetic markers and permanence of taxonomic nomenclature should give researchers cause for concern. Given the information provided by studies like this one, we feel that systematic biologist should employ a more robust, integrative approach to taxonomic change and strongly consider taxonomic modifications.

*Taxonomy*

Despite similarities in morphology, geographical distribution and life history, the Purple-naped Sunbird (*H. hypogrammicum*) is not embedded within *Arachnothera*, but rather is the sister taxon of the genus. Nevertheless, our findings do not mandate a taxonomic change. *Hypogramma hypogrammicum* can be maintained as a monotypic genus in agreements with traditional taxonomy, or it can remain in an expanded *Arachnothera* following the changes made based on Moyle et al. (2011).

Approximately 85% of Nectariniidae species exhibit sexual dimorphism in plumage coloration and body size, with the males being markedly larger than females (del Hoyo, Elliot, & Christie, 2008). In sexually dichromatic species, males display iridescent plumage patches; in four sexual monomorphic species, both sexes show iridescent patches. Contrary to the morphological condition of most sunbirds, spiderhunters do not show sexual dichromatism in plumage coloration and only slight sexual dimorphism in body size. However, dichromatism is being assessed from a human perspective and there may be appreciable differences in plumage coloration from an avian visual perspective (Eaton & Lanyon, 2003; Burns & Shultz, 2012). In contrast, *Hypogramma* exhibits sexually dichromatic plumage, with the male possessing multiple iridescent patches.

*Arachnothera* is further unified by a tongue morphology that is distinct among sunbirds. Additionally, *Hypogramma* possess a tongue morphology unlike any other species. Given that sunbirds rely on their tongue to probe fruits and flowers while foraging, and the large amount of overlap in the distribution of these two genera, differences in tongue morphology could be the result of resource allocation or niche partitioning. Lastly, *Arachnothera* exhibits different

parental strategies than all other sunbirds, including a unique nest structure and shared parental care by males and females (Cheke and Mann, 2001). This suite of behavioral and physiological differences between *Hypogramma* and *Arachnothera* further substantiate traditional taxonomic delimitation of these particular sunbirds (del Hoyo, Elliot, & Christie, 2008). In a family of birds where genera are distinguished by far fewer traits (e.g. length of primary feathers in *Dicaeum* and *Prionocholus*; see Cheke & Mann, 2001), the difference in multiple life history traits for *Hypogramma* and *Arachnothera* corroborate our molecular findings that support an independent evolutionary history for each of these lineages.

**Conclusion**

Our data demonstrate the utility of reduced-representation genomic datasets to produce well-supported species trees despite substantial gene tree conflict, even among groups that displayed strong discordance among fewer genetic markers. In this study, all concatenation and MSC estimations of the species phylogeny produced a single species tree topology, with *Arachnothera* shown to be a monophyletic genus sister to *Hypogramma*. The consistency and high support across all nodes in each species tree estimation gives lends confidence to our hypothesis of species relationships within *Arachnothera*.

**Funding**

This work was funded in part by a Panorama Grant awarded to LCC. Additional funding was provided by the National Science Foundation (DEB 1146345 to RGM).

**Acknowledgements**

We wish to thank the following people and institutions for providing tissue loans: Steve Cardiff, Louisiana State University Museum of Natural Science and Mark Robbins, University of Kansas Natural History Museum. We are grateful to Brant Faircloth for guidance on UCE laboratory protocols and data processing. We thank the KU Genome Sequencing Core (supported by NIH grant 5P20GM103638 to E.A. Lundquist) and the KU Advanced Computing Facility (partially funded by NSF grant CNS 1337899 to A.T. Peterson) for the use of their facilities.

**Literature Cited**

Andersen, M.J., Shult, H.T., Cibois, A., Thibault, J.-C., Filardi, C.E., Moyle, R.G., 2015. Rapid diversification and secondary sympatry in Australo-Pacific kingfishers (Aves: Alcedinidae: Todiramphus). Royal Society Open Science 2, 140375–140375. doi:10.1098/rsos.140375

Bates, J.M., Hackett, S.J., Goerck, J.M., 1999. High Levels of Mitochondrial DNA Differentiation in Two Lineages of Antbirds (Drymophila and Hypocnemis). The Auk 116, 1093–1106. doi:10.2307/4089688

Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L., Ward, P.S., 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evolutionary Biology 15. doi:10.1186/s12862-015-0552-5

Brennan, I.G., Bauer, A.M., Jackman, T.R., 2016. Mitochondrial introgression via ancient hybridization, and systematics of the Australian endemic pygopodid gecko genus Delma. Molecular Phylogenetics and Evolution 94, 577–590. doi:10.1016/j.ympev.2015.10.005

Burns, K.J., Shultz, A.J., 2012. Widespread cryptic dichromatism and ultraviolet reflectance in the largest radiation of Neotropical songbirds: implications of accounting for avian vision in the study of plumage evolution. The Auk 129, 211–221.

Cheke, R., Mann, C., 2001. Sunbirds: A guide to the Sunbirds, Flowerpeckers, Spiderhunters and Sugarbirds of the World. Yale University Press, New Haven.

Chifman, J., Kubatko, L., 2014. Quartet Inference from SNP Data Under the Coalescent Model. Bioinformatics 30, 3317–3324. doi:10.1093/bioinformatics/btu530

Clements, J. F., T. S. Schulenberg, M. J. Iliff, D. Roberson, T. A. Fredericks, B. L. Sullivan, and C. L. Wood. 2015. The eBird/Clements checklist of birds of the world: v2015. Downloaded from http://www.birds.cornell.edu/clementschecklist/download/

Davidson, R., Vachaspati, P., Mirarab, S., Warrnow, T., 2015. Phylogenomic species tree estimation in the presence of incomplete lineage sorting and horizontal gene transfer. BMC Genomics 16(S10). doi:10.1186/1471-2164-16-S10-S1

Degnan, J.H., Rosenberg, N.A., 2006. Discordance of Species Trees with Their Most Likely Gene Trees. PLoS Genetics 2, e68. doi:10.1371/journal.pgen.0020068

Eaton, M.D., Lanyon, S.M., 2003. The ubiquity of avian ultraviolet plumage reflectance. Proceedings of the Royal Society B: Biological Sciences 270, 1721–1726. doi:10.1098/rspb.2003.2431

Faircloth, B.C., 2015. PHYLUCE is a software package for the analysis of conserved genomic loci (No. biorxiv;027904v2).

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. Systematic Biology 61, 717–726. doi:10.1093/sysbio/sys004

Gill, F & D Donsker (Eds). 2016. IOC World Bird List (v 6.1). doi : 10.14344/IOC.ML.6.1

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. Nature Biotechnology 29, 644–652. doi:10.1038/nbt.1883

del Hoyo, J., Elliot, A., Christie, D. (Eds.), 2008. Penduline-tits to Shrikes, Handbook of the birds of the world. Lynx Ed, Barcelona.

Katoh, K., Standley, D.M., 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. Molecular Biology and Evolution 30, 772–780. doi:10.1093/molbev/mst010

Liu, L., Xi, Z., Wu, S., Davis, C.C., Edwards, S.V., 2015. Estimating phylogenetic trees from genome-scale data: Phylogenies and the multispecies coalecsent model. Annals of the New York Academy of Sciences 1360, 36–53. doi:10.1111/nyas.12747

Liu, L., Yu, L., 2011. Estimating Species Trees from Unrooted Gene Trees. Systematic Biology 60, 661–667. doi:10.1093/sysbio/syr027

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

Liu, L., Yu, L., Kubatko, L., Pearl, D.K., Edwards, S.V., 2009. Coalescent methods for estimating phylogenetic trees. Molecular Phylogenetics and Evolution 53, 320–328. doi:10.1016/j.ympev.2009.05.033

Maddison, W.P., 1997. Gene trees in species trees. Systematic biology 46, 523–536.

Manthey, J.D., Campillo, L.C., Burns, K.J., Moyle, R.G., 2016. Comparison of Target-capture and Restriction-site Associated DNA Sequencing for Phylogenomics: a Test in Cardinalid Tanagers (Aves, Genus: Piranga). Systematic biology syw005.

McGuire, J.A., Linkem, C.W., Koo, M.S., Hutchison, D.W., Lappin, A.K., Orange, D.I., Lemos-Espinal, J., Riddle, B.R., Jaeger, J.R., 2007. Mitochondrial Introgression and Incomplete Lineage Sorting Through Space and Time: Phylogenetics of Crotaphytid Lizards. Evolution 61, 2879–2897. doi:10.1111/j.1558-5646.2007.00239.x

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A., 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research 17, 240–248. doi:10.1101/gr.5681207

Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., Warnow, T., 2014. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics 30, i541–i548. doi:10.1093/bioinformatics/btu462

Moyle, R.G., Taylor, S.S., Oliveros, C.H., Lim, H.C., Haines, C.L., Rahman, M.A., Sheldon, F.H., 2011. Diversification of an endemic Southeast Asian genus: Phylogenetic relationships of the spiderhunters (Nectariniidae: *Arachnothera* ). The Auk 128, 777–788. doi:10.1525/auk.2011.11019

Pease, J.B., Haak, D.C., Hahn, M.W., Moyle, L.C., 2016. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol 14, e1002379.

Philippe, H., Brinkmann, H., Lavrov, D.V., Timothy J Littlewood, D., Manuel, M., Wörheide, G., Baurain, D., 2011. Resolving difficult phylogenetic questions: why more sequences are not enough. PLoS-Biology 9, 402.

Price, M.N., Dehal, P.S., Arkin, A.P., 2009. FastTree: Computing Large Minimum Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution 26, 1641–1650. doi:10.1093/molbev/msp077

Seo, T.-K., 2008. Calculating Bootstrap Probabilities of Phylogeny Using Multilocus Sequence Data. Molecular Biology and Evolution 25, 960–971. doi:10.1093/molbev/msn043

Springer, M.S., Gatesy, J., 2016. The gene tree delusion. Molecular Phylogenetics and Evolution 94, 1–33. doi:10.1016/j.ympev.2015.07.018

Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics btu033.

Swofford, DL. 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Toews, D.P.L., Brelsford, A., 2012. The biogeography of mitochondrial and nuclear discordance in animals: Biogeography of Mito-Nuclear Discordance. Molecular Ecology 21, 3907–3930. doi:10.1111/j.1365-294X.2012.05664.x

Wiens, J.J., Reeder, T.W., 1997. Phylogeny of the Spiny Lizards (Sceloporus) Based on Molecular and Morphological Evidence. Herpetological Monographs 11, 1. doi:10.2307/1467007

**Figures & Tables**

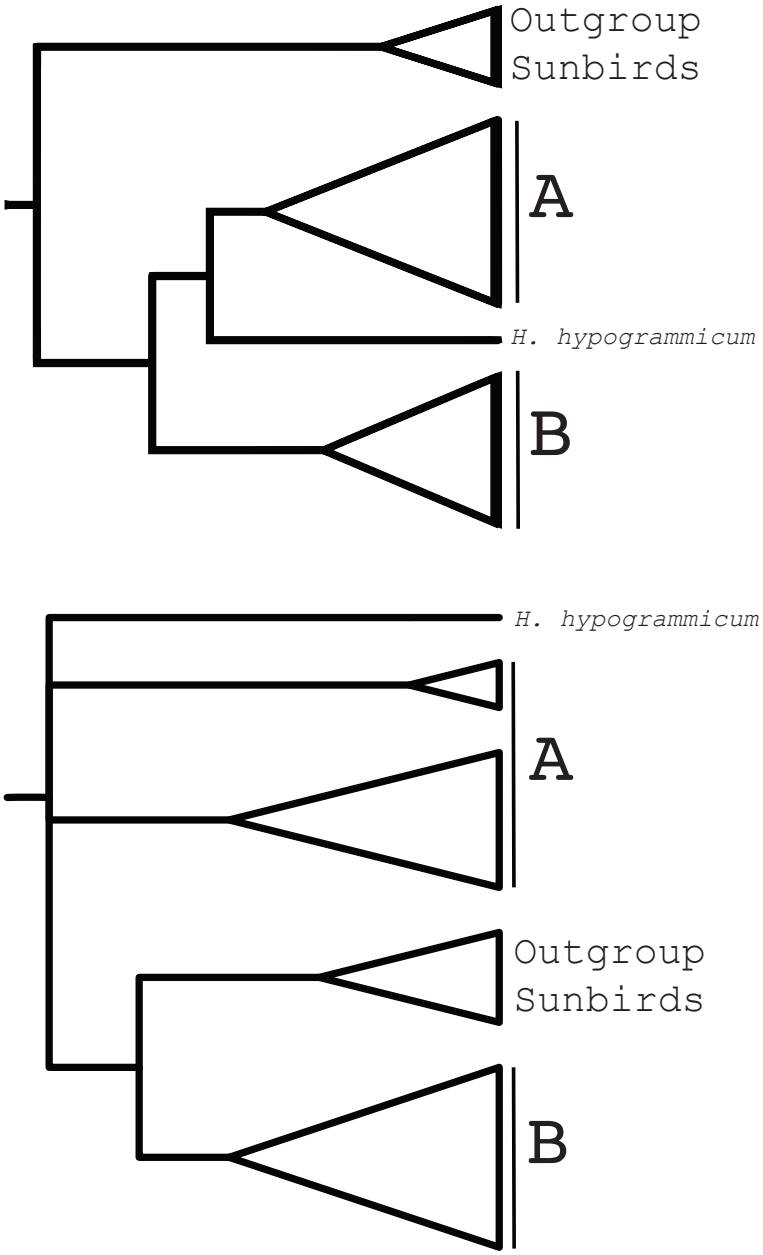



Figure 1: Cartoon adaptation from Moyle et al. (2011)

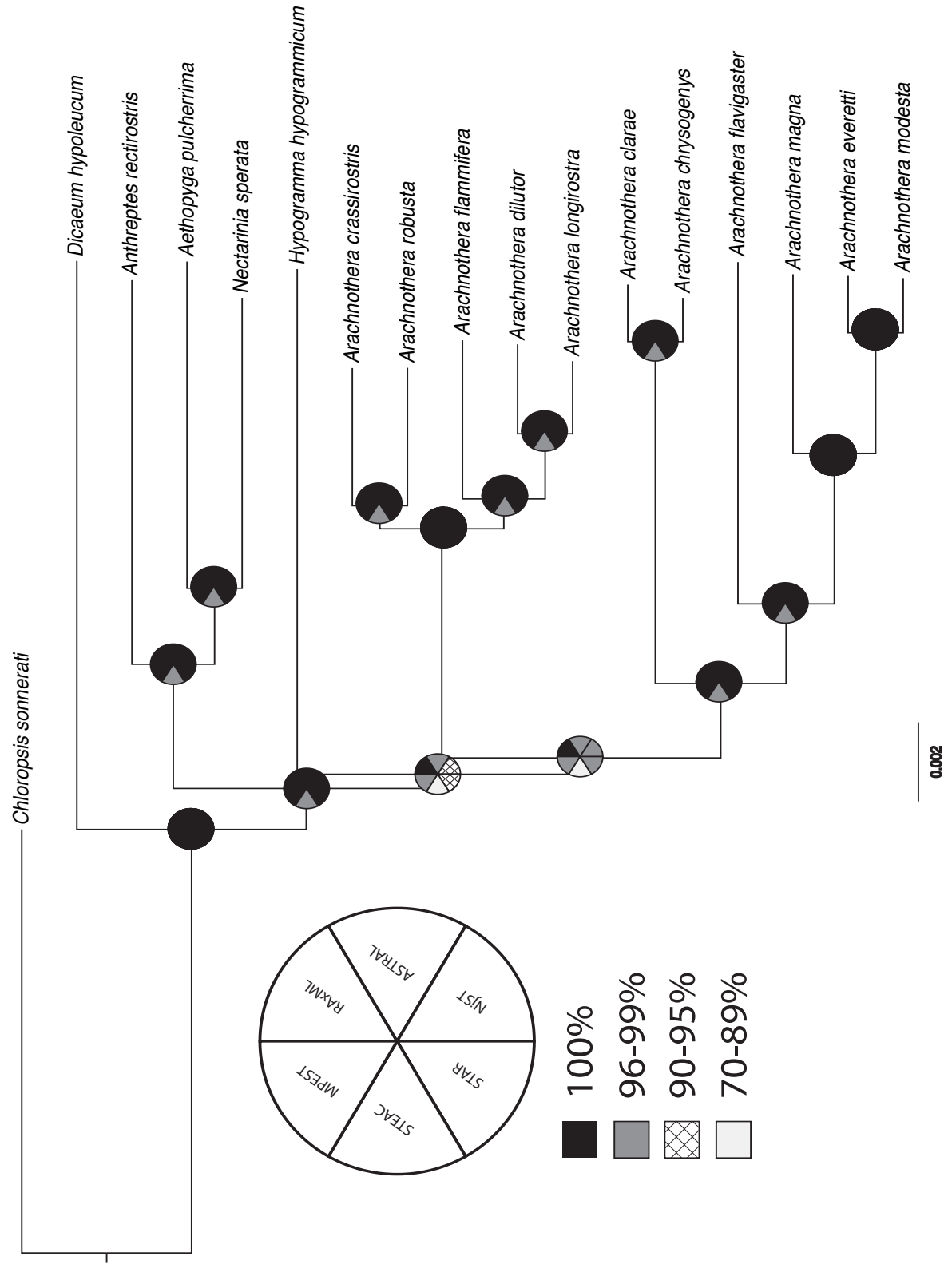



Figure 2: RAxML tree, including nodal support for MSC analyses

Table 1: Locality information and coverage statistics

| Species | Tissue # | Locality | Number of contigs | Mean contig length | Mean Coverage |
|---|---|---|---|---|---|
| *Arachnothera chrysogenys* | KU 17783 | Crooker Range Park, Sabah, Malaysia | 4213 | 842.0 | 35.8 |
| *Arachnothera clarae* | KU 19622 | San Luis, Luzon, Philippines | 3941 | 647.7 | 27.6 |
| *Arachnothera crassirostris* | KU 24436 | Gununy Mulu Natl. Park, Sarawak, Malaysia | 3980 | 851.3 | 34.1 |
| *Arachnothera dilutior* | KU 12621 | Puerto Princesa, Palawan, Philippines | 3794 | 903.0 | 31.3 |
| *Arachnothera everetti* | KU 17761 | Crooker Range Park, Sabah, Malaysia | 4141 | 951.3 | 35.7 |
| *Arachnothera flammifera* | KU 19010 | Mt. Magdiwata, Mindanao, Philippines | 4037 | 874.5 | 35.7 |
| *Arachnothera flavigaster* | KU 17772 | Crooker Range Park, Sabah, Malaysia | 3963 | 930.1 | 36.3 |
| *Arachnothera longirostra* | KU 12343 | Samarakan, Sarawak, Malaysia | 4232 | 564.6 | 45.4 |
| *Arachnothera magna* | KU 10194 | Guanxi, China | 4052 | 989.0 | 37.5 |
| *Arachnothera modesta* | LSU B79500 | Mt. Penrissen, Sarawak, Malaysia | 4350 | 866.9 | 46.6 |
| *Arachnothera robusta* | LSU B51150 | Tawau Hills Park, Sabah, Malaysia | 4160 | 872.9 | 37.8 |
| *Hypogramma hypogrammicum* | KU 17785 | Crooker Range Park, Sabah, Malaysia | 4042 | 954.2 | 32.8 |
| *Nectarinia sperata* | KU 20350 | Baler, Luzon, Philippines | 4187 | 653.3 | 41.5 |
| *Aethopyga pulcherrima* | KU 28306 | Mt. Hilong-hilong, Mindanao, Philippines | 4243 | 882.8 | 35.9 |
| *Anthreptes rectirostris* | KU 29161 | Luki Biosphere, Dem. Republic of Congo | 4063 | 964.2 | 28.8 |
| *Dicaeum hypoleucum* | KU 20176 | San Luis, Luzon, Philippines | 3951 | 971.5 | 46.0 |
| *Chloropsis soneratti* | KU 24451 | Gununy Mulu Natl. Park, Sarawak, Malaysia | 4342 | 758.0 | 54.1 |
| | | | | | |
| | | Overall Means: | 4099.5 | 851.6 | 37.8 |

Table 2: Unique relationships across gene trees

| | Proportion recovery across all gene trees (n=2,107) | | | | |
|---|---|---|---|---|---|
| **Program (Test)** | Clade A | Clade B | Arachnothera monophyly | Clade A + Hypogramma | Clade B + Sunbirds |
| PAUP (bipartitian table) | 0.45 | 0.7561 | 0.1334 | 0.0783 | 0.0641 |
| FastTree (CompareToTree) | 0.45 | 0.756 | 0.133 | 0.078 | 0.064 |

Figure & Table Captions

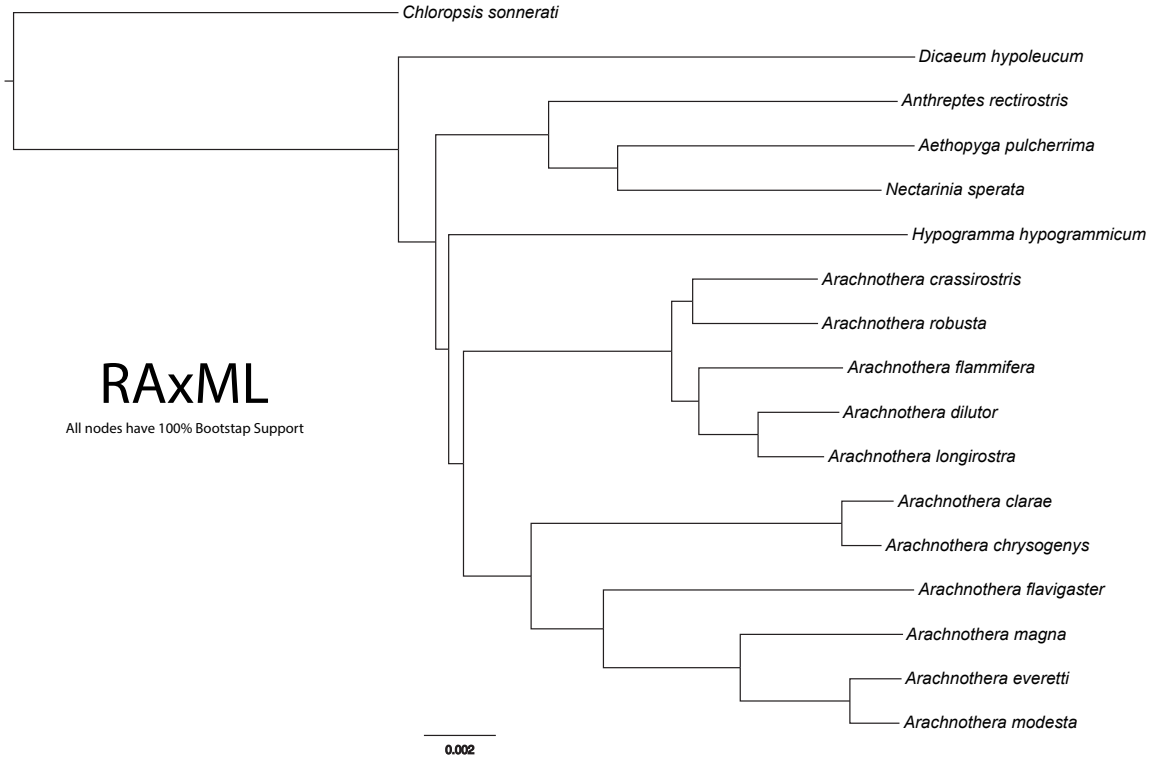
**Figure 1:** The above trees are cartoon adaptations of the trees produced by Moyle et al. 2011. The top tree depicts the topology for the mitochondrial marker and the concatenated ML tree. The bottom tree depicts the topology produced from the nuclear marker**. Clade A includes:** *Arachnothera clarae, A. chrysogenys, A. flavigaster, A. magna, A. everetti*, and *A. modesta*. **Clade B includes:** *A. robusta, A. crassirostris, A. flammiferma, A. longirostra,* and *A. dilutor.*

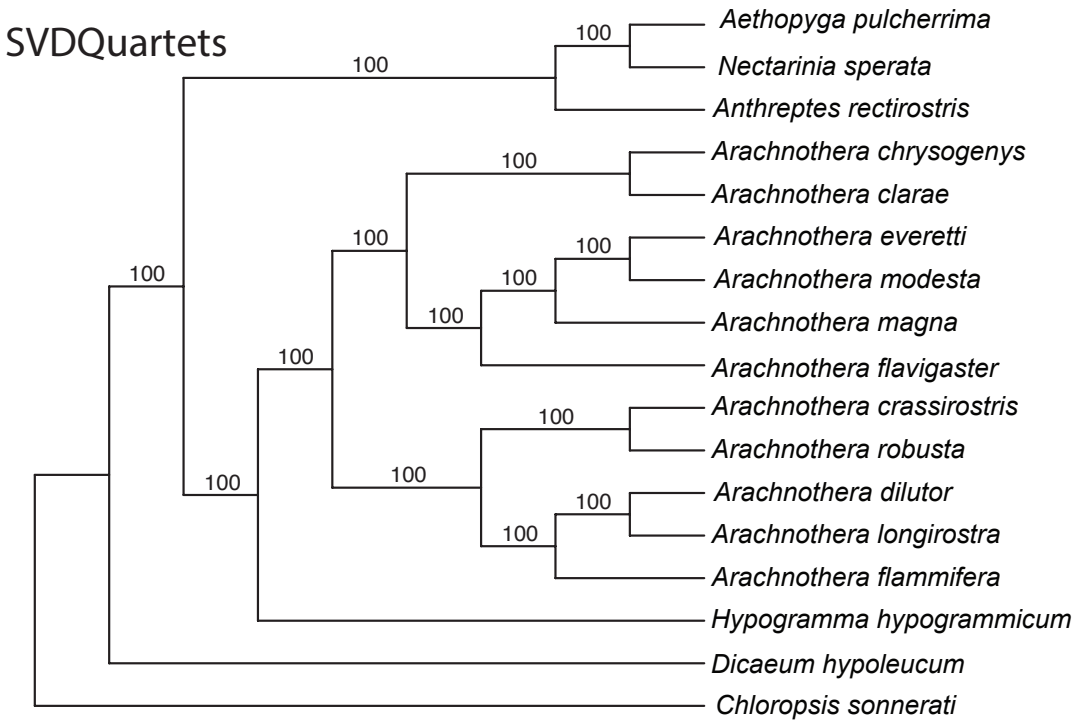
**Figure 2:** Concatenated species tree (produced in RAxML; GTR+GAMMA substitution model) for *Arachnothera*. Pie diagrams at each node correspond to the arrangement of the large pie diagram depicting multiple species tree estimation methods. Colors and patterns represent support for each method, values correspond to the legend under the large pie diagram.

**Table 1:** Samples used in this study, the associated tissue number including institution (KU: Kansas University, LSU: Louisiana State University), locality information, and UCE coverage statistics.

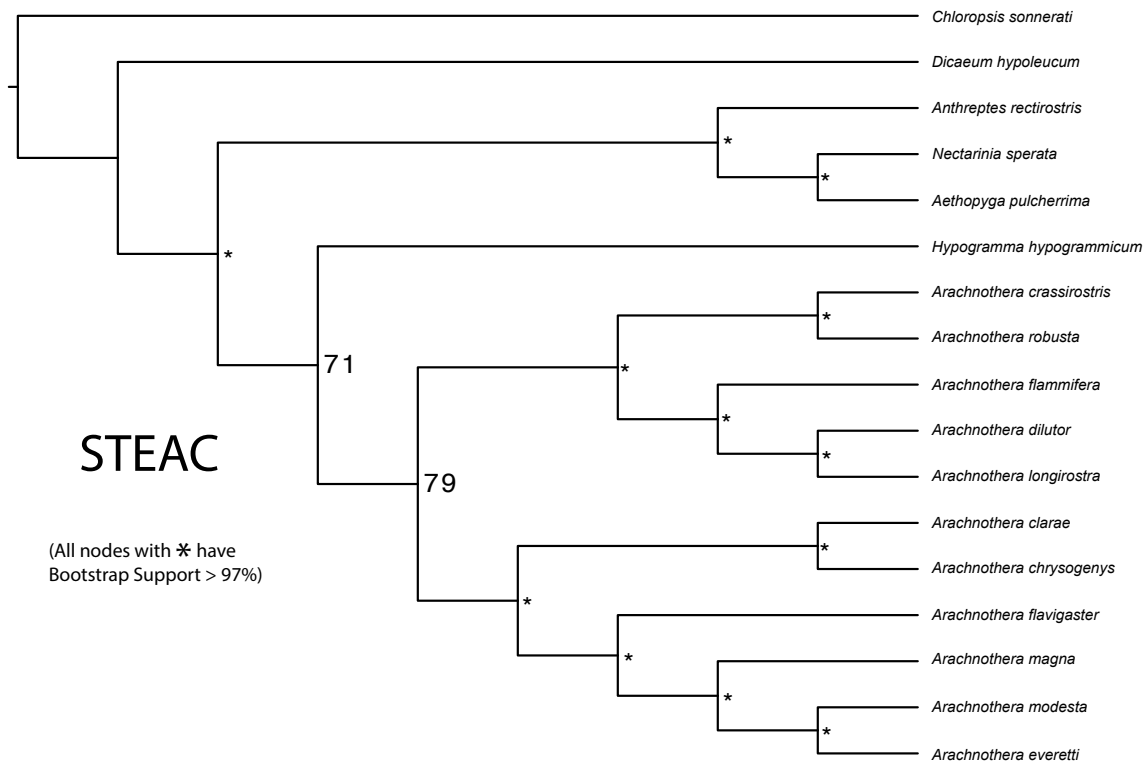
**Table 2:** Proportion of gene trees recovering specific topologies. Clade A and Clade B follow the naming scheme in Fig. 1. The columns comparing "Clade A + *Hypogramma*" and "Clade B + Sunbirds" represent the paraphyletic relationships recovered by Moyle et al. (2011).

## Appendix 1: SVDQuartet & RAxML Trees

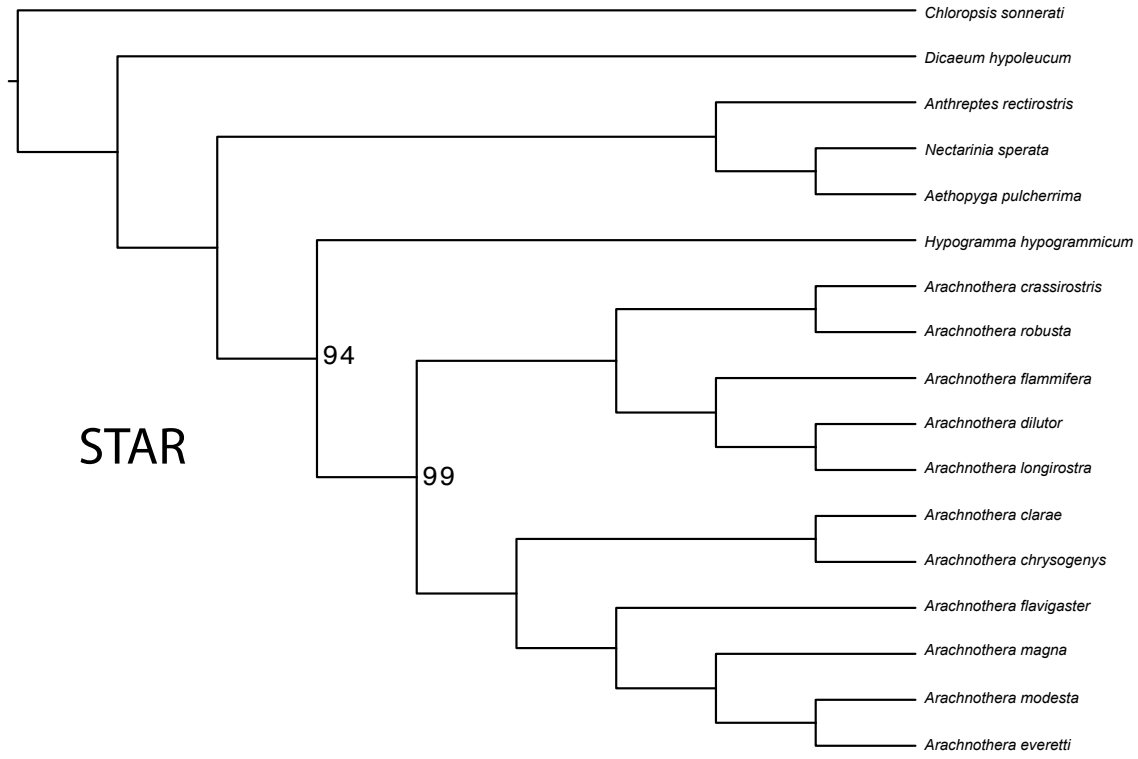


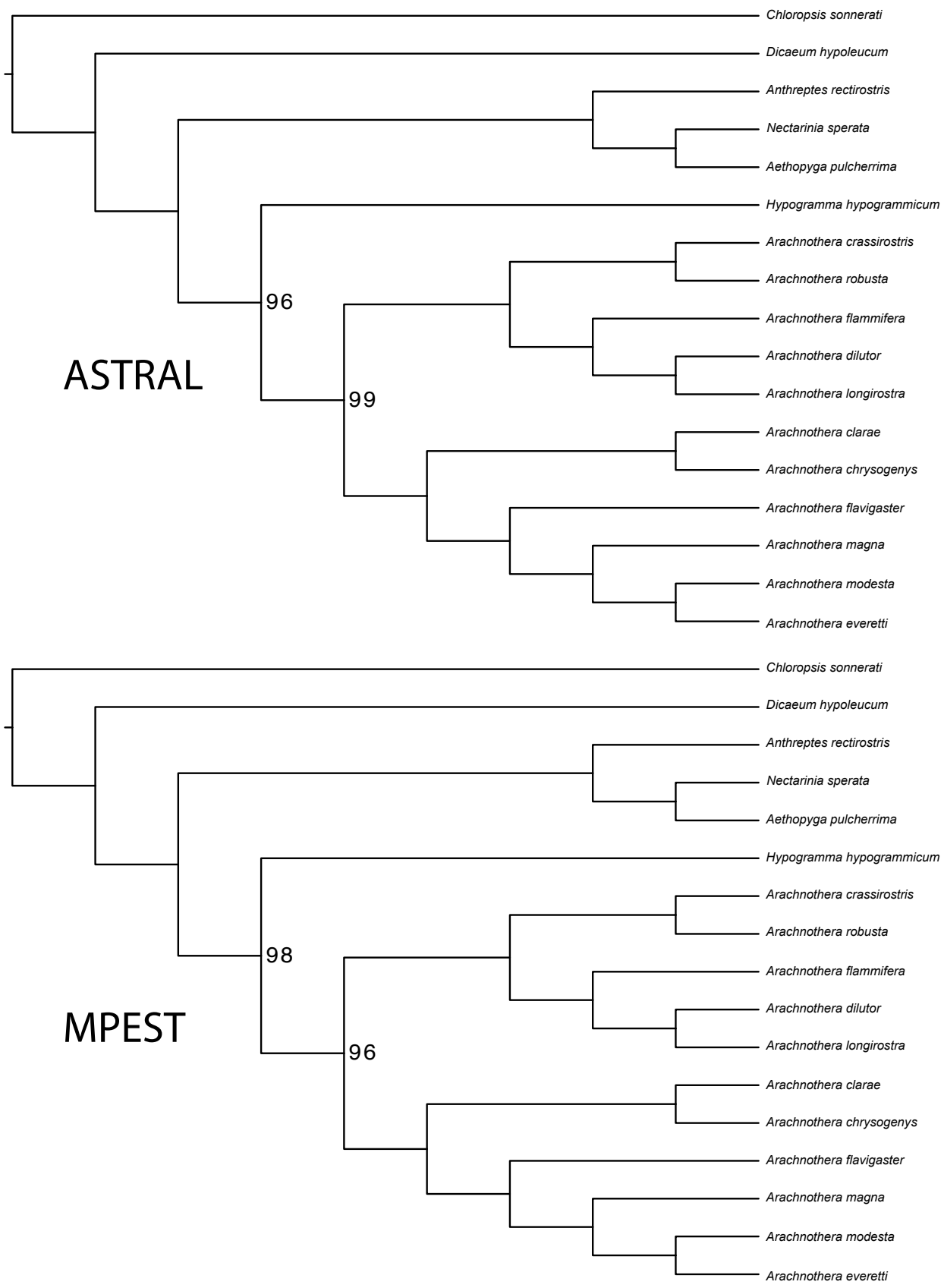



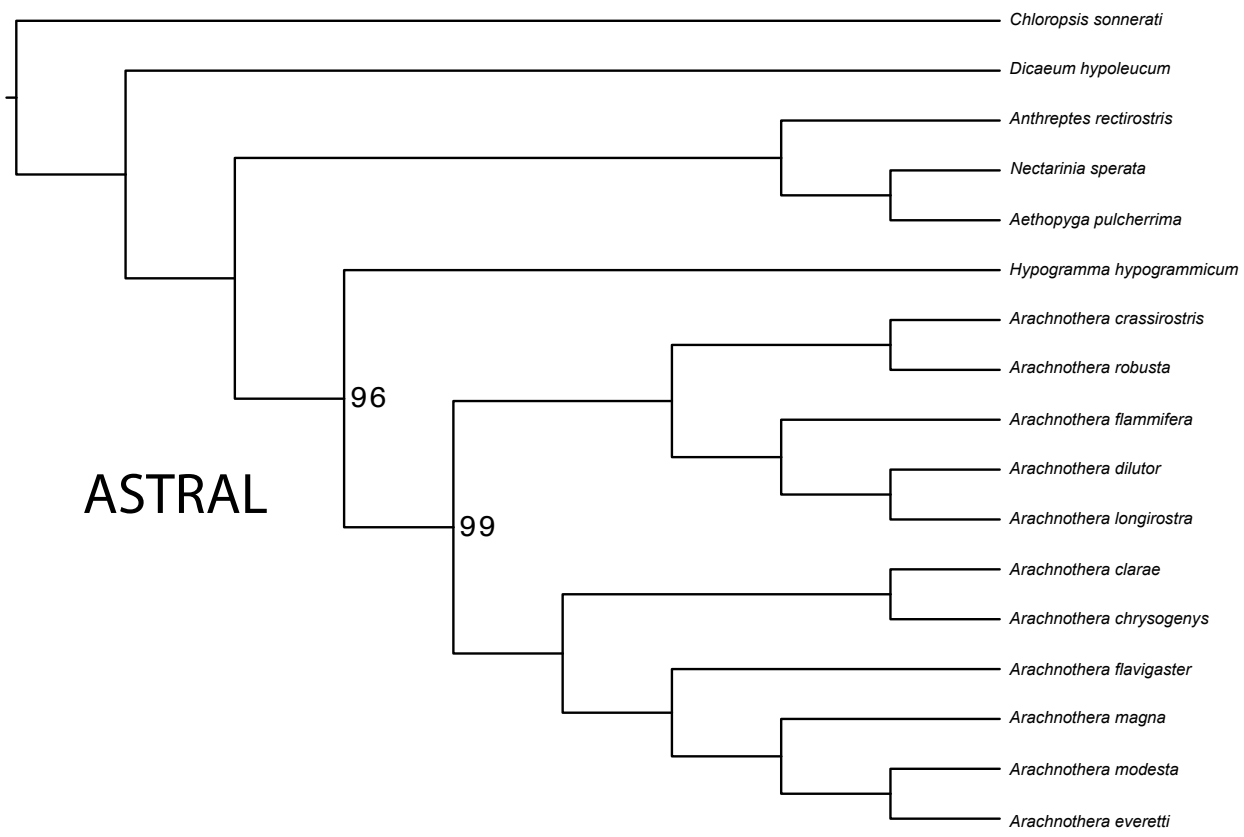
**Appendix 2: MSC Trees**



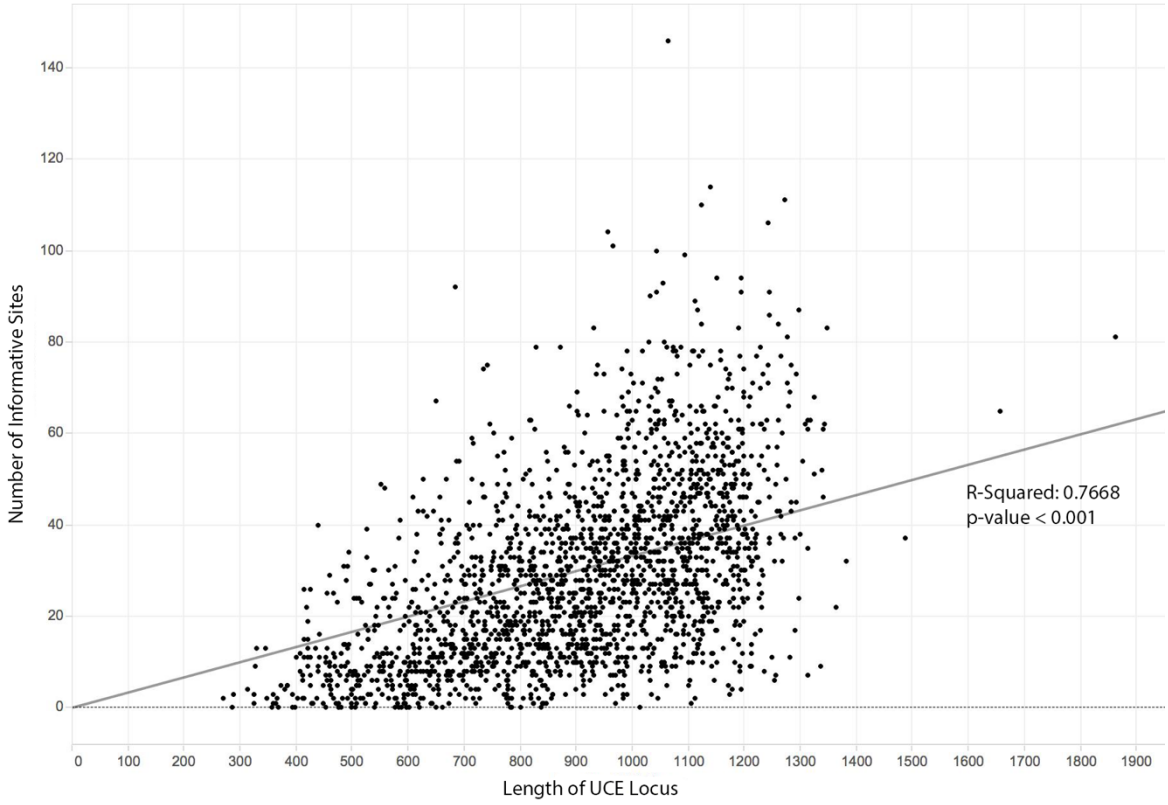

For all subsequent trees, nodes without a number indicate 100% BS

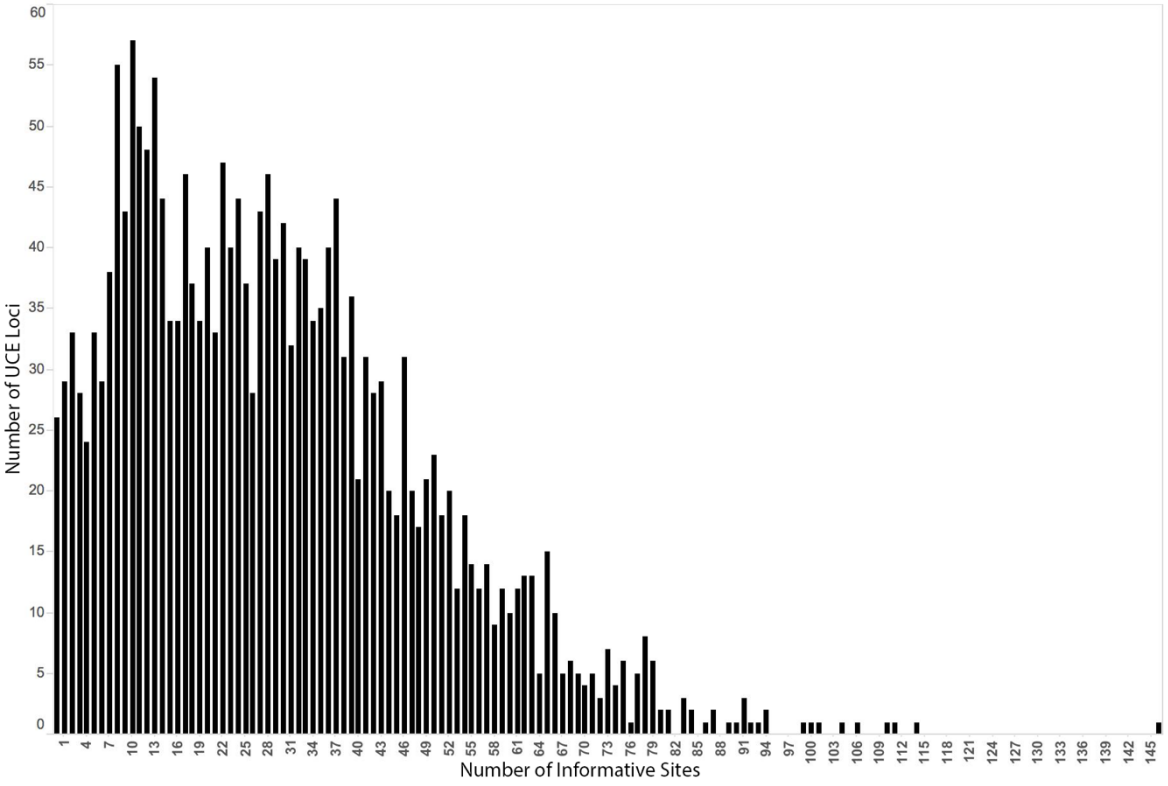


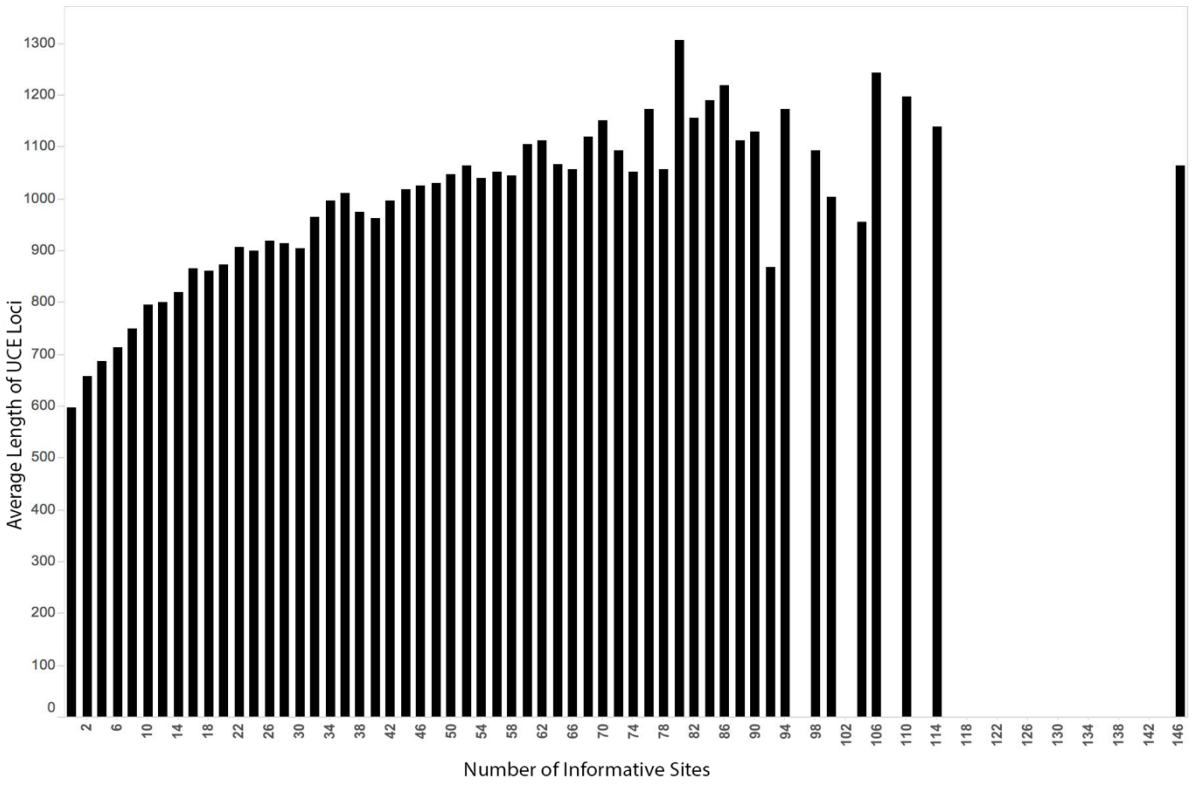

Chloropsis sonnerati
Dicaeum hypoleucum
Anthreptes rectirostris
Nectarinia sperata
Aethopyga pulcherrima
Hypogramma hypogrammicum
Arachnothera crassirostris
Arachnothera robusta
Arachnothera flammifera
Arachnothera dilutor
Arachnothera longirostra
Arachnothera clarae
Arachnothera chrysogenys
Arachnothera flavigaster
Arachnothera magna
Arachnothera modesta
Arachnothera everetti
96
99
ASTRAL
Chloropsis sonnerati
Dicaeum hypoleucum
Anthreptes rectirostris
Nectarinia sperata
Aethopyga pulcherrima
Hypogramma hypogrammicum
Arachnothera crassirostris
Arachnothera robusta
Arachnothera flammifera
Arachnothera dilutor
Arachnothera longirostra
Arachnothera clarae
Arachnothera chrysogenys
Arachnothera flavigaster
Arachnothera magna
Arachnothera modesta
Arachnothera everetti
98
96
MPEST


ASTRAL
96
99
Chloropsis sonnerati
Dicaeum hypoleucum
Anthreptes rectirostris
Nectarinia sperata
Aethopyga pulcherrima
Hypogramma hypogrammicum
Arachnothera crassirostris
Arachnothera robusta
Arachnothera flammifera
Arachnothera dilutor
Arachnothera longirostra
Arachnothera clarae
Arachnothera chrysogenys
Arachnothera flavigaster
Arachnothera magna
Arachnothera modesta
Arachnothera everetti

## Appendix 3: Comparing UCE locus length and Informative Sites




Average Length of UCE Loci
0
100
200
300
400
500
600
700
800
900
1000
1100
1200
1300
2
6
10
14
18
22
26
30
34
38
42
46
50
54
58
62
66
70
74
78
82
86
90
94
98
102
106
110
114
118
122
126
130
134
138
142
146
Number of Informative Sites

**Appendix 4: Unique relationship recovered from a reduced gene tree dataset**

Recognizing that uninformative loci (i.e. those loci without many informative sites, if any at all) may bias the proportion of gene trees recovering unique topologies, we investigated relationships among reduced datasets (see Manthey et al., 2016). The full data matrix (2,107 gene trees) was culled according to the number of phylogenetically informative sites (PIS) per locus. We constructed two reduced datasets by removing UCE loci that did not meet the following cutoffs: PIS ≤ 10 and PIS ≤ 20, the number of gene trees left after pruning the data set were 1712 and 1291, respectively. We again constructed a bipartition table in PAUP and conducted comparisons FastTree, the results are as follows:

| | Proportion recovery with PIS ≤ 10 removed (n = 1712) | | | | |
|---|---|---|---|---|---|
| **Program (Test)** | Clade A | Clade B | Arachnothera monophyly | Clade A + Hypogramma | Clade B + Sunbirds |
| PAUP (bipartitian table) | 0.531 | 0.862 | 0.156 | 0.094 | 0.077 |
| FastTree (CompareToTree) | 0.531 | 0.869 | 0.156 | 0.093 | 0.077 |

| | Proportion recovery with PIS ≤ 20 removed (n = 1291) | | | | |
|---|---|---|---|---|---|
| **Program (Test)** | Clade A | Clade B | Arachnothera monophyly | Clade A + Hypogramma | Clade B + Sunbirds |
| PAUP (bipartitian table) | 0.596 | 0.934 | 0.181 | 0.109 | 0.0868 |
| FastTree (CompareToTree) | 0.596 | 0.936 | 0.181 | 0.109 | 0.087 |

(Special thanks to Bailey Carges for help with formatting figures)