# ABSTRACT


| | |
|---|---|
| Title of Dissertation: | UNRAVELING THE EVOLUTIONARY HISTORY OF NOCTURNALITY IN BIRDS, WITH A FOCUS ON STRISORES |
| | Noor Daniella White, Doctor of Philosophy, 2017 |
| Dissertation directed by: | Michael J. Braun, PhD, Smithsonian Institution<br>Karen L. Carleton, PhD, Department of Biology |



Tracing the processes of adaptation is a fundamental practice in the study of evolutionary biology. By combining multiple lines of evidence, I can elucidate the processes of diversification, speciation, and ultimately, evolution. For my doctoral dissertation, I studied the evolutionary history of a superorder of birds (*Strisores*) that have undergone a dramatic life history transition, the shift from a day-living (diurnal) to a night-living (nocturnal) lifestyle. Previous study found that the diurnal *Apodiformes* (swifts and hummingbirds) are nested deep within the clade of nocturnal or crepuscular *Caprimulgiformes* (nightbirds). However, resolution of the other major lineages eluded previous efforts, precluding analysis of the evolution of nocturnality in this group. To resolve the phylogeny of *Strisores*, I utilized a novel class of genome-scale markers, ultraconserved elements (UCEs). UCEs are operationally defined regions of extreme conservation between two or more genomes. I collected

and sequenced ~4,000 UCEs from each of 191 species of birds representing every major extant lineage, plus two crocodilian outgroups—a greater number of elements than had ever been collected or studied before. With these data, I have resolved the phylogeny of the largest and oldest (Caprimulgidae and Nyctibiidae, respectively) lineages of nightbirds, as well as the superorder *Strisores*, and exemplified some rigorous methods for the phylogenomic analysis of UCEs. With a phylogeny representing the evolutionary history of *Strisores* I then ask when, and where, potential adaptations to nocturnality occurred. To this end, I have developed a molecular tool to efficiently enrich 47 genes comprising the phototransduction cascade, a network of genes that converts the absorption of a photon by an opsin into a neural signal. I demonstrated that this tool is effective in 33 bird species chosen to cover extant avian diversity. The data captured using this array will facilitate the identification of potential molecular adaptations to nocturnality, enable the improvement of models predicting opsin sensitivity from sequence data, and allow strong inference about the perception of color across birds and other vertebrates.

# UNRAVELING THE EVOLUTIONARY HISTORY OF NOCTURNALITY IN BIRDS, WITH A FOCUS ON STRISORES

by

Noor Daniella White

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2017

Advisory Committee:
Professor Michael J. Braun, Co-Chair
Professor Karen L. Carleton, Co-Chair
Professor Charles F. Delwiche
Professor Charles Mitter
Professor David L. Swofford
Professor Gerald S. Wilkinson

# Dedication

To my parents,

Georgette and Glenn White,

who raised me to believe that I could do anything.

This is all their fault.

# Foreword

Two chapters herein have been previous published. The first chapter, "A Multi-Gene Estimate of Higher-Level Phylogenetic Relationships Among Nightjars (Aves: Caprimulgidae)," was published as "White N. D., Barrowclough G. F., Groth J. G. and Braun M. J. (2016). A multi-gene estimate of higher-level phylogenetic relationships among nightjars (Aves: Caprimulgidae). Ornitología Neotropical, 27: 223-236." White conducted all analyses described in that paper with the exception of generating the heterozygosity and base composition plots (Figures 1 and 2), which were made by Barrowclough. White, Barrowclough and Braun wrote the text.

The second chapter, "Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae)" was published as "White, N. D., Mitter, C. and Braun, M. J. (2017), Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae). Journal of Avian Biology, 48: 872–880." White generated all sequences described in this work and conducted all analyses described. White, Mitter and Braun wrote the text.

UNIVERSITY OF MARYLAND
BIOLOGICAL SCIENCES GRADUATE PROGRAM

2101 Bioscience Research Building
College Park, Maryland 20742-4415
301.405.6905/6991 TEL, 301.314.9921 FAX

Alexander Chen
Dean of the Graduate School
The Graduate School
2123 Lee Building
University of Maryland
College Park, MD 20742

Dear Dean Chen,

This letter is written to signify that the dissertation committee, committee chair, and the graduate director have all approved the use of previously published co-authored work in the final dissertation of Noor D. White, Biological Sciences, UID 108906495.

Citations for the published work(s):

1. White ND, Mitter CM and Braun MJ (2017). Ultraconserved Elements Resolve the Phylogeny of Potoos (Aves: Nyctibiidae). Journal of Avian Biology, 48: 872–880.

2. White ND, Barrowclough GF, Groth JG and Braun MJ (2016). A Multi-Gene Estimate of Higher-Level Phylogenetic Relationships among Nightjars (Aves: Caprimulgidae). Ornitología Neotropical, 27: 223-236.

In accordance with the Graduate School's policy the dissertation committee has determined that they made substantial contributions to the included work.

Per Graduate School policy the dissertation foreword will identify the scope and nature of the student's contributions to the jointly authored work included in the dissertation and a copy of this letter will be submitted with the dissertation.

Sincerely,

Karen L. Carleton, Dissertation Committee Chair,
Professor, Biology

Michael J Braun, Co-Advisor,
Research Scientist, Smithsonian Institution National Museum of Natural History

Dr. Michelle Brooks,
Associate Director, Biological Sciences Graduate Program

# Table of Contents

# List of Tables

# List of Figures

# Introduction

Everything in evolution makes better sense in the light of phylogenetics (to paraphrase Dobzhansky, 1973; Sytsma & Pires 2001). Indeed, the process of depicting the evolutionary history of life on a branching diagram enables researchers to trace the patterns and processes of adaptation, and elucidate mechanisms of diversification, speciation, and ultimately evolution itself. With the advent of next-generation sequencing (NGS) technologies, molecular techniques to efficiently capture thousands of loci across hundreds of samples have been developed to take advantage of this sequencing capacity. This has enabled a suite of new genome-scale phylogenetic markers, such as genotyping by sequencing (GBS), restriction associated DNA sequencing (RADseq), anchored phylogenetics and ultraconserved elements, that are now widely used (Elshire et al. 2011; Faircloth et al. 2012; Lemmon et al. 2012; Miller et al. 2007). Forward progress in this area has been so great that the rate at which one can generate a genome-scale molecular data matrix is outpacing the development of methods with which to analyze it. The field is particularly lacking in appropriate models of sequence evolution for such large datasets, or means by which to sensibly partition thousands of loci (Bull et al. 1993; de Queiroz, 1993; Lanfear et al. 2012). Applying currently available phylogenetic tools to large NGS matrices can inflate statistical support, resulting in a tree with apparently high support at every node, and placing the onus on the researcher to accurately interpret that support and topology (Kumar et al. 2012).

One major focus of my doctoral work has been to investigate the use of a novel class of genome-scale markers, ultraconserved elements (UCEs), for phylogenetic inference. UCEs are operationally-defined regions of extreme conservation between two or more genomes (Bejerano et al. 2004b). UCE function is an active area of research in molecular genetics; they are believed to play a variety of roles. Empirical studies have so far identified UCEs with functions in development and regulation of transcription, as well as UCEs that have no effect when deleted from the mouse genome (Calle-Mustienes et al. 2005; Navratilova et al. 2009; Nóbrega et al. 2004; Pennacchio et al. 2006; Woolfe et al. 2005). UCEs, or regions of similar conservation, have been identified in a wide variety of taxonomic groups, including vertebrates, insects, yeasts, and plants (Faircloth et al. 2015; Lockton & Gaut, 2005; McCormack et al. 2012; Siepel et al. 2005; Stephen et al. 2008)

The utility of UCEs for phylogenetics is not in the sequence of the locus itself, but in the ability to use the extreme sequence conservation of these elements as a reliable signal of orthology across variable depths of time, as well as an efficient way to enrich genomic libraries for UCE sequences in the lab. The informative data for phylogenetic inference is derived from regions flanking the conserved element, which Faircloth et al. (2012) demonstrated shows increasing sequence variation across taxa with increasing distance from the UCE core (the core itself typically shows ~2% variation across the relevant taxonomic scale). The use of UCEs can thus be tailored to the targeted phylogenetic depth by altering the length of the flanking sequence recovered, and their utility has been demonstrated at multiple timescales (Crawford et al. 2012; Faircloth et al. 2013a; Smith et al. 2013). Perhaps due to their flexibility, the

most appropriate methods by which to analyze UCE data can vary from dataset to dataset. For example, the phylogeny presented in Chapter 2 was derived from a dataset which showed little to no evidence of topological conflict, whereas the dataset presented in Chapter 3 showed substantial hidden conflict. This variation means that analysis of UCE data must be tailored to each individual dataset, a phenomenon that will be made easier when more detailed analyses of UCE data (such as in Chapter 3) are published and available as example pipelines.

The UCE probe set I use targets the phyletic diversity of extant birds, and was identified by searching an alignment of the chicken (*Gallus gallus*), zebra finch (*Taeniopyga guttata*), and Carolina anole (*Anolis carolinensis*) genomes for regions ≥60 bp in length that share 98-100% identity (Faircloth et al. 2012). The resulting 5,060 loci are present as a single copy in each genome, and represent putatively orthologous core regions. The goal of my investigation of UCEs is to improve the accurate use of UCE data in phylogenomic inference. I hope through these efforts to also shed light on their function, and best methods for modeling their evolution in phylogenetic analyses (see Functional Annotation efforts in Ch. 3).

To this end, I have collected and sequenced ~4,000 UCEs from each of 193 species of birds representing every major extant lineage, plus two crocodilian outgroups. This dataset is unique to the study of UCEs thus far as it represents a broader sampling of diversity, and greater number of elements than have ever been collected or studied before. The avian tree of life is a pertinent system to reconstruct using UCEs for two main reasons; 1) resolution of the rapid radiation of avian diversity has eluded previous efforts made even with 48 whole genome sequences

(Jarvis et al. 2014), and 2) because birds are so widely studied, a robust phylogeny of birds will inform questions related to many major themes in evolutionary biology, including sexual selection, biogeography and speciation.

The other focus of my doctoral work was resolution of the phylogeny of the *Strisores*, a superorder of birds containing both nocturnal and diurnal members. The nightbirds (previously *Caprimulgiformes*) are an enigmatic, but much-storied group of five, largely nocturnal, families. Due to the exceptionally wide mouths of the nightjars and nighthawks (Caprimulgidae), they are commonly referred to as "goatsuckers," as legend has it that their wide mouths allowed them to suckle goat's milk at night. Nightbirds are characterized by their mottled or cryptic plumage, which aids camouflage in the group but confounds taxonomy. There are a number of unusual traits present in the nightbirds, particularly striking among these is the presence of echolocation in the oilbird (Steatornithidae), and a distinctive roosting posture taken on by the potoos (Nyctibiidae) and frogmouths (Podargidae), in which they extend their beaks and heads upwards when sitting on the end of a vertical snag so that their body takes on the shape of their roost (reviewed in Han et al. 2010 and White et al. 2017). Vocalizations in the group are varied but are largely comprised of hoarse, buzzy tones, rather than melodic song. Two lineages, the nightjars and potoos, are particularly known for their loud, broadcasting calls. Most nightbirds are insectivorous, hunting their prey on the wing, but exceptions include the oilbird, which feeds primarily on the fruits of palm and laurel trees, and frogmouths, which have been described eating anything from worms to carrion (Cleere 1998; 2010; Martin et al. 2004; Snow 1961; 1962). Most nightbirds have rictal bristles around

their beaks which are thought to aid in the localization of insects when caught in flight (Cleere 1998; 2010; Han et al. 2010). Potoos do not have true rictal bristles, but do have bristles emanating from the corners of their mouths, which may serve the same purpose (Cohn-Haft 1999; 2016). Nightbirds display a diversity of nesting strategies, from the caprimulgids which roost directly on the ground, where they blend in with foliage and build their nests, to the oilbird, a cave-dweller, which builds its nests on steep cliff faces. Owlet-nightjars (Aegothelidae) nest and roost in holes in trees, while frogmouths nest on horizontal tree limbs and potoos lay their eggs on limbs or in the open ends of vertical dead stumps.

While nightbirds have a very distinctive natural history, it is their association with a group of diurnal birds, the *Apodiformes* (swifts and hummingbirds), which prompted me to study them. In 2001, sequence data from nuclear and mitochondrial DNA found that the diurnal *Apodiformes* are sister to the owlet-nightjars (Braun & Huddleston, 2001). This topology has since been confirmed by multiple independent studies using both molecular and morphological data, and the placement of the *Apodiformes* within the *Caprimulgiformes* is now referred to as the superorder *Strisores* (Barrowclough et al. 2006; Braun & Huddleston, 2009; Ericson et al. 2006; Hackett et al. 2008; Mayr 2002; 2010; Prum et al. 2015; Reddy et al. 2017). Despite the studies listed here, placement of the other major *Strisores* lineages within the phylogeny has not been resolved. Lack of resolution of this phylogeny precludes analysis of the evolutionary history of any of the adaptations present in this group. Here, I present resolved phylogenies of the families Caprimulgidae (Chapter 1) and

Nyctibiidae (Chapter 2), the largest and oldest (respectively) lineages of nightbirds, as well as the superorder *Strisores* (Chapter 3).

The last chapter of my dissertation describes preliminary testing of a molecular tool I developed to enable further inference of the evolutionary history of nocturnality across all birds (Chapter 4). It is a target capture probe array for efficient enrichment of 47 genes comprising (or related to) the phototransduction cascade (PTC). The PTC is a network of genes that converts the absorption of a photon by an opsin into a neural signal. I demonstrate that sequence capture via in-solution hybridization with my probe array is effective, recovering full coding sequence for the rhodopsin gene in 27 of the 33 taxa tested. The data captured using this array will facilitate the identification of many potential molecular adaptations, enable the improvement of models predicting opsin sensitivity from sequence data, and allow strong inference about the perception of color in birds and other vertebrates (Ödeen et al. 2008). Study of the PTC in the context of nocturnality has been rare, but holds great potential for uncovering mechanisms of adaptation (Borges et al. 2015; Wu et al. 2016). The PTC is highly conserved across vertebrates, and thus findings within birds are translatable across the tree of life. Lastly, the PTC has been co-opted to serve other non-visual functions, and this tool could impact non-visual research (Kelley & Davies, 2016).

# Chapter 1: A Multi-Gene Estimate of Higher-Level Phylogenetic Relationships Among Nightjars (Aves: Caprimulgidae)

This chapter was published as: White N. D., Barrowclough G. F., Groth J. G. and Braun M. J. (2016). A multi-gene estimate of higher-level phylogenetic relationships among nightjars (Aves: Caprimulgidae). Ornitología Neotropical, 27: 223-236.

**Abstract**

The higher-level phylogenetic relationships of the nightjars and nighthawks (Caprimulgidae) have been challenging for traditional systematics due to their cryptic plumage and conservative morphology. I explored these relationships by combining two previously published molecular datasets with new data to generate a complete matrix (7,104 bp) of evolutionarily disparate sequence elements from four genes for 36 taxa. I analyzed each of the genes separately for base composition heterogeneity and heterozygosity. I analyzed the concatenated matrix in a likelihood framework using seven different partitioning schemes. As the number of subsets in a given partitioning scheme increased, tree length and likelihood score also increased; however, the branching topology was little affected by increasingly complex partitioning schemes. My best maximum likelihood tree has increased bootstrap support at 13 of 30 ingroup nodes compared with previous analyses, a result likely due to doubling the length of the sequence data. Coalescent-based species tree inference produced a tree congruent with all strongly supported nodes in the

maximum likelihood tree. This topology agrees with previous molecular studies in identifying three small, early branching Old World genera (*Eurostopodus*, *Lyncornis*, and *Gactornis*) and four more speciose terminal clades, representing the New World nighthawks (genus *Chordeiles*) and three nightjar radiations centered in South America, Central America and the Old World, respectively. Increased node support across the tree reinforces a historical scenario with origins in the region surrounding the Indian Ocean, followed by diversification in the New World and subsequent recolonization and radiation in the Old World. Future work on this group should incorporate additional members of the genera *Lyncornis* and *Eurostopodus*, to determine which is the basal lineage of Caprimulgidae.

**Introduction**

The field of molecular phylogenetics has changed dramatically in the past 15 years. Whereas datasets of less than a kilobase from a single mitochondrial gene were frequently published in the 1990's, the size and complexity of datasets have advanced rapidly to include multiple nuclear genes, whole mitochondrial genomes and even entire nuclear genomes (e.g., Jarvis et al. 2014). Today it is relatively straightforward to amass datasets consisting of hundreds to thousands of nuclear markers for dozens of taxa, due to the advent of high-throughput sequencing platforms and the development of efficient genome reduction techniques (e.g., McCormack et al. 2012, Faircloth et al. 2012, Lemmon et al. 2012). However, drawing conclusions from analyses of these datasets requires caution, as they can produce trees with high statistical support that conflict with independent analyses bearing equally high

support (e.g., note conflict between the Bayesian trees in Jarvis et al. 2014 and Prum et al. 2015, see also Hahn & Nakhleh 2015). These examples illustrate that, while analytical methods have been advancing rapidly, the growth of datasets has outpaced the development of software with which to analyze them (see discussion in Kumar et al. 2012). Thus, it is important to examine the complexities of phylogenetic inference on datasets of moderate size where more comprehensive analyses can be undertaken, in order to both test emerging analytical methods and to provide topological comparisons for genome-scale work. Here I explore two important analytical issues for which relatively new software has been developed: 1) data partitioning (Bull et al. 1993, de Queiroz 1993) whereby different models of sequence evolution are applied to distinct subsets of a data matrix evolving under different functional constraints, and 2) incongruence between gene trees and species trees (reviewed in Liu et al. 2015). I apply these methods to address the deeper relationships in the Caprimulgidae (nightjars and nighthawks), a family with a striking but understudied evolutionary history.

The Caprimulgidae were long divided into two subfamilies, nightjars (Caprimulginae) and nighthawks (Chordeilinae) based on several morphological characters including wing shape, palate structure, and rictal bristles (e.g., Oberholser 1914, Ridgway 1914, Peters 1940, Hoff 1966, Cleere 1998). However, the exact composition of the two subfamilies was never settled, with several genera (*Podager*, *Eurostopodus*, *Veles*, *Nyctiprogne*) being shifted back and forth due to presence or absence of some of these characters (e.g., Holyoak 2001, Whitney et al. 2003). Moreover, it was clear that these traits might be prone to convergence because they

were associated with foraging ecology— nightjars typically sally after flying insects from an exposed perch at night, while nighthawks pursue flying insects during sustained flight at dusk and dawn. A second major issue concerned the large genus *Caprimulgus* (*sensu lato*), with 55–57 species and a cosmopolitan distribution, which appeared to be a grab bag of taxa with an ancestral body plan and few derived features (Cleere 1998).

Although a number of authors have commented on the morphology and anatomy of various exemplars of Caprimulgidae, most did not have sufficient taxon sampling to address relationships across the family in any detail (e.g., Oberholser 1914, Wetmore 1919, Hoff 1966, Bühler 1970, Schodde & Mason 1980, Mayr 2002, Mayr et al. 2003). An exception was Mayr (2010) who examined eight caprimulgid genera but did not find or did not analyze phylogenetically informative variation within the family. The only morphological study with truly extensive sampling of Caprimulgidae is the recent osteological analysis of Costa (2014), who examined nearly 50 species and all genera but *Veles*.

Molecular studies have begun to clarify caprimulgid phylogeny, suggesting a complex biogeographic and evolutionary history. The DNA hybridization data of Sibley & Ahlquist (1990) and mitochondrial cytochrome b (MT□CYB) sequence data of Mariaux & Braun (1996) first indicated that the two traditional subfamilies were not monophyletic. Barrowclough et al. (2006) used recombination activating gene-1 (RAG□1) to investigate the Caprimulgidae at the generic level, finding strong support for the placement of *Eurostopodus* sister to the rest of Caprimulgidae and for polyphyly of *Caprimulgus* (*sensu lato*). They also showed that most caprimulgid

species belong to one of four major geographically relevant clades, either restricted to the New World or the Old World. More data from MT CYB and MYC, the cellular homolog of the myelocytomatosis viral oncogene, reinforced these conclusions (Larsen et al. 2007; Braun & Huddleston 2009).

The phylogeny and classification of Caprimulgidae underwent a significant overhaul with the work of Han et al. (2010), including new generic designations for many groups. These authors used data from three genes: MT CYB, MYC, and growth hormone (GH). Their work mostly confirmed the findings of Barrowclough et al. (2006), although the relative placement of the four major geographic clades differed. With more comprehensive taxon sampling, Han et al. (2010) detected two previously unappreciated long branches: a deep split within *Eurostopodus* (justifying resurrection of the genus *Lyncornis*) and the Malagasy endemic "*Caprimulgus*" *enarratus*, for which they erected the new genus *Gactornis*. Most recently, Sigurdsson & Cracraft (2014) studied the phylogeny of New World Caprimulgidae at the species, and in some cases subspecies, level with data from four genes (including RAG 1 and MT CYB). Their resolution of relationships between the major geographic clades is compatible with those found by Barrowclough et al. (2006), though no study has yet found strong support for one of the alternative topologies. I will use the generic nomenclature of Han et al. (2010) and follow Sigurdsson & Cracraft (2014) in referring to the four major geographic clades as 1) the Poorwill Clade, containing mostly North and Central American nightjars (genera *Antrostomus*, *Nyctiphrynus* and *Siphonorhis*); 2) the Nighthawk Clade (the New World genus

*Chordeiles*); 3) the South American Clade (the Neotropical genus *Hydropsalis*); and 4) the Old World Clade, a radiation including African, Asian, Australian, and European lineages (genus *Caprimulgus sensu stricto*).

In this study, I sought to provide an improved estimate of the higher-level phylogeny of Caprimulgidae by combining the datasets of Barrowclough et al. (2006) and Han et al. (2010). The expanded dataset effectively doubled the number of characters in either previous one, and incorporates heterogeneous molecular marker types, including mitochondrial and nuclear genes, introns, exons and untranslated regions (UTRs). I added new RAG-1 sequences to generate a complete data matrix for 36 taxa addressing key basal nodes in the family and the placement of all major geographic clades. I partitioned and analyzed the matrix using a variety of *a priori* and *a posteriori* partitioning methods, and attempted to identify the species tree despite gene tree-species tree incongruence with a recently described quartet assembly approach, SVDquartets (Chifman & Kubatko 2014, 2015). Lastly, I followed up on the report by Barrowclough et al. (2006) that some caprimulgids have elevated GC content and excessive heterozygosity at the RAG-1 locus by exploring these parameters in all of my nuclear loci.

## Methods

### *Sequencing*

Barrowclough et al. (2006) previously obtained RAG-1 sequences for 24 species examined in this study. Using methods described by Groth & Barrowclough

(1999), and Barrowclough et al. (2006), I sequenced 12 additional taxa for the RAG-1 exon to obtain a common set of critical species for comparison with the Han et al. (2010) study (see Table 1.1 for voucher and GenBank accession numbers). The placement of *Hydropsalis parvulus* differs between Han et al. (2010) and Barrowclough et al. (2006), so both of those vouchered specimens were included here to test for possible contamination or mislabeling. The new RAG-1 sequences were assembled using Sequencher software (version 5.1; Gene Codes: Ann Arbor, MI), and aligned manually. All new RAG-1 sequences were examined for 1) indels that were not a multiple of three base pairs in length, 2) unexpected stop codons in the reading frame, and 3) unexpectedly similar (chimeric or contaminated) portions of sequence between taxa before inclusion in this study.

*Dataset generation*

A complete matrix of four loci for 36 taxa was generated by combining the aligned dataset of RAG-1 with aligned data from Han et al. (2010) using PAUP* (version 4.0a130; Swofford 2003). The four loci include: the entire MT-CYB coding sequence; parts of exons 2 and 3, and all of intron 2 from GH; part of intron B, all of exon 3 and part of the 3' UTR of MYC; and most of the exonic region of RAG-1. Alignments of MT-CYB, GH, and MYC were initially done in Clustal X (version 1.8.3; Thompson et al. 1997), then edited manually by Han et al. (2010). The resulting 7,104 base pair (bp) aligned dataset doubles the size of either original, and

combines multiple lines of genetic evidence. Representatives from every major nocturnal lineage of *Strisores* were included to allow outgroup rooting of all trees. The aligned data matrix is deposited in Treebase (ID # 19469).

I did not incorporate the data from Sigurdsson & Cracraft (2014) in this analysis for two reasons. First, due to their focus on New World taxa, the differences in taxon sampling—especially of outgroups—would require substantial further sequencing. Second, I do not expect the addition of mitochondrial loci to help elucidate phylogeny at this evolutionary depth (further discussed later). Therefore, including the Sigurdsson & Cracraft (2014) data would have added limited data from only one additional locus (intron 9 from the nuclear aconitase gene) to this study.

*Heterozygosity and Base Composition*

In order to determine whether the previously reported (Barrowclough et al. 2006) high GC content and heterozygosity at the RAG-1 locus extended to other loci, I estimated heterozygosity and overall base composition on my data. Barrowclough et al. (2006) noted that heterogeneity in base composition was largely driven by elevated GC content at third codon positions (GC3) in some species. However, two of the three nuclear loci used here include extensive noncoding regions. Therefore, I compared the distributions of overall GC content, rather than just GC3. I used contingency G-tests of A + T versus G + C proportions among species, separately for each of the three genes, to calculate base composition heterogeneity, and employed non-parametric Kolmogorov-Smirnov (KS) tests to calculate distributions of observed heterozygosity between genes. To visualize GC differences among the

genes, I used standard McGill et al. (1977) box plots to summarize their distributions; medians, upper and lower quartiles (box), and ranges (whiskers) were found. Species with GC content greater than, or less than, 1.5 times the inter-quartile range from their respective quartile, were indicated by dots. The mitochondrial MT-CYB locus is not relevant for analyses of heterozygosity, and hence was excluded from these analyses.

*Partitioning Schemes and Alternative Models*

*A priori* partitions of the sequence data were chosen based on the expectation that rates and patterns of molecular evolution will vary among loci, subcellular compartments, and distinct functional regions of genes. For example, a model of sequence evolution applied to the slowly-evolving exons may be inappropriate to apply to the more quickly-evolving introns. I tested six *a priori* partitioning schemes that subdivide the dataset in various ways, ranging from unpartitioned, to very simple subdivisions, to a more complex scheme (Table 1.2). For comparison, I tested the *a posteriori* method of partitioning implemented in PartitionFinder (version 1.1.1; Lanfear et al. 2012). Given user specified subsets of the data, PartitionFinder uses a heuristic search to find an optimal partitioning scheme by searching for the best model for each subset, and combining subsets that conform to similar models of sequence evolution. I specified 18 possible subsets of the data, including the first, second and third positions of each exon in each gene, as well as subsets for the UTR and each intron. The best partitioning scheme was selected by PartitionFinder using the corrected Akaike information criterion (AICc; Akaike 1973, Hurvich & Tsai

1989). I employed the 'greedy' search algorithm, for the sake of computational tractability.

To test the effect of alternative models on phylogenetic analysis, I conducted maximum-likelihood (ML) analyses of each partitioning scheme and each individual gene under two models. First, an independent general-time reversible model with estimated proportion of invariant sites and gamma-distributed rate variation among sites (GTR + I + G; the most highly parameterized model available) was applied to each subset. Second, I set user-defined partitioning schemes, and allowed PartitionFinder to select the best model for each subset in a given scheme. All PartitionFinder runs used unlinked branch lengths, and searched all models of sequence evolution implemented in the program. Model descriptions may be found in Posada (2008). The number of parameters in each partitioning scheme was calculated automatically by PartitionFinder. To calculate that number for my schemes under GTR+I+G, I ran each scheme through PartitionFinder with both the scheme and model (GTR+I+G for each subset) user-defined.

*Phylogenetic analyses*

ML tree searches were conducted for each individual gene and all partitioning schemes of the combined dataset using GARLI (version 2.01) with partitions unlinked (Zwickl 2006). In order to ensure thorough searches for optimal tree topologies, I used the 'searchreps' option to vary the number of search replicates performed within a GARLI run. After 100 GARLI runs with a given number of search replicates, I compared topologies between best trees from each run using the

'treedist' function in PAUP* (version 4.0a130; Robinson & Foulds 1981). When the best topologies were identical for all 100 runs, I assumed the number of search replicates was sufficient to produce the optimal tree topology. The number of search replicates required to satisfy this criterion was two in all cases except the individual gene trees for GH and MT□CYB, which required 40 search replicates. Overall tree lengths were calculated in PAUP* (version 4.0a146) using the function 'describetrees/brlens'.

To evaluate nodal support, 100 non-parametric bootstrap datasets were generated and subjected to GARLI tree searches, with 1 search replicate for each bootstrap run. Nodal support values were tabulated as the number of bootstrap runs in which a particular node appears, and plotted on the optimal topology for each gene or partitioning scheme using the SumTrees program (version 3.3.1) in the python library DendroPy (version 3.12.0; Sukumaran & Holder 2010).

To address gene-tree/species-tree discordance, I applied a new coalescent-based species tree method which uses the full data matrix directly, without estimating individual gene trees or utilizing computationally-inefficient Bayesian statistics (SVDquartets; Chifman & Kubatko 2014, 2015). This method computes the probability distribution of site patterns at the tips of the tree by integrating over the probability distribution of gene trees under the coalescent model. It was designed for single nucleotide polymorphism data, but has been demonstrated to perform well on multi□locus datasets, such as ours. I employed the SVDquartets code implemented in PAUP* (version 4.0a147), conducting exhaustive sampling of quartets ('eval=all')

and 100 non-parametric bootstrap replicates. All trees were rooted to known outgroups representing each of the other nocturnal families of *Strisores*, and have been deposited in Treebase (ID # 19469).

**Results & Discussion**

*Heterozygosity and base composition*

In an earlier report on RAG-1 variation in Caprimulgiformes, Barrowclough et al. (2006) found that GC3 composition was correlated with heterozygosity in a clade of Old World nightjars. I extended that investigation to the two additional nuclear loci examined here to see if the RAG-1 results represented a general, perhaps genome-wide, phenomenon. The number of species showing high levels of heterozygosity for RAG-1 was greater than for GH, which in turn exceeded that of MYC (Figure 1.1). The RAG-1 distribution of heterozygosity was significantly different from that of GH ($P < 0.05$) and of MYC ($P < 0.01$), but the distributions for GH and MYC were not statistically divergent ($P > 0.05$).

Median GC composition was highest in GH (0.49) and lowest in MYC (0.44) (Figure 1.2). Each of the pairwise comparisons of the overall distributions was significant at the 0.01 level (Figure 1.2A). There was significant heterogeneity in GC composition among taxa for RAG-1 ($\chi^2 = 67.6$, df = 35, $P < 0.01$), but not for either MYC ($\chi^2 = 7.4$, df = 35, $P > 0.5$) or GH ($\chi^2 = 38.8$, df = 35, $P > 0.1$). Although heterogeneity in base composition was not statistically significant for GH, its range (0.064) was nearly equal to that of RG-1 (0.068), and the correlation in base

composition between the two loci was significant ($P = 0.01$; Figure 1.2B). However, the amount of variation explained by the correlation was not very large ($R^2 = 0.18$).

As reported by Barrowclough et al. (2006), correlation analysis verified that heterozygosity and base composition were correlated across species for RAG-1 ($R^2 = 0.26$, $P < 0.01$), but the correlation did not extend to the other two genes (MYC: $R^2 = 0.02$, $P > 0.1$; GH: $R^2 = 0.00$, $P > 0.5$). These results strongly suggest that the condition identified in RAG-1 is not a genome-wide phenomenon attributable to such causes as larger effective population sizes, interspecific hybridization, etc. In the Domestic Fowl (*Gallus gallus*), these three loci reside on separate chromosomes and so it appears that one of the more probable causes is increased mutation associated with a GC-rich isochore encompassing the region surrounding the RAG-1 gene in some taxa. This is indicated by the comparatively long branch lengths seen in the Old World Clade of the RAG-1 gene tree relative to the other caprimulgid clades (Supplemental Material Figure 1), and perhaps to a lesser extent in the GH gene tree (Supplemental Material Figure 2). This information would potentially be of considerable importance in phylogenetic reconstruction as it informs us about the substitution process for RAG-1. Unfortunately, most current phylogenetic inference algorithms, such as those used here, utilize DNA sequence partitions taken across the entire set of taxa in the study. Partition heterogeneity across taxa (i.e., failure of base composition stationarity) is a much more difficult process to model for all but the smallest datasets (Galtier & Gouy 1998). In this case, the fact that the individual phylogenies produced by the three nuclear genes are largely concordant suggests the lack of stationarity is a small effect relative to the overall historical signal (see also

Supplemental Material Figure 3 for the MYC gene tree).

*Data partitioning and alternative models*

Despite testing seven different partitioning schemes ranging between one (unpartitioned) to six subsets of the concatenated dataset, the resulting trees were very similar in topology and bootstrap support across all treatments. The major effects of my partitioning tests were on tree length and likelihood score, which both increased with increased number of partitions (Table 1.3). This can be attributed to an improvement in the ability of a more parameterized model to account for sequence evolution. With my data, applying as few as three partitions (roughly tripling the number of parameters) was enough to improve estimation of the model. The greatest likelihood score was seen using the automated partitioning software on my dataset. The scheme chosen by PartitionFinder had the same number of data subsets as my most complex *a priori* scheme (6), but it fit a model that was slightly less complex (459 vs. 462 parameters), and produced a substantial improvement in likelihood score (~ 270 units). This demonstrates the value of the automated search for combination of subsets, and of sorting data by quantitative patterns that may not have been expected given preconceived notions of molecular evolution. With the exception of the 'Unpartitioned' scheme, and individual genes MT□CYB and RAG□1, the models chosen by PartitionFinder all had fewer parameters than GTR + I + G. Despite this, differences in likelihood score and tree length for a given partitioning scheme were modest (Table 1.3), and identical tree topologies were found under both models tested. Similarly, both the partitioning schemes and the models applied had slight but

inconsistent effects on bootstrap support values (Supplemental Material Table S1, Supplemental Material Figures S1–S4).

The seven partitioning schemes tested resulted in only two topological changes (Figure 1.3). The two simplest schemes ('Unpartitioned' and partitioned 'By Gene') produced trees that differed by both changes, while all other schemes produced a single topology agreeing with the 'Unpartitioned' scheme in one area and the 'By Gene' partitioning scheme in the second area. All schemes with more than one partition agreed on the resolution of the three *Hydropsalis* species, but the 'By Gene' partitioning scheme differs from all others in outgroup topology, placing the Oil bird (*Steatornis caripensis*) sister to owlet nightjars (*Aegotheles insignis*) instead of potoos (*Nyctibius grandis;* Figure 1.3). The 'By Gene' scheme is perhaps the least sophisticated partitioning of my dataset, and may overemphasize the signal of the mitochondrial locus relative to the other partitions. Mitochondrial loci evolve quickly, and may be too saturated to resolve phylogeny at this evolutionary depth. The other partitioning schemes tested here, as well as the unpartitioned analysis, may average out the signal of the mitochondrial locus with the nuclear loci and provide a better estimate of evolutionary history. In addition, a recent analysis that incorporates other relevant taxa from the *Strisores* also finds oilbird and potoos to be sister taxa (Prum et al. 2015).

*The maximum likelihood topology*

Our best estimate tree from the concatenated dataset is based on the 'PartitionFinder' partitioning scheme run under the model selected by that program

(Figure 1.4). It resolves the relationships among the four major clades of caprimulgids, placing the South American Clade basal with 87% bootstrap support, and the Old World and Nighthawk Clades sister with 96% bootstrap support. *Gactornis* is firmly placed sister to the four major clades with 98% support, with *Lyncornis* and *Eurostopodus* successively more basal. *Hydropsalis rufiventris* and *H. leucopyga,* two taxa formerly considered to belong to the ‘‘nighthawk” subfamily Chordeilinae, are firmly placed within the South American Clade as successive basal branches.

This tree shows increased resolution and bootstrap support relative to comparable prior analyses (summarized in Figure 1.5, tabulated in Supplementary Material Table S2). Of 30 ingroup nodes in the current tree, 12 have increased bootstrap support over the study of Han et al. (2010), while only four have decreased support. Similar increases in nodal support were seen over Barrowclough et al. (2006), and, in comparison to Sigurdsson & Cracraft (2014), support increased for nine nodes and decreased for five. Decreased support in the latter case was concentrated on shallow nodes and can be attributed to the smaller amount of mtDNA sequence included in my dataset.

*The SVDquartets topology*

Our SVDquartets tree (Supplementary Material Figure S5) has the same topology for the major clades of Caprimulgidae as the best ML tree, though with lower bootstrap support overall. The positions of *Eurostopodus* and *Lyncornis* are switched in this tree and the outgroup topology changes, again with lower bootstrap support. There are substantial differences within the South American Clade.

*Hydropsalis maculicaudus* and *H. climacocerca* are sister as found previously in the 'Unpartitioned' analysis (Figure 1.3), but now with bootstrap support of 74%, the highest seen for this node in any of my analyses. *Hydropsalis rufiventris* and *H. leucopyga* are nested well within the South American Clade in the SVDquartets tree, as opposed to being successive basal branches as in the best ML tree from the concatenated dataset.

This analysis provides an interesting perspective on phylogeny in this group – it confirms some key features of the tree, but it also differs in important ways. However, I consider any gene-tree/species-tree type analyses based on these data preliminary for two reasons. First, four genes is a limited sampling of loci, subject to potential sampling error. Second, one of my genes is a rapidly evolving mitochondrial locus that can be expected to have weak phylogenetic signal deep in the caprimulgid tree due to mutational saturation effects. This was documented by Larsen et al. (2007) for their caprimulgid MT-CYB data. The relative strengths and weaknesses of gene tree/species tree vs. concatenated analyses of molecular sequence data are topics of much debate (e.g., Gatesy & Springer 2014, Hahn & Nakhleh 2015, Liu et al. 2015, Simmons & Gatesy 2015, Tonini et al. 2015, Edwards et al. 2016), and I view my results here with caution.

*Individual gene trees*

All individual gene tree analyses identified the four major geographic clades, with the exception of the MT□CYB tree, which places *Siphonorhis brewsteri* within the South American Clade, instead of the Poorwill Clade (Supplemental Material

Figure S4). The RAG-1 and GH gene trees are largely congruent, with the exception that GH places *Lyncornis* and *Eurostopodus* as sister taxa, rather than successive basal branches (Supplemental Material Figures S1, S2). The MYC gene tree has the branching order of *Lyncornis* and *Eurostopodus* reversed from that seen in the RAG-1 tree, and fails to resolve the branching order of the Old World, Nighthawk and Poorwill Clades (Supplemental Material Figure S3). Overall, analyses of MT-CYB yield a very different topology than the nuclear genes, with much lower support (several nodes < 50% bootstrap). The lack of phylogenetic signal in this locus (at this depth) can be attributed to substitutional saturation. Also of note is the extremely long branch leading to the frogmouths *Batrachostomus septimus* and *Podargus strigoides* in the MT-CYB tree, suggesting rapid evolution of mtDNA in Podargidae. Earlier authors have also noted very high levels of mtDNA divergence in Podargidae (Cleere et al. 2007). The same phenomenon exists to a lesser extent in the MYC tree, but not in the RAG-1 or GH trees.

*Basal taxa*

I confirm that the genera *Lyncornis* and *Eurostopodus* are the earliest branching taxa in the Caprimulgidae. These genera were formerly lumped together in the genus *Eurostopodus,* and Sibley & Ahlquist (1990) suggested that this group be treated as a separate family. However, with increased taxon sampling, Han et al. (2010) uncovered additional complexity, detecting another early branching lineage, *Gactornis*, and the deep split in *Eurostopodus sensu lato* that justified resurrection of *Lyncornis.* While the 63% bootstrap support I found for *Eurostopodus sensu stricto* as the earliest branch is the highest support seen in any study so far, it still does not

confidently resolve the trichotomy with *Lyncornis,* and my SVDquartets tree has the branching order reversed (Figure 1.5). The morphological evidence presented in Costa (2014; Figures 4 and 5 of that study) places *Lyncornis* as the most basal lineage with 31 % bootstrap support, with a paraphyletic group of three *Eurostopodus* species branching next, and *Gactornis enarratus* sister to the Old World Clade. Thus, separation of any of these genera as distinct families or subfamilies cannot yet be justified. Both *Lyncornis* and *Eurostopodus* are on relatively long branches in all molecular trees, so sampling additional species from both genera may facilitate resolution.

*Core caprimulgids*

Barrowclough et al. (2006) and Sigurdsson & Cracraft (2014) had the South American Clade sister to the other three core caprimulgid clades, but with < 50% and < 60% bootstrap support, respectively. In contrast, Han et al. (2010) had the Poorwill Clade sister to the other three core clades, with equally low support (Figure 1.5). By increasing the size of the dataset, I have achieved stronger resolution of the major caprimulgid clades, and find the South American Clade to be sister to the other three. This is likely due to the congruent signal provided by RAG-1 and MYC, which are both slowly-evolving nuclear genes. Another nuclear locus, intron 9 from the aconitase gene, also yields this topology (Sigurdsson & Cracraft 2014), as does my SVDquartets analysis, leaving the GH gene tree as the only nuclear locus that does not support it. While my MT-CYB gene tree does not have this topology, it also does not agree with the GH gene tree nor with that of another mitochondrial locus, NADH

dehydrogenase subunit 2 (see Sigurdsson & Cracraft 2014, Figure 1). As in previous studies, I find the Nighthawk Clade to be sister to the Old World Clade, with increased support (Figure 1.5).

Comparing the core caprimulgid lineages between my best tree and the Sigurdsson & Cracraft (2014) tree, I see a few slight differences in topology that are likely due to taxon sampling and/or genes used. Rearrangements of three terminal taxa (*Caprimulgus climacurus, europaeus* and *vexillarius*) are apparent within the Old World Clade and three others in the Poorwill Clade (*Antrostomus vociferus, carolinensis* and *ridgwayi*), though the topology of the early-branching members of each respective clade are the same between the trees. Identical topology and bootstrap support values can be seen in both studies within the Nighthawk Clade. Basal within the South American Clade, my tree has *Hydropsalis leucopyga* and *H. rufiventris* branching in succession. This topology is present in Barrowclough et al. (2006), and Han et al. (2010), but the order is reversed in Sigurdsson & Cracraft (2014). *H. maculicaudus* is sister to *H. anomalus* in my best estimate tree, but sister to *H. climacocerca* in Sigurdsson & Cracraft (2014) and in my SVDquartets tree. Most of these differences are not strong conflicts in terms of nodal support, and may be addressed by increasing the size of the data matrix.

*Nighthawks vs. nightjars*

*Hydropsalis leucopyga* and *H. rufiventris* were formerly placed with the nighthawks in the subfamily Chordeilinae. All of these taxa have pointed wings and reduced rictal bristles, and forage on the wing at dusk and dawn by coursing rapidly

over open spaces (*H. leucopyga* over water, *H. rufiventris* over forest, and *Chordeiles* over open country). *Eurostopodus* and *Lyncornis* share some of these traits (Cleere 1998) and were also placed in Chordeilinae by some authors (e.g., Holyoak 2001). However, all molecular evidence agrees in firmly placing *Eurostopodus* and *Lyncornis* as the earliest branches in the family, and *H. leucopyga* and *H. rufiventris* within the South American Clade. Thus, the morphological similarities among these taxa are homoplasious and likely represent independently derived adaptations to aerial foraging. The osteological study of Costa (2014) also recovered *Eurostopodus* and *Lyncornis* as early branches and *H. leucopyga* and *H. rufiventris* (represented by *H. semitorquatus*) as lineages distinct from *Chordeiles* (but not monophyletic with the South American Clade).

*Biogeography*

The three earliest branches of the caprimulgid phylogeny have current distributions around the Indian Ocean; *Eurostopodus* in Australo-Papua, *Lyncornis* in South Asia and *Gactornis* in Madagascar (Cleere 1998). I can thus infer that the family may have originated in this general region. On the other hand, the two earliest branching of the four major caprimulgid clades are restricted to the New World (South American and Poorwill), while the Old World Clade is nested in the phylogeny sister to the Nighthawk Clade. Thus, diversification of the core caprimulgids in the New World appears likely, followed by a re-colonization and secondary radiation in the Old World. These scenarios were previously envisioned by Barrowclough et al. (2006), Han et al. (2010), and Sigurdsson & Cracraft (2014), and

are reinforced here with increased resolution and support at key nodes in the phylogeny.

**Acknowledgements**

I thank Matthew Kweskin for technical assistance on computational analyses, Laura Kubatko for discussion of SVDquartets, and David L. Swofford and Robert Lanfear for discussion of issues related to partitioning. I thank Charles Mitter for constructive comments on an early draft of this manuscript, Minh Le for assistance with some of the RAG□1 sequencing for this project, Thiago V. V. da Costa for sharing his unpublished dissertation, and Natalia Agudelo and Gustavo S. Cabanne for translating the abstract. NDW received Smithsonian Institution support from the Scholarly Studies Program, from the Consortium for Understanding and Sustaining a Biodiverse Planet, and as a Predoctoral Fellow. Computations in this paper were run on the Smithsonian Institution's High Performance Cluster.

## Tables

Table 1.1 GenBank accession numbers and voucher information for all new sequences generated for this study. Acronyms: CONACYT = Consejo Nacional de Ciencia y Tecnología, FMNH = Field Museum of Natural History, USNM = US National Museum of Natural History, AMNH = American Museum of Natural History, KUNHM = University of Kansas Natural History Museum, ANSP = The Academy of Natural Sciences, Philadelphia.

| Species | GenBank accession number | Voucher information |
|---|---|---|
| *Antrostomus ridgwayi* | KU361177 | CONACYT (Mexico) 415 |
| *Caprimulgus affinis* | KU361174 | FMNH 358300 |
| *Caprimulgus manillensis* | KU361180 | USNM B6090 |
| *Chordeiles pusillus* | KU361178 | USNM B12993 |
| *Eurostopodus argus* | KU361170 | AMNH DOT2401 |
| *Gactornis enarratus* | KU361171 | FMNH 438654 |
| *Hydropsalis anomalus* | KU361179 | KUNHM 3275 |
| *Hydropsalis anthonyi* | KU361173 | ANSP 4580 |
| *Hydropsalis nigrescens* | KU361175 | USNM B4478 |
| *Hydropsalis parvulus* | KU361176 | KUNHM 106 |
| *Hydropsalis whitelyi* | KU361181 | USNM B19022 |
| *Siphonorhis brewsteri* | KU361172 | KUNHM 8149 |

Table 2.1 Partitioning schemes used to study phylogenetic relationships across 36 species of Caprimulgiformes. The length of each partition is given in nucleotide base pairs (bp). Model chosen by PartitionFinder for each subset, as well as individual genes, are given. Model descriptions may be found in Posada (2008).

| Scheme | Partitions | Partition length (bp) | PartitionFinder model |
|---|---|---|---|
| Unpartitioned | n / a | 7,104 | GTR + I + G |
| Coding vs. non-coding | Codon positions 1 & 2 | 3,137 | GTR + I + G |
| | Codon position 3 | 1,565 | GTR + I + G |
| | Introns & UTR | 2,402 | TVMef + G |
| By gene | MYC | 1,318 | TVM + I + G |
| | MT-CYB | 1,143 | GTR + I + G |
| | GH | 1,765 | SYM + G |
| | RAG-1 | 2,878 | GTR + I + G |
| Coding positions vs. non-coding | Codon position 1 | 1,569 | SYM + I + G |
| | Codon position 2 | 1,568 | GTR + I + G |
| | Codon position 3 | 1,565 | GTR + I + G |
| | Introns & UTR | 2,402 | TVMef + G |
| Nuclear vs. mito vs. non-coding | Nuclear codon positions 1 & 2 | 2,375 | GTR + I + G |
| | Nuclear codon position 3 | 1,184 | TVM + I + G |
| | MT-CYB positions 1 & 2 | 762 | TVM + I + G |
| | MT-CYB position 3 | 381 | TIM + I + G |
| | Introns & UTR | 2,402 | TVMef + G |
| Nuclear vs. mito vs. introns vs. UTR | Nuclear codon positions 1 & 2 | 2,375 | GTR + I + G |
| | Nuclear codon position 3 | 1,184 | TVM + I + G |
| | MT-CYB positions 1 & 2 | 762 | TVM + I + G |
| | MT-CYB position 3 | 381 | TIM + I + G |
| | Intron | 2,012 | TVMef + G |
| | UTR | 390 | TIM + G |
| Partition-Finder | Introns | 2,012 | TVMef + G |
| | MYC UTR, MYC Exon 1st pos, MYC Exon 3rd pos, RAG-1 2nd pos, RAG-1 3rd pos, GH Exon1 1st pos, GH Exon1 3rd pos, GH Exon2 1st pos, GH Exon2 3rd pos | 2,764 | TVM + I + G |
| | MYC Exon 2nd pos, MT-CYB 1st pos | 573 | GTR + I + G |
| | MT-CYB 2nd pos, GH Exon1 2nd pos | 391 | TrN + I + G |
| | MT-CYB 3rd pos | 380 | TIM + I + G |
| | RAG-1 1st pos, GH Exon2 2nd pos | 984 | SYM + G |
| **Individual genes** | | | |
| MYC | n / a | 1,318 | TVM + I + G |
| MT-CYB | n / a | 1,143 | GTR + I + G |
| GH | n / a | 1,765 | SYM + G |
| RAG-1 | n / a | 2,878 | GTR + I + G |

Table 3.1 Results of alternate model analysis for each partitioning scheme, as well as individual genes used to determine phylogenetic relationships across 36 species of Caprimulgiformes. Number of parameters per scheme calculated by PartitionFinder. Maximum likelihood scores (-lnL) are those reported by GARLI. Overall tree length is presented in substitutions per site summed over the whole tree.

| Scheme | PartitionFinder Model | | | GTR + I + G | | |
|---|---|---|---|---|---|---|
| | # of Params | -lnL | Tree Length | # of Params | -lnL | Tree Length |
| Unpartitioned | 79 | 38,507.37 | 1.315 | 79 | 38,507.37 | 1.315 |
| Coding vs. Non-coding | 232 | 37,740.03 | 2.240 | 237 | 37,733.67 | 2.238 |
| By Gene | 311 | 37,362.41 | 2.213 | 316 | 37,361.02 | 2.212 |
| Coding Positions vs. Non-coding | 308 | 37,631.55 | 2.231 | 316 | 37,623.16 | 2.230 |
| Nuclear vs. Mito vs. Non-coding | 386 | 36,642.52 | 2.374 | 395 | 36,635.04 | 2.391 |
| Nuclear vs. Mito vs. Introns vs. UTR | 462 | 36,520.87 | 2.400 | 474 | 36,515.67 | 2.425 |
| PartitionFinder | 459 | 36,250.03 | 2.465 | 474 | 36,240.87 | 2.487 |
| **Individual Genes** | | | | | | |
| MYC | 78 | 5,088.23 | 0.545 | 79 | 5,087.70 | 0.547 |
| MT-CYB | 79 | 12,448.95 | 12.588 | 79 | 12,448.95 | 12.588 |
| GH | 75 | 7,339.50 | 1.125 | 79 | 7,338.93 | 1.125 |
| RAG-1 | 79 | 12,085.61 | 0.550 | 79 | 12,085.61 | 0.550 |

**Figures**

Figure 1.1 Distributions of observed heterozygosity for 36 species of caprimulgiform birds at three nuclear loci; distribution for RAG-1 is significantly different from that of both GH ($P < 0.05$) and MYC ($P < 0.01$); distributions for GH and MYC do not differ ($P > 0.05$): Kolmogorov-Smirnov two-sample tests.

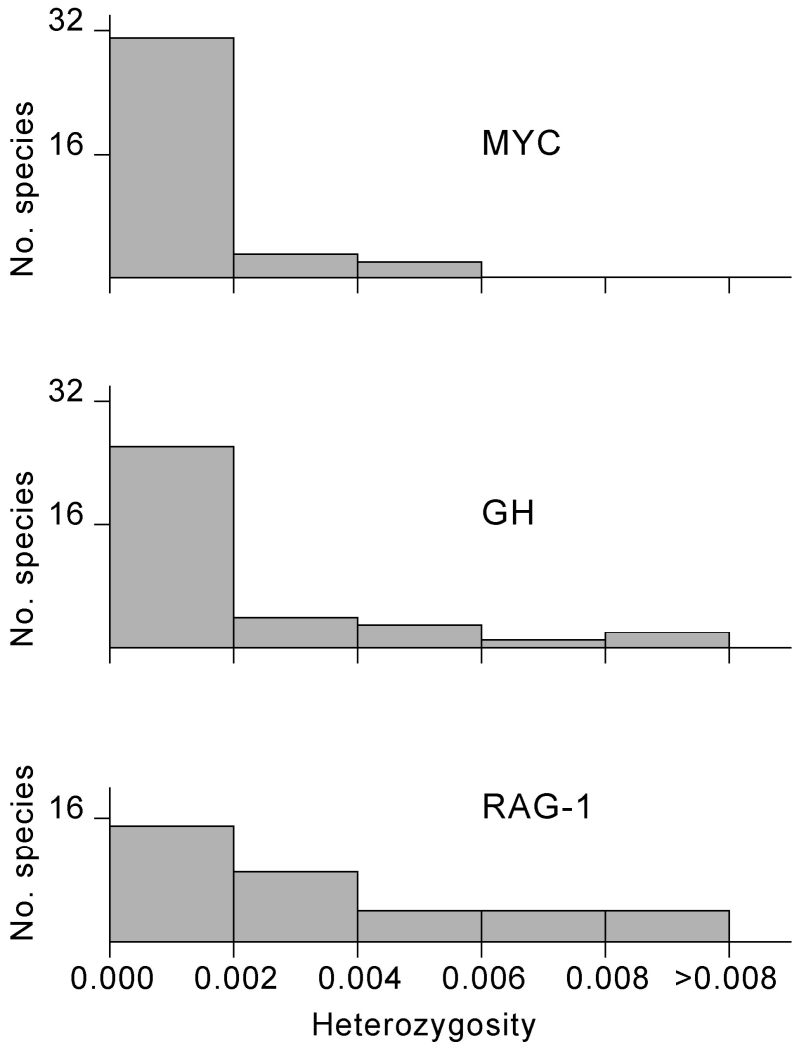


Figure 1.2 Base composition at three nuclear loci in 36 caprimulgiform birds. A: Box plots of distribution of GC proportions at each locus (box indicates first to third quartiles, interior line is the median; whiskers extend up to an additional 1.5 interquartile ranges; points farther from a quartile are indicated by dots); all three comparisons are statistically significant ($P < 0.01$: Kolmogorov–Smirnov two-sample tests). B: Scatter plot and linear regression ($R^2 = 0.18$) of GC proportion at GH gene versus RAG-1 gene.

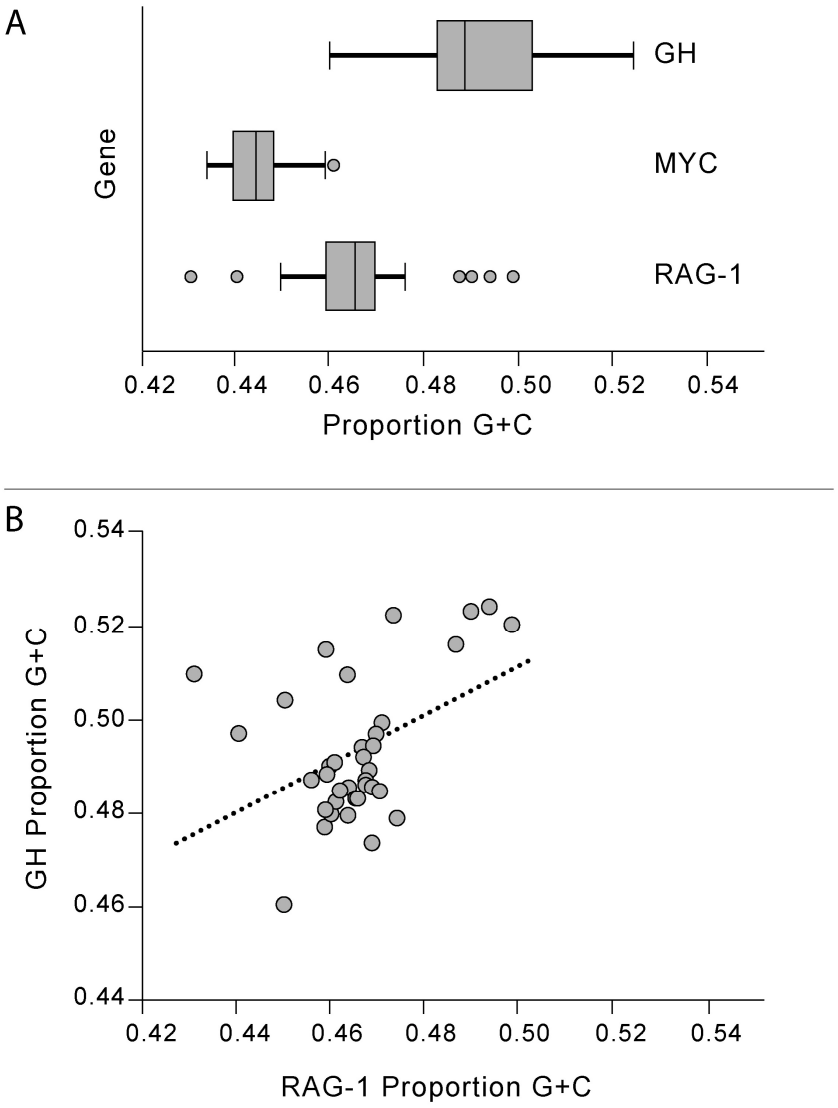



Figure 1.3 Cladogram representation of the three topologies found with the full dataset under alternative partitioning schemes for 36 caprimulgiform species. Branches are collapsed where topologies are identical, broadened tips represent multiple taxa. Differences between the three topologies are highlighted in grey. Bootstrap support values present on relevant nodes for 'Unpartitioned' and 'By Gene' analyses. For bootstrap support values of all partitioning schemes, see Supplemental Material Table S1.

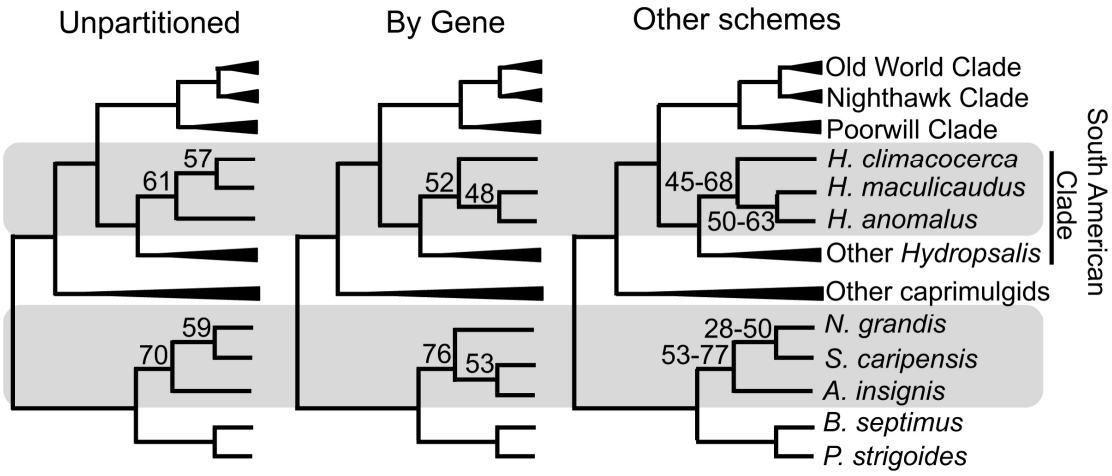

Unpartitioned
By Gene
Other schemes
Old World Clade
Nighthawk Clade
Poorwill Clade
H. climacocerca
H. maculicaudus
H. anomalus
Other Hydropsalis
Other caprimulgids
N. grandis
S. caripensis
A. insignis
B. septimus
P. strigoides
South American Clade
57
61
59
70
52
48
76
53
45-68
50-63
28-50
53-77

Figure 1.4 Best estimate of the phylogenetic tree for 36 caprimulgiform species from the data matrix partitioned by PartitionFinder as chosen by maximum likelihood score. Bootstrap support values shown for analysis under GTR + I + G above branches (black), below for analysis under model selected by PartitionFinder (grey) by AICc. Topologies were identical. Scale units are substitutions per site.

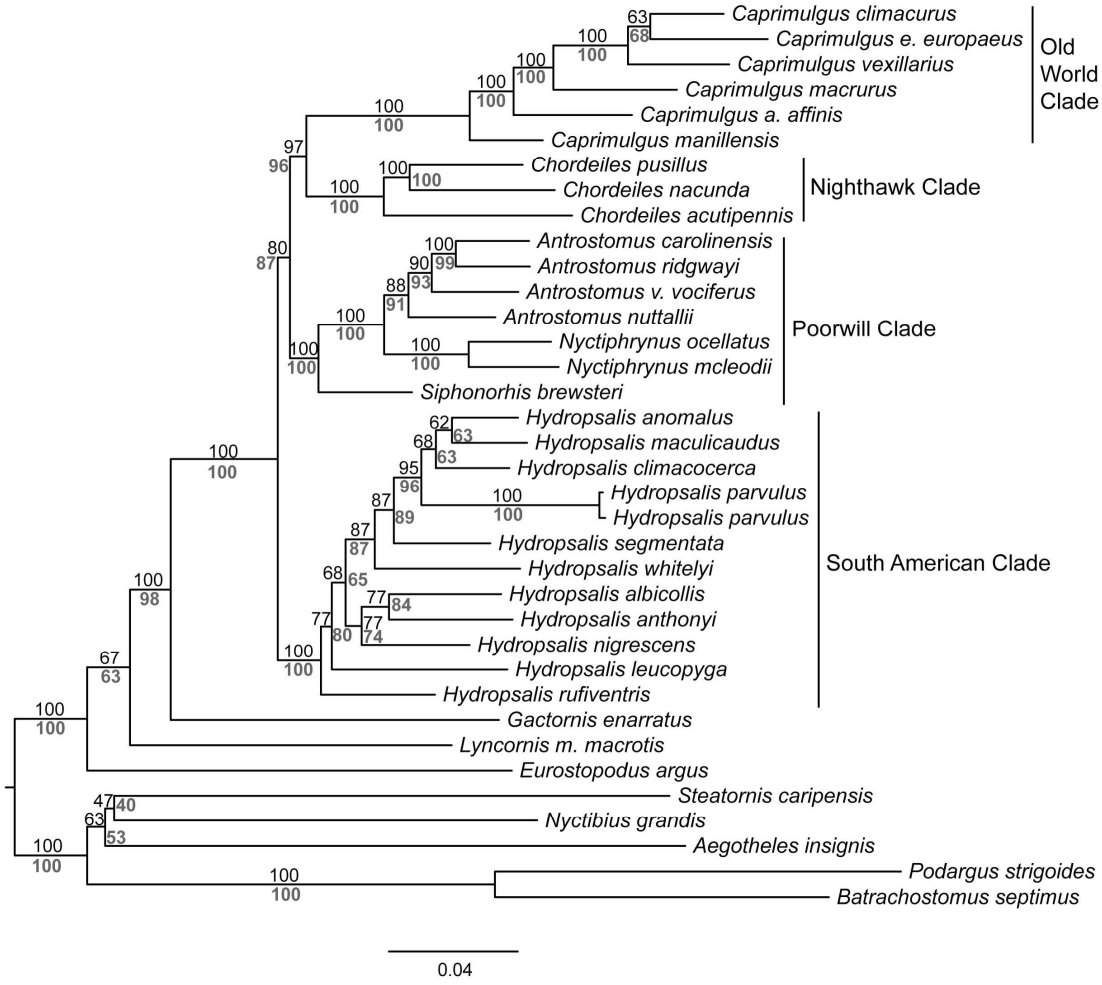
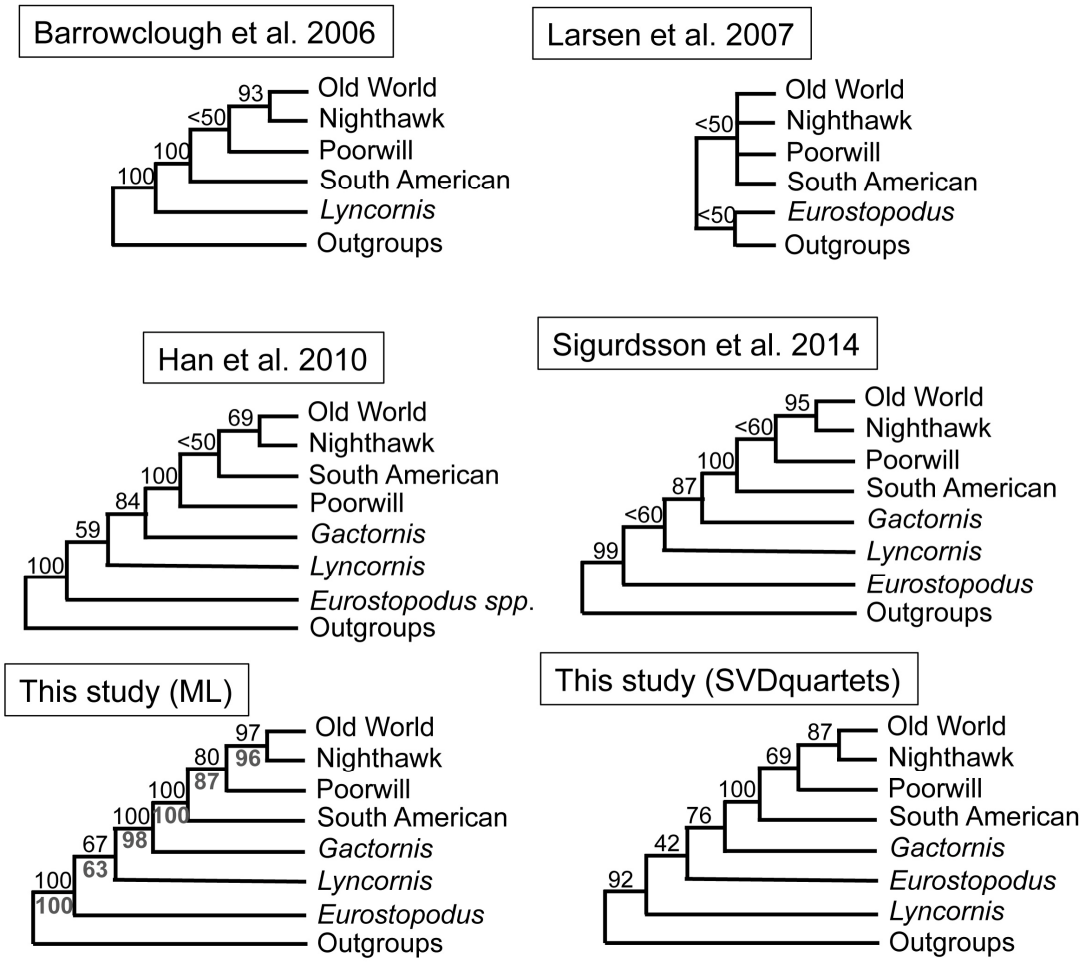


Figure 1.5 Caprimulgiform backbone trees with bootstrap support values from prior publications and this study obtained using maximum likelihood and coalescent (SVDquartets) analyses. Best estimate from this study depicted with bootstrap support values from analysis under GTR + I + G shown above branches (black), and with models selected by PartitionFinder shown below branches (grey). The word 'Clade' is omitted from major clade names for simplicity.



# Chapter 2: Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae)

This chapter was published as: White, N. D., Mitter, C. and Braun, M. J. (2017), Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae). Journal of Avian Biology, 48: 872–880.

**Abstract**

In this study, I apply a genome-scale set of molecular markers, ultraconserved elements, to fully resolve the phylogeny of a family of secretive, nocturnal birds, the potoos (Nyctibiidae). This dataset provides an opportunity to explore some challenges of phylogenetic analyses of genome-scale datasets, which I address in several ways. I generate data matrices ranging between 2,610–4,175 loci (1,477,319–3,848,295 aligned base pairs) that represent versions of the data differing in whether or not alignments were trimmed prior to concatenation, and whether 100 or 75% of all taxa were required to be represented by data for inclusion of a given locus. These matrices are analyzed with both maximum likelihood and coalescent algorithms, to check for artifacts of concatenation. Then, I subsample my data matrix by locus into randomly-selected replicates of 125–1,000 loci, and compare the topologies and statistical support of the resulting trees to look for evidence of systematic error. In analyses of complete matrices, I find strong statistical support for all ingroup nodes of the tree with no evidence for systematic error introduced by alignment trimming, missing data, or concatenation. I find further support for that topology in my subsampling analyses and statistical topology tests. The earliest branch of the tree separates

*Nyctibius bracteatus* from the rest of the potoos, followed successively by *N. grandis* and *N. aethereus*. Two pairs of species, *N. jamaicensis* plus *N. griseus*, and *N. leucopterus* plus *N. aethereus* comprise the distal tips of the tree. Finally, I compare my strongly supported topology to those of previous studies, and use the phylogeny to examine the evolutionary history of potoos.

**Introduction**

The family Nyctibiidae, the potoos, consists of seven currently recognized species, subdivided into 15 subspecies (Cohn-Haft 1999, Cohn-Haft 2016, Remsen et al. 2016). Potoos are part of a radiation of nocturnal birds including the nightjars, nighthawks, oilbird, frogmouths, owlet- nightjars and allies. The diurnal clade *Apodiformes* (swifts and hummingbirds) is embedded within this radiation, and together they comprise the strongly established clade *Strisores* (Barrowclough et al. 2006, Ericson et al. 2006, Hackett et al. 2008; reviewed by Braun and Huddleston 2009).

Like most other nocturnal birds of this group, potoos rest during the day, protected by their cryptic coloration and body form, and forage for insects at night (Cohn-Haft 1999). Also like other nightbirds, they show adaptations for night vision, including possession of a tapetum lucidum (reviewed by Braun and Huddleston 2009). They differ from most other nightbirds in lacking rictal bristles around the beak, having a 'toothed' bill, and possessing notches in their upper eyelids, which are thought to allow some vision during the day without revealing their bulbous eyes to potential predators (Borrero 1974, Cohn-Haft 1999). Most species of potoos are also known for perching upright at the end of dead snags, assuming a cryptic pose in

which they bear a remarkable resemblance to the snag itself. Some descriptions even suggest that their plumage patterns resemble lichens growing on dead wood, enhancing their camouflage (Cohn- Haft 1999). This posture is similar to one taken by some species of frogmouths (family Podargidae), which are currently restricted to the Old World.

The current distribution of Nyctibiidae is entirely Neotropical, although the fossil record suggests that they were previously more widespread (Mayr 2005a,b). Several species overlap extensively in geographic range (Fig. 1.1). Potoos primarily inhabit lowland moist forest except for *Nyctibius maculosus*, which is found in montane moist forest. The potoos are conservative in external appearance, and plague systematists with their cryptic coloration and high degree of individual variation. Some species boundaries are based largely on differences in vocalizations, such as that between the common *Nyctibius griseus* and northern *Nyctibius jamaicensis* potoos. However, genetic divergence among potoos is high, suggesting that these species are quite old, despite their conservative morphology. For example, mitochondrial cytochrome b (MT-CYB) sequence divergence ranges from 11.1–16.2% in potoos (Mariaux and Braun 1996), greatly exceeding the range of □1–10% more typically observed among congeneric bird species (Brumfield et al. 1997, Braun and Huddleston 2009).

A strongly resolved phylogeny, prerequisite for understanding the contrasting patterns of phenotypic and genetic diversity within *Nyctibius*, is not yet available. The first molecular study, based on partial MT-CYB sequences, was unable to fully resolve relationships among the six species sampled, but did suggest that *Nyctibius*

*leucopterus* and *N. maculosus* are sister taxa (Mariaux and Braun 1996; Fig. 1.2). Allozyme data subsequently confirmed the grouping of *N. leucopterus* and *N. maculosus*, additionally finding that *N. griseus* is probably sister to that pair, and that *N. bracteatus* is likely sister to all other potoos (Brumfield et al. 1997; Fig. 1.2). In an analysis of all seven species based on combining data from MT-CYB with the cellular homolog of the myelocytomatosis viral oncogene (MYC; 2,274 bp total), Braun and Huddleston (2009) found strong support for the pairing of *N. griseus* with *N. jamaicensis*, and *N. leucopterus* with *N. maculosus*, but only weak evidence regarding the remaining relationships (Fig. 1.2). The most recent study of potoo relationships firmly places *Nyctibius bracteatus* as sister to all other potoos based on osteological evidence, but lacks strong resolution of the rest of the taxa (Costa 2014; Fig. 1.2).

In this study, I established a robust phylogeny for potoos using ultraconserved elements (UCEs; Faircloth et al. 2012, McCormack et al. 2012). The advent of high-throughput DNA sequencing technology (also referred to as 'next generation' or 'massively parallel') has enabled the rapid generation of genome-scale datasets. In the present case, the size of my UCE dataset represents a roughly hundred-fold increase over the most recent study of potoo phylogeny (Braun and Huddleston 2009). The extraordinary information content of these genome-scale molecular matrices has allowed for unprecedented resolution of phylogenetic relationships, but has also brought to light new issues in data analysis. These issues are particularly problematic when they are due to systematic error – errors that give increasing statistical confidence in the wrong answer when increasing amounts of data are analyzed

(Kumar et al. 2012, Hahn and Nakhleh 2015, Hosner et al. 2016). The possible artifacts most relevant to my dataset and the phylogeny in question are the effect of trimming sequence alignments, discussed by Capella-Gutierrez et al. (2009), and the effect of missing data, discussed by Hosner et al. (2016).

The information content of UCEs comes largely from the flanking regions that are captured as a result of targeting the ultra-conserved core. Trimming of UCE loci after alignment is potentially a double-edged sword in that trimming away the 'ragged' ends of alignments should reduce phylogenetic error due to sequence error or poor alignment, but possibly at the cost of losing informative sequence. Trimming algorithms were originally introduced to compensate for any potential noise introduced at the ends of alignments. These regions generally represent areas of lower sequencing coverage, and thus more sequencing error. These regions are also more difficult to align due to variation in sequence lengths across taxa. In the past, alignments were checked by eye to remove any potentially misaligned regions that could include non-homologous characters. However, this is not possible with a genome-scale dataset, and thus I rely on computational methods to scan alignments for us. Potential artifacts due to missing data are another concern in phylogenomic analysis (Hosner et al. 2016). A recent study applying UCEs to lizards found that varying matrix completeness (where completeness refers to percentage of taxa for which data are available for a given locus) can affect support for inferred relationships, and that maximum likelihood analyses of concatenated datasets benefitted from the inclusion of more genes, even though missing data increased (Streicher et al. 2015). Thirdly, high overall bootstrap support in genome-scale

datasets can mask strong conflicts among loci or groups thereof, which might suggest lower confidence in the total evidence tree (Salichos and Rokas 2014). One indicator of the existence of strong conflict would be marked variation in topology and support among trees built from subsets of loci.

Here, I gauged the impact of these potential artifacts on my results in several ways. I constructed datasets with varying levels of missing data and analyzed them before and after trimming. To look for conflict among loci or groups thereof, I randomly subsampled my loci, and compared the resulting trees. This exercise also enabled me to estimate the minimum number of loci required to find the total evidence tree with strong support. Having detected no evidence for artifacts (i.e. systematic error) due to analyzing genome-scale matrices, I compare my results to those of previous studies, and use the phylogeny to examine the evolutionary history of potoos.

## Methods

### *Taxon Sampling*

Taxon sampling for this study consisted of all seven currently recognized potoo species plus five representatives of the other major lineages of *Strisores* (one individual each, 12 species total; Table 1.1). Frozen tissue samples were assembled through my own fieldwork and loans from major museum collections. Purified DNA samples were available from my previous work on these species (Hackett et al. 2008, Braun and Huddleston 2009).

*Data collection and alignment*

UCE data were collected by sequence capture using the tetrapod probe set of Faircloth et al. (2012), following the laboratory protocols described in that paper and available at www.ultraconserved.org. This probe set is designed to enrich sequencing libraries for 5,060 UCE loci conserved across tetrapods. Sequencing of paired-end, 100 base pair (bp) reads was conducted on Illumina platforms (HiScan and HiSeq2000), recovering an average of 1,456,836 reads per species (600,814–2,995,105). Sequences were processed with the PHYLUCE software package (Faircloth 2016) including quality control, assembly using Velvet (ver. 1.2.09; Zerbino 2008) and matching of contigs to UCE locus probes using LASTZ (Harris 2007). Individual loci were aligned using SATé-II based on MAFFT (ver. 2.2.7; Liu et al. 2011).

*Matrix completeness, alignment trimming and subsampling of loci*

To test the effect of missing data, I generated two datasets varying in 'matrix completeness', a term I use to refer to the proportion of taxa represented for a given locus. In one dataset, every taxon is represented in every locus alignment (termed '100% complete'), while the other dataset has at least 75% of all taxa represented for every locus ('75% complete'). For the purposes of this study, 75% completeness represented a threshold where I could both test the effects of missing data by introducing a lot of loci (nearly doubling the dataset relative to 100% complete) where one or more taxa were not represented, but also maintain computational tractability for the variety of analyses I wanted to run.

For each of these datasets, I also subjected individual locus alignments to the trimming algorithm included in the PHYLUCE software in order to test the effect of alignment trimming on phylogenetic inference. The PHYLUCE trimmer parses a given alignment in 20 bp windows, removing 'ragged' ends, so that at least 65% of taxa are present in a given column, and at least 65% of residues are present within that window. For each window, it allows no more than 20% divergence between any row of the alignment and the consensus of all sequences in that window. Lastly, it removes any loci with a trimmed length shorter than 100 bp. Resulting datasets varied from 2,610–4,175 loci, were □1.5–3.8 million bp in length, and contained □43,000–98,000 parsimony informative sites (Table 1.2).

To test the resolving power of UCE data, I randomly selected subsets of the 100% complete, PHYLUCE-trimmed matrix to create matrices of 125, 250, 500, 750 and 1,000 loci. Three replicate matrices of each size were created. These matrices ranged from □69–572 kilobases in length, and contained □2,000–17,000 parsimony-informative sites (Table 1.3). All data matrices generated by this study are available in Treebase (ID # 20258).

*Phylogenetic analysis*

Model selection was conducted using jModeltest 2.1.10 (ver. 20160303; Darriba et al. 2012). A general-time reversible model with estimated proportion of invariant sites and gamma-distributed rate variation among sites (GTR+I+G) was selected as the most appropriate model according to the corrected Akaike information criterion for all my datasets (Akaike 1973, Hurvich and Tsai 1989). Maximum likelihood (ML) phylogenetic analyses of the 100% complete and 75% complete

datasets were conducted on concatenated matrices with GARLI (ver. 2.1; Zwickl 2006). One hundred independent runs of the program were conducted, using two search replicates in each run ('searchreps 2'). I used the 'treedist' function of PAUP* (ver. 4.0a150; Swofford 2003) to ensure that the same topology was found for the best replicate in each of the 100 runs. This method is used as a means of assessing the thoroughness of the search of treespace, and is described in White et al. (2016). To evaluate nodal support, I conducted 100 non-parametric bootstrap replicates, each with one search replicate. The relatively small number of taxa in this dataset (n = 12) enabled me to run tree searches guaranteed to find the optimal topology. I ran both branch and bound and exhaustive tree searches in PAUP* (ver. 4.0a150; Swofford 2003) under GTR+I+G, and the resulting topologies were identical to that found by GARLI, ensuring that I have found the optimal tree.

ML analyses of the randomly subsampled matrices were conducted using both RAxML (ver. 8.2.7; Stamatakis 2014) and GARLI (ver. 2.1; Zwickl 2006), under the GTR+I+G model. In RAxML, searches for the best topology consisted of 20 replicates, and bootstraps were run using the bootstopping criterion (option – 'autoMRE'). In GARLI, searches for the best topology consisted of 100 independent runs of 2 search replicates each, and 100 non-parametric bootstraps (one search replicate each) were run.

To test for the possibility that gene tree/species tree discordance introduces artifacts in my concatenated analyses, I applied a coalescent-based species tree method, SVDquartets (Chifman and Kubatko 2014, 2015). I employed the

SVDquartets code implemented in PAUP* (ver. 4.0a150; Swofford 2003), evaluating all possible quartets and conducting 100 non-parametric bootstrap replicates.

In all cases, bootstrap values were plotted on the optimal topology using the SumTrees program (ver. 4.0.0) in the python library DendroPy (ver. 4.0.2; Sukumaran and Holder 2010). All trees generated in this study were rooted to the *Strisores* outgroups and have been deposited in Treebase (ID no. 20258).

*Topology tests*

I employed a statistical topology test to assess the significance in likelihood score of my preferred tree over all plausible trees. To make the search computationally tractable, I constructed a plausible set of trees that incorporates *a priori* information on potoo phylogeny presented in Fig. 1.2, including the grouping of *N. leucopterus* with *N. maculosus*, as well as *N. griseus* with *N. jamaicensis*. *Nyctibius leucopterus* and *N. maculosus* form a sister clade in all prior analyses, and *N. jamaicensis* was previously classified as a subspecies of *N. griseus*, so these groupings are justified. I also fixed the positions of all outgroup nodes except for *Steatornis caripensis* and *Eurostopodus mystacalis*, again based on prior knowledge (Barrowclough et al. 2006, Hackett et al. 2008, Prum et al. 2015). I devised a constraint tree reflecting these known relationships, and used the constraint topology to generate all plausible dichotomous trees in PAUP* (ver. 4.0a150; Swofford 2003), resulting in 315 candidate topologies. To test the significance of differences in likelihood among trees, I used the approximately unbiased (AU) test of Shimodaira (2002), which is less conservative than the Shimodaira–Hasegawa test (Shimodaira & Hasegawa 1999), and was deemed valid for my application. The AU test was run in

PAUP* (ver. 4.0a150; Swofford 2003) using resampling of estimated log likelihoods (RELL) to generate 1,000 bootstraps. Tests were run for each of my four full matrices.

## Results & Discussion

### *Matrix completeness, alignment trimming and subsampling of loci*

PHYLUCE trimming had little effect on the number of loci included (only one locus was eliminated, in the 75% complete matrix), but total alignment length was reduced by about 40%. This reduction had no apparent effect, however, on the phylogenetic result. In each case, the trimmed and untrimmed matrices yielded the same topology, with 100% bootstrap support for every ingroup node in the ML analyses, and a minimum of 98% in the coalescent analyses (Fig. 1.3, 1.4). Matrix completeness similarly had no discernable effect on the phylogeny. The 75% complete matrices include □1.5 times as many loci as the 100% complete datasets, yet both yielded the same topology with extremely high bootstrap support for every ingroup node in both ML and coalescent analyses. Thus, I find no evidence that my results are sensitive to possible artifacts of alignment trimming or missing data, nor to artifacts of concatenation.

The random subset matrix trees agreed with each other, as well as with those from the complete matrices, with two exceptions. These exceptions represent the other possible resolutions of the trichotomy between *N. aethereus*, *N. grandis* and a clade of four small potoos (topologies B and C in Fig. 1.5). Topology B was optimal for one replicate subset of the smallest sample size (125 loci) and topology C was

optimal for one replicate subset of 500 loci (Fig. 1.5). In neither case was the node conflicting with topology A strongly supported. I originally used RAxML to analyze the subset matrices, and re-ran them with GARLI, which yielded the same topologies and similarly low support (Fig. 1.5). The results indicate that the alternative topologies reflect differing signals in the subsampled matrices, and not a bias of the software used. In sum, I find no evidence that the 100% bootstrap support for all nodes in the analysis of the full dataset is masking pervasive strong conflict – subsampling of genes generally produced the same topology with diminished signal. My subsampling results also suggest that substantially fewer loci than the 4,175 employed here would suffice to definitively resolve the potoo phylogeny; alternative topologies only appeared with subsets of 500 or fewer loci.

*Phylogeny and evolution of potoos*

As detailed above, ML and coalescent analyses of genome-scale UCE data yielded a very strongly supported phylogeny of potoos (Fig. 1.3, 1.4). The congruence of these two fundamentally different methods enhances confidence in my result. The tree agrees with all strongly supported nodes in previous molecular analyses (Fig. 1.2), and mirrors that found by Braun and Huddleston (2009), but with much stronger bootstrap support. My topology differs from that found in the osteological study by Costa (2014), but character support for conflicting nodes in that study is limited. In fact, it would only take one step to make the osteological tree identical to the UCE tree. Specifically, Costa's character 9 requires one more step on the UCE tree, while his character 24 is equally parsimonious on both topologies. I found no evidence for a variety of possible artifacts to which genome-scale datasets

are potentially susceptible. AU tests of my four full datasets reject all but 2 or 3 trees (Fig. 1.6). The trees which cannot be rejected all have the same ingroup topology, which mirrors that found in my ML and coalescent analyses (Fig. 1.3, 1.4). I therefore believe that the ingroup topology based on UCE data provides a robust framework for discussion of potoo evolution.

The earliest branch in the potoo tree separates the rufous potoo *N. bracteatus* from all other species. *Nyctibius bracteatus* is arguably the most divergent potoo on several counts. Previous molecular studies all found the highest genetic divergences within potoos to be those between *N. bracteatus* and all other species (Mariaux and Braun 1996, Brumfield et al. 1997). It is the smallest potoo, and bears a unique plumage pattern that more closely resembles the oilbird *S. caripensis* than any of the other potoos. The deeply rufous plumage of *N. bracteatus* is said to mimic dead foliage or dead wood in the forest understory where the species typically nests and roosts during the day, with the white spots resembling dappled sunlight or whitish fungi (Cohn-Haft 1999, B. Whitney pers. comm.). It also has a unique dark wedge in its otherwise yellow iris, the function of which is unknown (Cohn-Haft 1999), and bears a number of unique osteological traits, described in Costa (2014).

Although rarely encountered, *N. bracteatus* appears to be widespread in the tropical lowlands of South America. Behaviorally, it is unique in roosting during the day on thin, roughly horizontal branches, and perpendicular to the substrate, rather than adopting the branch/stub-mimicry roosting posture (frozen with bill up) for which potoos are famous. Furthermore, it is reported to remain in a ‘hunched’ posture (head down, bill nearly parallel with ground), and to regularly rock back and forth on

a vertical axis, assuming the motion of a dead leaf in the breeze (B. Whitney pers. comm.), which may boost blood circulation in this relatively sedentary bird (M. Cohn-Haft pers. comm.). *Nyctibius bracteatus* forages inside forests, principally in the understory, whereas other potoos forage mostly in the canopy and subcanopy, often at forest edges and around larger open clearings within forests (B. Whitney pers. comm.). Lastly, the principal 'song' of *N. bracteatus* is more complex and multi-syllabic than are the presumably homologous songs of other potoos (T. Costa and B. Whitney pers. comm.).

Among the remaining species, the next divergences separate first the great potoo *N. grandis*, then the long-tailed potoo *N. aethereus*, from the others. These two are the largest of the potoos. They are biogeographically similar in having disjunct populations in the Brazilian Atlantic Forest, as well as extensive distributions in lowland tropical South America, including both sides of the Andes. The great potoo is found on the borders of, or in openings in, dry to humid forests, while the long-tailed potoo is restricted to the interior of humid tropical forests. The final four species form a clade, which divides into two sister-group pairs. One of these consists of the similarly sized white-winged *N. leucopterus* and Andean *N. maculosus* potoos, as first reported by Mariaux and Braun (1996). Both are rarely encountered species of tropical South America. The Andean potoo is unique to the family in preferring the habitat of montane cloud forests, whereas the white-winged potoo is a secretive denizen of the canopy of tropical forests. The remaining sister species pair consists of the morphologically similar common *N. griseus* and northern *N. jamaicensis* potoos. Historically *N. jamaicensis* had been treated as a subspecies of *N. griseus*. The two

are now split based on their distinct vocalizations. *Nyctibius griseus* has perhaps the broadest distribution of all potoos, ranging over much of South America and into central America as far as Nicaragua. *Nyctibius jamaicensis* occurs in northern central America and on Jamaica and Hispaniola. Both are more commonly encountered than other potoos and may be found in a variety of habitats, from the borders of humid forests and river edges to dry woodlands and open areas.

Given the conserved cryptic plumage of potoos and the high degree of genetic divergence observed in all molecular studies to date, further research on geographic variation in potoos is in order, beginning with previously recognized subspecies and disjunct populations. This would include trans-Andean populations of *N. griseus*, *N. grandis* and *N. aethereus*, and populations of *N. leucopterus*, *N. aethereus* and *N. grandis* isolated in the Atlantic Forest. Research should include studies of molecular, morphological and behavioral (especially vocal) variation, as significant divergence in all of these character suites has been discovered in the examples cited above. While a paucity of traditional museum specimens and genetic resource materials has hampered the study of this elusive group, new technologies such as computed tomography scanning and next generation sequencing promise to facilitate extraction of new information from the specimens that do exist (Lim and Braun 2016).

Of special interest are the island subspecies *N. j. jamaicensis* of Jamaica and *N. j. abbotti* of Hispaniola. Their distributions conform to stage V of the taxon cycle theory (Ricklefs 1970), which would suggest that these populations may be old and substantially divergent from their mainland counterparts. Several other island endemic nightbirds have also proven to be much more divergent than previously

suspected. Examples include the Malagasy endemic *Gactornis enarratus* (Han et al. 2010), the Hispaniolan endemic *Siphonorhis brewsteri* (Han et al. 2010) and the Solomon Islands endemic *Rigidapenna inexpectata* (Cleere et al. 2007). *Rigidapenna*, which clearly belongs in a separate genus, was previously treated as a subspecies of *Podargus ocellatus*, and *Gactornis* had always been treated as a species of *Caprimulgus* until molecular data demonstrated that it was one of the oldest lineages of Caprimulgidae.

**Acknowledgements**

I acknowledge the genetic repositories listed in Table 1.1 for loans of the tissues used herein. I especially thank the intrepid field collectors and diligent curatorial staff whose work makes studies such as this possible. I thank Thiago V. V. Costa for sharing his unpublished dissertation, Bret Whitney and Mario Cohn-Haft for insight on the behavior and vocalizations of *N. bracteatus*, and Handbook of the Birds of the World for allowing use of the potoo plate images. Computations in this paper were run on the Smithsonian Institution's High Performance Computing Cluster, Hydra. Laboratory work was conducted at and with the support of the US National Museum of Natural History.

**Tables**

Table 2.1. Collection details for all specimens used in this study. Localities appear as given in their respective collection databases. Abbreviations: ABTC = Australian Biological Tissue Collection (South Australian Museum, Adelaide); ANSP = Academy of Natural Sciences, Philadelphia; ANWC = Australian National Wildlife Collection, Canberra; FMNH = Field Museum of Natural History, Chicago; KUNHM = Univ. of Kansas Natural History Museum, Lawrence; LSUMZ = Louisiana State Univ. Museum of Zoology, Baton Rouge; WLK = preparator L. K. Wang; UWBM = Univ. of Washington, Burke Museum, Seattle; SAMA = South Australian Museum, Adelaide.

| Common Name | Species | Tissue No. | Voucher No. | Locality | Collector |
|---|---|---|---|---|---|
| Long-tailed potoo | *Nyctibius aethereus* | LSUMZ B10877 | LSUMZ 156210 | PERU; Depto. Ucayali, SE slope of Cerro Tahuayo | A. S. Meyer |
| Rufous potoo | *Nyctibius bracteatus* | LSUMZ B4509 | LSUMZ 114641 | PERU; Depto. Loreto, lower Rio Napo region, E bank of Rio Yanayacu, ~90 km N of Iquitos | S. W. Cardiff |
| Great potoo | *Nyctibius grandis* | LSUMZ B15415 | LSUMZ 150545 | BOLIVIA; Depto. Santa Cruz, Velasco, Pre-Parque Nacional: "Noel Kempff Mercado," 30 km E of Aserradero Moira | J. M. Bates |
| Common potoo | *Nyctibius griseus* | ANSP 18217 (B3238) | ANSP 183090 | ECUADOR; Prov. Sucumbios, Imuya Cocha 0°34'S, 75°17'W, 200 m | F. Sornoza |
| Northern potoo | *Nyctibius jamaicensis* | KUNHM 2116 | KUNHM 92957 | MEXICO; Silvituc, 24 km S; Mexico, Campeche; 18.233, -90.200 | --- |
| White-winged potoo | *Nyctibius leucopterus* | LSUMZ B20315 | LSUMZ 165693 | BRAZIL; Amazonas, Munic. Manaus; km 34 ZF-3, Faz. Esteio, ~80 km N of Manaus | M. Cohn-Haft |
| Andean potoo | *Nyctibius maculosus* | LSUMZ B271 | LSUMZ 97586 | PERU; Depto. Cajamarca, Lucuma on the Sapalache-Carmen Trail | M. J. Braun |
| Oilbird | *Steatornis caripensis* | LSUMZ B32579 | LSUMZ 169580 | PERU; Cajamarca Region | R. C. Faucett |
| Philippine frogmouth | *Batrachostomus septimus* | FMNH 429205 | FMNH 429205 | PHILIPPINES; Luzon County, Kalinga 17.4417N, 121.0708E | --- |
| Australian owlet-nightjar | *Aegotheles cristatus* | ABTC 24643 | SAMA B39188 | AUSTRALIA; South Australia, 15 km from Lagoon Witt, Mabel Creek | --- |
| Moustached treeswift | *Hemiprocne mystacea* | WLK 385 | UWBM 68087 | SOLOMON ISLANDS; New Georgia Island, Arara, 8°29.5'S, 157°38.8'E (-8.492; 157.647) | C. E. Filardi |
| White-throated nightjar | *Eurostopodus mystacalis* | ANWC B40813 | ANWC B40813 | AUSTRALIA; Australian Capital Territory, Canberra, ~35°21'S, 149°058'E | --- |

Table 2.2. Number of loci, alignment length (in base pairs), and number of parsimony-informative sites in the full UCE data matrices.

| Completeness | Matrix | Loci | Alignment Length (bp) | Informative Sites |
|---|---|---|---|---|
| 100% | Untrimmed | 2,610 | 2,541,839 | 69,740 |
| | Trimmed | 2,610 | 1,477,319 | 43,571 |
| 75% | Untrimmed | 4,175 | 3,848,295 | 98,576 |
| | Trimmed | 4,174 | 2,219,207 | 62,496 |

Table 2.3. Number of loci, alignment length (in base pairs), and number of parsimony-informative sites in the UCE subset matrices.

| Loci | Replicate | Alignment Length (bp) | Informative Sites |
|---|---|---|---|
| **125** | 1 | 71,189 | 2,082 |
| | 2 | 72,700 | 2,003 |
| | 3 | 69,359 | 2,124 |
| **250** | 1 | 143,882 | 4,527 |
| | 2 | 138,773 | 3,900 |
| | 3 | 140,230 | 4,224 |
| **500** | 1 | 284,365 | 8,047 |
| | 2 | 281,767 | 8,228 |
| | 3 | 280,130 | 8,111 |
| **750** | 1 | 425,434 | 12,613 |
| | 2 | 428,598 | 12,685 |
| | 3 | 427,130 | 12,497 |
| **1000** | 1 | 572,481 | 16,812 |
| | 2 | 566,594 | 16,693 |
| | 3 | 568,380 | 16,663 |

## Figures

Figure 2.1. Maps of Central and South America depicting geographic distributions of potoo species. Range maps based on Cohn-Haft (2016). Collection localities of specimens used in this study are indicated by numbered arrows.

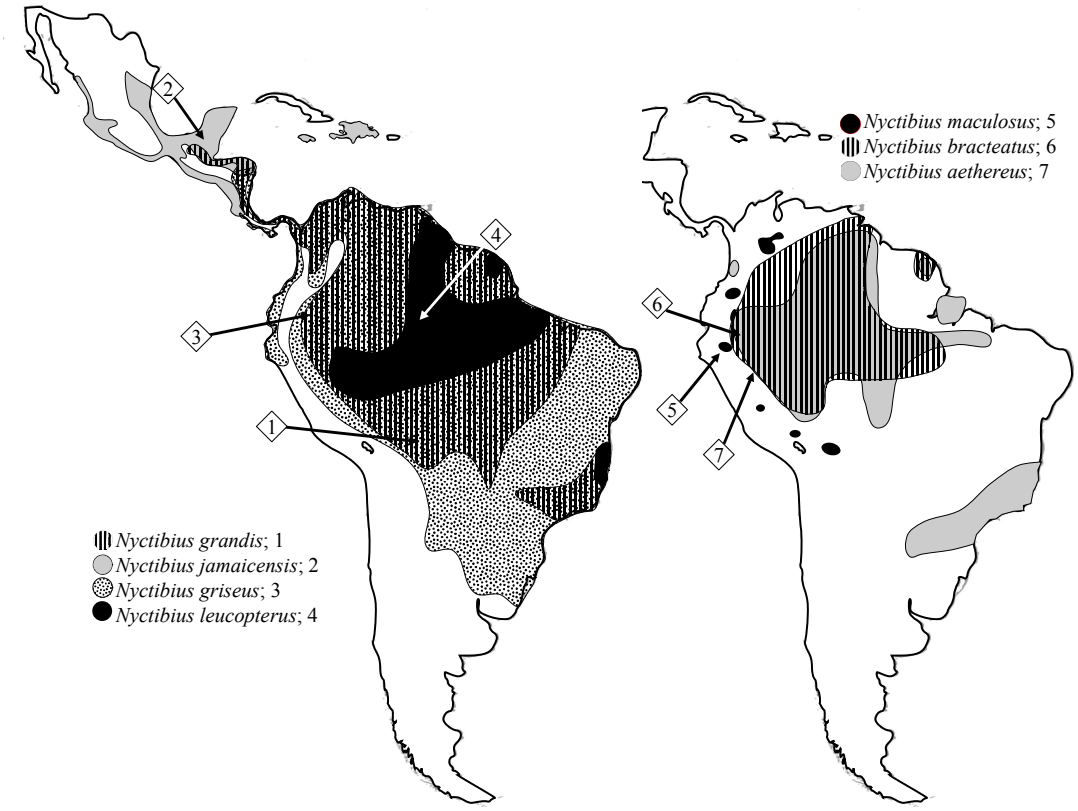


Figure 2.2. Cladograms of previous estimates of potoo phylogeny. Data used for each study is shown in the labels. MT-CYB = mitochondrial cytochrome b; MYC = cellular homolog of the myelocytomatosis viral oncogene. Where available, bootstrap support is depicted. The isozyme tree of Brumfield et al. (1997) is from Fig. 2b of that paper.

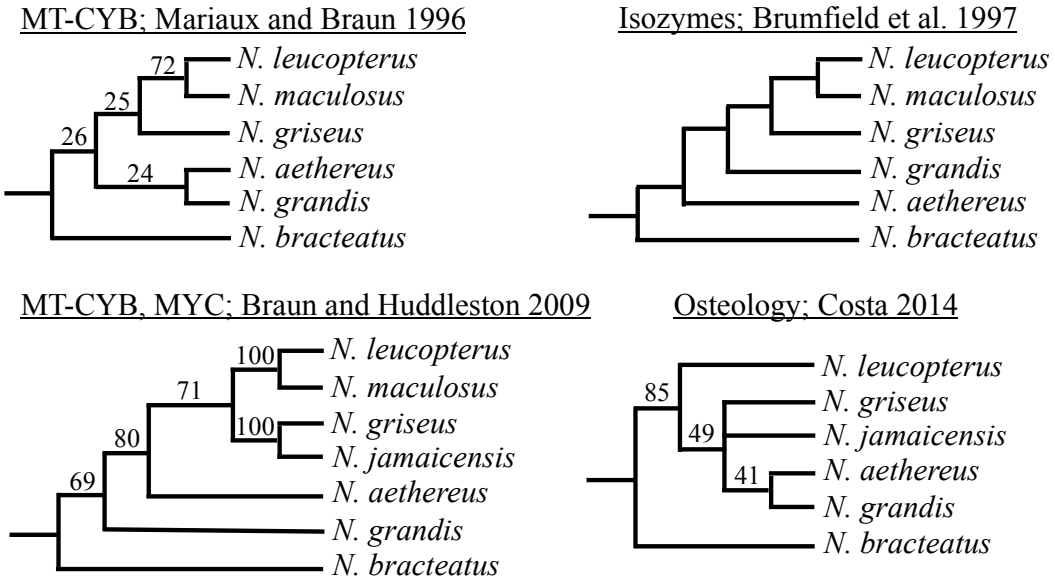


Figure 2.3. Maximum likelihood phylogram of potoo relationships from GARLI analyses of four UCE datasets (100% and 75% complete, trimmed and untrimmed). Bootstrap support values were 100% for all nodes in all analyses. Branch lengths are from the 100% complete, trimmed dataset. Outgroups not shown. Scale units are substitutions per site. Potoo illustrations represent closest geographic form to specimen used available in Cohn-Haft (2016), and are roughly scaled to size.

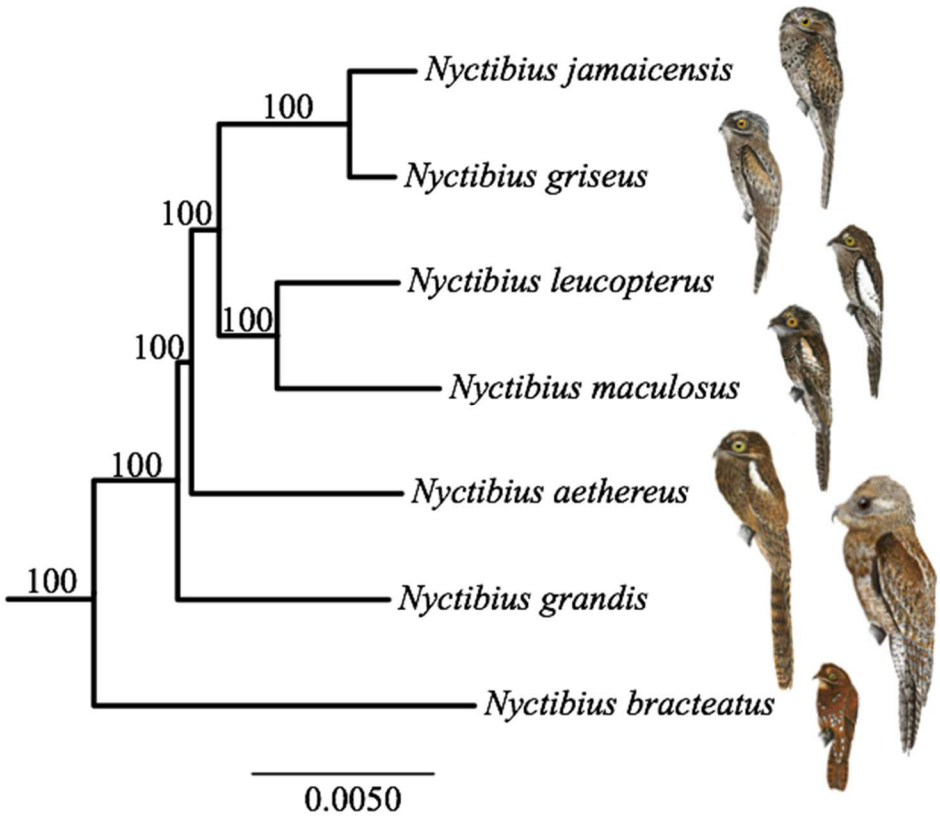


Figure 2.4. SVDquartets topology with bootstrap support from analyses of 4 datasets as follows: 100% complete, trimmed/100% complete, untrimmed/75% complete, trimmed/75% complete, untrimmed. Outgroups are collapsed.

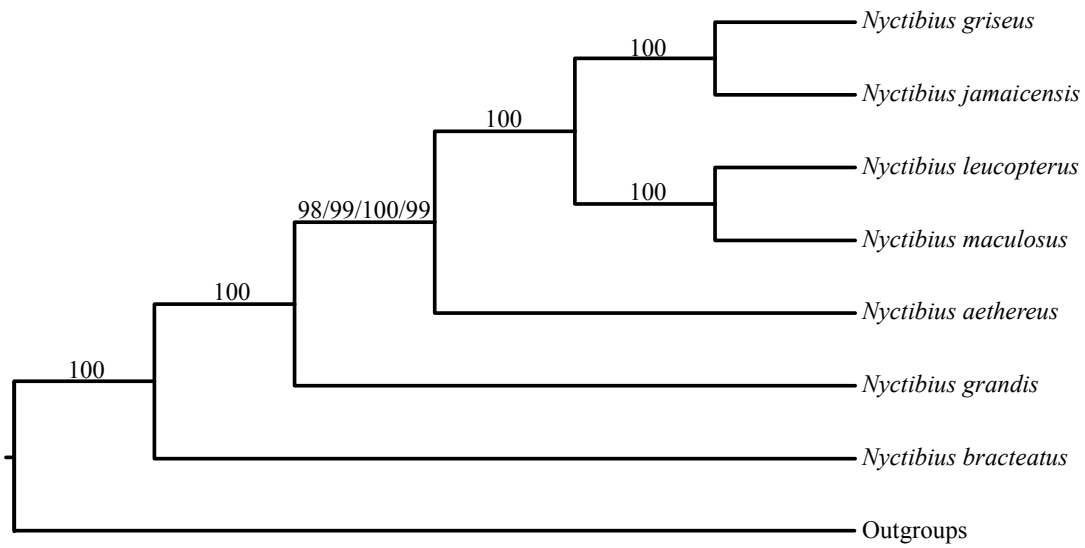


Figure 2.5. Resolving power of UCE data illustrated by matrix subset tests. Table includes the number of loci in each subset, the replicate number, and the RAxML and GARLI bootstrap support values for the node with an asterisk in topology A. All other nodes had bootstrap support of 100% in all cases. Where indicated, a replicate subset produced topology B or topology C. Bootstrap support values for those topologies are presented for RAxML above branches, and for GARLI below them.

| # of Loci | Replicate | RAxML | GARLI |
|---|---|---|---|
| **125** | 1 | Topology B | Topology B |
| | 2 | 73 | 73 |
| | 3 | 89 | 89 |
| **250** | 1 | 98 | 99 |
| | 2 | 98 | 100 |
| | 3 | 90 | 91 |
| **500** | 1 | 100 | 100 |
| | 2 | Topology C | Topology C |
| | 3 | 95 | 95 |
| **750** | 1 | 98 | 99 |
| | 2 | 100 | 100 |
| | 3 | 100 | 100 |
| **1000** | 1 | 100 | 99 |
| | 2 | 98 | 98 |
| | 3 | 100 | 100 |

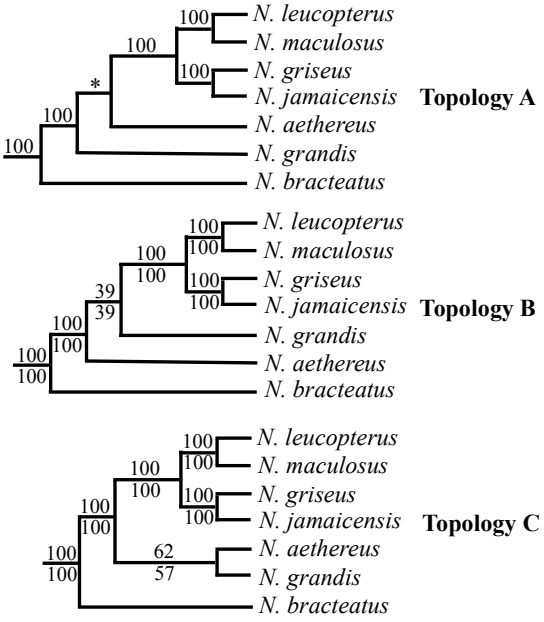
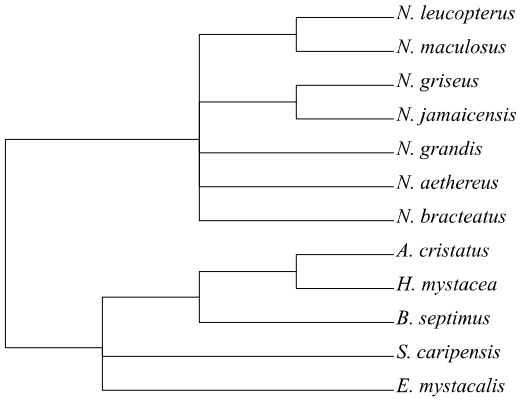


Figure 2.6. Results of the AU topology test. The constraint tree shown at left implies 315 plausible topologies which must be considered. The resulting likelihood scores (-lnL) are shown for the optimal tree and other trees that cannot be rejected by each of the four full datasets. Difference in likelihood (Δ -lnL) and p-values for suboptimal trees that cannot be rejected are also shown. All other trees had p-values below 0.05. All trees that cannot be rejected had ingroup topologies identical to Fig. 2.3 and 2.4, and differed only in the placement of outgroups *S. caripensis* and *E. mystacalis*.




| | Tree | -lnL | △ -lnL | p-value |
|---|---|---|---|---|
| **100% complete, trimmed** | 1 | 3,308,964.33 | (best) | |
| | 2 | 3,308,968.49 | 4.16 | 0.51 |
| **100% complete, untrimmed** | 1 | 5,842,639.33 | (best) | |
| | 2 | 5,842,775.34 | 136.01 | 0.06 |
| **75% complete, trimmed** | 2 | 4,875,405.89 | (best) | |
| | 1 | 4,875,416.96 | 11.08 | 0.55 |
| | 3 | 4,875,518.79 | 112.90 | 0.09 |
| **75% complete, untrimmed** | 1 | 8,633,856.67 | (best) | |
| | 2 | 8,633,983.25 | 126.58 | 0.19 |
| | 3 | 8,634,038.05 | 181.38 | 0.10 |

# Chapter 3: Phylogenomic Analysis of the Higher-Level Relationships of the Nightbirds and Relatives (*Strisores*)

**Abstract**

Among the considerations for using genome-scale, molecular sequence data in phylogenomic analysis are issues related to GC content, GC variance and their effects on model selection. In this study, I employed a variety of analytical techniques to empirically investigate those issues as well as biases and errors resulting from alignment trimming, taxon selection and matrix completeness in large datasets composed of ultraconserved elements (UCEs). I generated datasets of 2,289 – 4,243 UCE loci for 23 taxa representing the avian superorder *Strisores*. Extensive analyses revealed strong conflict within the data, and especially in regard to variation in GC content, that would not have been detected with more cursory analyses. Additionally, my results indicated that readily available models of molecular evolution are insufficient to encapsulate all phenomena present in genome-scale matrices. This problem may be at the root of current issues in phylogenomic analysis. *Strisores* includes both nocturnal and diurnal lineages, and with my current best estimate of the tree and knowledge of visual adaptations to nocturnality in these birds, I hypothesized that there were multiple independent origins of nocturnality in this group. The results of this study are relevant to phylogenomic analysis of any large, heterogeneous matrix.

## Introduction

Targeted sequence capture methods have enabled the efficient generation of genome-scale phylogenetic data matrices for dozens or hundreds of taxa (Glenn & Faircloth, 2016). However, the rate at which these data matrices are being generated is outpacing the development of software with which to analyze them. It has been well documented that use of currently available methods with genome-scale matrices can lead to systematic error—increased statistical confidence in the wrong answer with increased data points (reviewed in Kumar et al. 2012; Hahn & Nakhleh 2015; Hosner et al. 2016). Thus, phylogenomic analyses often produce trees with apparently strong statistical support at every node, but can these values always be trusted? The onus is on the researcher to design a rigorous set of analyses that will allow for accurate interpretation of support values, including looking for sources of topological bias or methodological error. Such errors are beginning to garner a lot of attention, which is crucial to the progression of the field (e.g. Kubatko & Degnan 2007; Philippe et al. 2011; Kumar et al. 2012; Salichos & Rokas 2014; Zwickl et al. 2014; Kocot et al. 2016; Suh 2016). In some cases, data type (e.g., coding vs. non-coding DNA sequence) may have a greater practical effect on phylogenetic inference than taxon sampling due to issues related to model selection (Reddy et al. 2017). However, best practices for analyzing a heterogeneous matrix (comprised of both coding and non-coding sequence) have not been established. Among the popular marker types that produce such matrices are ultraconserved elements.

Ultraconserved elements (UCEs) are attractive phylogenomic markers due to their ease of generation and flexibility at multiple evolutionary timescales (e.g.

(Crawford et al. 2012; Harvey et al. 2016; Lim & Braun 2016; Branstetter et al. 2017; White et al. 2017). UCE function is an active area of research in molecular genetics; they are believed to play a variety of roles. Empirical studies have so far identified UCEs with functions in development and regulation of transcription, as well as UCEs that have no effect when deleted from the mouse genome (Nóbrega et al. 2004; Calle-Mustienes et al. 2005; Woolfe et al. 2005; Pennacchio et al. 2006; Navratilova et al. 2009). UCEs, or regions of similar conservation, have been identified in a wide variety of taxonomic groups, including vertebrates, insects, yeasts, and plants (Lockton & Gaut, 2005; Siepel et al. 2005; Stephen et al. 2008; McCormack et al. 2012; Faircloth et al. 2015;), and hold great potential for the resolution of the phylogenies of these groups. However, examples of thorough phylogenomic analyses of UCE data at a variety of evolutionary timescales are lacking. The majority of published UCE studies lack a substantial investigation of the effect of methodological choices, such as trimming algorithm, or investigation of the source of phylogenetic support. Further necessitating these analyses, it has been shown that data type (e.g. coding, non-coding, UCEs, exons, etc.) can significantly influence tree topology (Reddy et al. 2017). Thus, investigations of the source of signal, and potential hidden error, in UCE datasets are in order.

The basis of a phylogenetic analysis is the alignment, in which hypotheses of homology are established and provide the groundwork for all further assumptions made in any analysis (Higgins & Lemey 2009). One of the greatest changes the advent of phylogenomics has brought is that researchers no longer visually inspect and manually edit alignments. With hundreds or thousands of loci in an individual

analysis, manual inspection isn't feasible, and thus I rely on automated methods of alignment processing and trimming.

Trimming is an important issue with sequence capture datasets. Untrimmed alignments of sequence capture data have ragged ends because random laboratory and sequencing effects result in a great variety of contig lengths across taxa. Trimming removes these ragged ends, which may be poorly aligned. Trimming can also remove internal regions that are misaligned or randomly aligned, which could influence phylogenetic analyses (Misof & Misof 2009). However, UCEs have high variation in information content among loci (Hosner et al. 2016; Meiklejohn et al. 2016), and that information is not randomly distributed. Most variable sites in a given locus lie not in the core UCE itself, but in the variable regions flanking the core that are captured as "by-catch" (Faircloth et al. 2012). Thus, judicious trimming is especially important with UCE data, because excessive trimming can reduce the number of informative sites in each locus. There are several potential trimming algorithms to choose from, which vary in a variety of parameters, including degree of subjectivity, whether or not they account for tree structure or use a substitution model. A comparative study of automated trimming software found that some algorithms lead to reduced accuracy in tree inference, highlighting the need for testing and comparing different trimming software for individual datasets (Tan et al. 2015).

Related to issues with alignment estimation and trimming is model selection. Phylogenomic datasets comprise so many individual loci that it is unreasonable to expect that they can be adequately analyzed under simple models of sequence evolution. This is especially true for UCEs which do not represent a single class of

biological elements with uniform function, and which represent both coding and non-coding sequence (e.g. Harmston et al. 2013; Pirnie et al. 2016; Warnefors et al. 2016). Partitioning the data, by selecting subsets of a data matrix that have evolved under different functional constraints and applying different models of sequence evolution to these subsets in a combined analysis, is a practical methodological choice (Bull et al. 1993; de Queiroz, 1993). However, partitioned analyses of UCEs are rare in the literature, largely because running an analysis of thousands of loci, subdivided into tens (or hundreds) of partitions, quickly becomes computationally intractable.

Here, I sought to conduct a thorough analysis of UCE data by using multiple analytical approaches that address known and potential issues in phylogenomic analysis, paying particular attention to issues that are especially pertinent to UCE data, including trimming, nucleotide composition and model selection. I utilized *a priori* methods of partitioning my data, focusing on GC content. This parameter was chosen as I observed great variation of GC content across individual UCE loci, as well as across taxa within a given alignment, and because GC content is an important component of classical models of sequence evolution. These models assume a constant GC content across the data, and it is now widely accepted that variation from this assumption can cause erroneous phylogenetic results, where taxa with similar base compositions are grouped in the tree, regardless of evolutionary history (e.g. Foster & Hickey, 1999; Springer et al. 1999; Griffiths et al. 2004; Delsuc et al. 2005; Jeffroy et al. 2006; Sheffield et al. 2009; Nabholz et al. 2011).

Lastly, I also investigated related issues of matrix completeness and taxon sampling. The results of this study are relevant and applicable to any genome-scale

dataset, whether those data result from target-capture methods, where thousands of loci are captured from hundreds of samples in one assay, or simply a very large, heterogeneous matrix.

*Background on Strisores*

I conducted these analytical inquiries with the ultimate goal of resolving the evolutionary history of the *Strisores*, a group of birds that comprises both diurnal and nocturnal members. It includes five lineages of nocturnal or crepuscular birds: the nightjars and nighthawks (Caprimulgidae), potoos (Nyctibiidae), frogmouths (Podargidae), owlet-nightjars (Aegothelidae), and the oilbird (Steatornithidae). These five lineages were previously placed in the order *Caprimulgiformes*, and collectively termed "nightbirds". They are an enigmatic, but much-storied group of birds. The most speciose group, nightjars, are also referred to as "goatsuckers," as legend has it that their wide mouths allowed them to suckle goat's milk at night. All nightbirds are characterized by their mottled or cryptic plumage and retiring daytime behavior, which aids camouflage in the group but confounds taxonomy.

Phylogenetic placement of the order of diurnal swifts and hummingbirds (previously *Apodiformes*) within the clade of entirely nocturnal or crepuscular nightbirds (previously *Caprimulgiformes*) ruffled some feathers in the ornithological community (Braun & Huddleston 2001; Mayr 2002). How could the swifts and hummingbirds—colorful, acrobatic fliers, clearly diurnally-adapted— have descended from the cryptically-colored, secretive nightbirds? Since the early work, several independent phylogenetic studies, based on both molecular and morphological data, have confirmed this topology (Mayr 2002; Barrowclough et al. 2006; Ericson et

al. 2006; Hackett et al, 2008; Braun & Huddleston 2009; 2010; Prum et al. 2015; Reddy et al. 2017). The group is now given the superordinal name *Strisores* (Mayr 2010), a title first used by the first curator and second secretary of the Smithsonian Institution, Spencer Fullerton Baird (Baird et al. 1858). *Strisores* are globally-distributed (Fig. 3.1), encompassing 602 currently recognized species, 470 of which are diurnal (Chantler 2017; Cleere 2017; Holyoak 2017a,b; Schuchmann & Bonan, 2017; Thomas 2017; Wells 2017). The oilbird and potoos are currently restricted to the Neotropics, while the frogmouths and owlet-nightjars are currently only found in Australasia (Cleere 1998; Peterson 2002; Hunt et al. 2009; Simpson et al. 2010; Cleere 2010), although the fossil record indicates that each of these groups was more widespread in the past (Mayr 2004; 2005b). Caprimulgids (nightjars and nighthawks) are found in tropical and temperate areas throughout the world, as are swifts (Fig. 3.1). Hummingbirds currently have a broad distribution restricted to the New World, but appear in the fossil record of Europe as well (Mayr 2004).

The inclusion of diurnal and nocturnal lineages within *Strisores* raises questions about the evolution of nocturnality in these birds, especially how many times it occurred, what adaptations made it possible, and what genetic and molecular variation underlies those adaptations. A well resolved phylogeny, would help to answer these questions, but previous molecular estimates of the *Strisores* tree have failed to reproducibly place the major lineages, with the exception of pairing the diurnal *Apodiformes* with the nocturnal owlet-nightjars (Fig. 3.2). The datasets used in these studies have progressed from a few genes (Ericson et al. 2006; Braun & Huddleston, 2009) to tens of genes (Hackett et al. 2008; Reddy et al. 2017) to

hundreds of genes (Prum et al. 2015), but have failed to reproducibly resolve the earliest branches in the phylogeny, especially the position of oilbird and potoos.

In all of these studies, the oilbird (*Steatornis caripensis*) and potoos were either sister taxa or their position was unresolved (Fig. 3.2). A Bayesian analysis of five nuclear genes resulted in an unresolved topology after collapsing all nodes with less than 95% posterior probability (Ericson et al. 2006). Braun and Huddleston (2009) found that sequence from mitochondrial cytochrome b resulted in an unresolved topology, but a nuclear marker sequence (cellular homolog of the myelocytomatosis viral oncogene) recovered oilbird and potoos as sister lineages (oilbird+potoos) with 55% bootstrap support (combined analysis was unresolved Braun & Huddleston, 2009). In studies that found the oilbird+potoos topology, bootstrap support varied from less than 50% to 77% (Hackett et al. 2008; Prum et al. 2015; Reddy et al. 2017). The Hackett et al. (2008) study utilized 28 genes, while Prum et al. (2015) used 259 genes and Reddy et al. (2017) used 54 genes.

Morphological results are also unresolved, and independent studies disagree both in results of their analysis, and the interpretation of characters by different authors. For example, there is disagreement between Cracraft (1981) and Mayr (2002) on whether the caudal margin of the sternum in Nyctibiidae bears a single lateral process or two pairs of incisions. Analysis of morphological data alone recovered a sister clade comprised of caprimulgids and potoos in several independent studies (Cracraft 1981; 1988; Mayr 2002; 2003; 2010; Nesbitt et al. 2011). However, most studies of molecular or combined morphological+molecular datasets have not recovered that group (Mayr 2003; Ericson et al. 2006; Hackett et al. 2008; Braun &

Huddleston, 2009; Prum et al. 2015; White et al. 2016; Reddy et al. 2017; White et al. 2017), the one exception being the morphological+molecular analysis in Nesbitt et al. (2011).

In contrast, a potential synapomorphy was identified by Hoff (1966) as unique to potoos and oilbird—the medial serratus superficialis, a sheet of muscle from the last cervical rib to the first through fourth thoracic ribs and their uncinated processes. The anterior and posterior muscle sheets are fused in potoos and oilbird, rather than separate or intermediate, as in other nightbirds (Hoff 1966). Indeed, common to all of the studies listed in this section is difficulty with placing the oilbird. Its phylogenetic position varies not only between datasets and studies, but shifts with regard to data type (e.g. see morphological, molecular, and combined results in Mayr 2003). The oilbird, an ancient monotypic lineage, is a frugivorous, echo-locating cave-dweller, unlike any other nightbird. It displays numerous morphological adaptations to this lifestyle, many apomorphic traits, and does not share many characters that are common to the other nightbird families (Cracraft 1988).

The most recent molecular estimates of *Strisores* topology, Prum et al. (2015) and Reddy et al. (2017) both recover a clade of oilbird and potoos, but with relatively low support, and differ in the placement of the caprimulgids. The Reddy et al. (2017) analysis demonstrated that data type has a strong effect on tree inference. The Prum et al. (2015) dataset is dominated by exons, while the Reddy et al. (2017) data is mostly intronic sequence. Incongruence between these two analyses indicates conflicting signal and warrants further investigation. A statistically sound phylogenetic tree will provide the groundwork necessary to elucidate the evolutionary origins of

nocturnality in the *Strisores*, and allow mapping of molecular adaptations that might have facilitated the transition(s) between nocturnality and diurnality in this clade.

## Materials & Methods

### *UCE Functional Characterization*

The UCE probe sequences used in this study are those targeting tetrapod/amniote species, described by Faircloth et al. (2012) and available from <ultraconserved.org>. They include 5,472 probes targeting 5,060 UCE loci. To investigate the potentially disparate evolutionary constraints on these loci, I sought to identify where UCEs overlapped with existing functional annotations in the chicken (*Gallus gallus*) and human genomes. The chicken results are most relevant to this set of UCEs, but I included human in order to take advantage of its relative wealth of annotations. The most recent assemblies of the chicken (galGal5; Hillier et al. 2004) and human (hg38; Lander et al. 2001) genomes were downloaded from the UCSC genome browser (Kent et al. 2002). The UCE probe sequences were mapped to each genome, using Bowtie2 (v. 2.3.0; Langmead & Salzberg, 2012), aligning 'end-to-end' and without an upper limit on number of matches ('-a'). This mapping generated BED-format genome coordinates for each UCE probe sequence, and I kept those that mapped to a unique place in each genome. Annotation files for CpG islands and RefSeq 3' untranslated regions (UTRs), 5' UTRs, protein-coding sequence (CDS), exons and introns, were downloaded from the University of California, Santa Cruz genome table browser for each genome, also in BED format (Rosenbloom et al. 2015). Exons include UTRs, whereas CDS represents only the translated portions of exons. Using the 'intersect' function of the software suite bedtools (v. 2.18.1; Quinlan

& Hall, 2010), I identified where coordinates of my UCE probes intersected coordinates of an annotation in either genome.

To test whether UCE probes intersected functional annotations more frequently than expected by chance, I created "simulated UCEs" by randomly selecting short sequences from each genome that are the same length as my probes (120 bp). I used the length of every contig within a genome relative to the total genome length to generate a probability that a simulated UCE would be located on that contig, and incorporated that information into the random selection of coordinates. This method is implemented in a custom Perl script `Locus_simulator.pl` (DOI:10.5281/zenodo.832352), and was repeated 100 times. These 100 "simulated" region sets (each comprised of 5,472 loci) were intersected with the annotation files as described above, and the results presented are an average over those 100 repetitions. Statistical significance was tested by calculating a binomial probability based on the simulated intersection results.

*Taxon Sampling*

Taxon sampling for this study included two members of each major lineage of the *Strisores*, except for the oilbird, which is a monotypic family, and the potoos, for which I included all seven currently recognized species (Table 3.1). Three unambiguous outgroups were chosen for a total of 23 species (specimen information in Supplementary Table S1). Cleaned, aligned UCE data were available for potoos from a previous publication (White et al. 2017), described here in Chapter 2. For all other species, frozen tissue samples were assembled through my own fieldwork and loans from major museum collections. DNA was extracted using a phenol-chloroform

protocol (Rosel & Block, 1996), and quality was assessed by agarose gel electrophoresis and fluorometric quantification.

*Data Collection and Alignment*

UCE data were collected following the laboratory protocols described in Faircloth et al. (2012), and available at <ultraconserved.org>. Briefly, DNA samples were sheared to 200-500 bp in length via sonication, and Illumina sequencing libraries were prepared using KAPA Biosystems kits (KAPA Biosystems, Inc.). Unique, single-index adapters were ligated to each sample. Libraries were enriched for the 5,472 probes targeting 5,060 UCE loci contained in the Tetrapod/Amniote probe set (commercially available from MYcroarray), using chicken Cot-1 DNA as Block #1 (Applied Genetics Laboratories, Inc.). Hybridization was carried out at 65°C for 24 hours, followed by amplification and preparation for sequencing.

Enrichment success was verified by qPCR. Custom primers targeting five individual UCE loci were used to quantify enrichment of those loci by comparing enriched versus unenriched libraries. Enrichment values of 100-fold or over were considered successful. Enriched libraries were pooled and submitted for paired-end, 100 base pair (bp) sequencing on Illumina platforms (HiScan and HiSeq2000). On average, ~2 million reads were recovered per species (600,814-6,338,255; Supplementary Table S2). I used the PHYLUCE pipeline (Faircloth 2015b) to process this data, beginning with quality control and trimming via Trimmomatic (Bolger et al. 2014). Reads were assembled *de novo* using Velvet (ver. 1.2.09; Zerbino & Birney, 2008), and contigs matched to UCE probe sequences using LASTZ (Harris 2007). Contigs without a UCE match were discarded. Alignments were conducted separately

on each set of contigs that included more than three taxa matching a single UCE locus using SATé-II based on MAFFT (ver. 2.2.7; Liu et al. 2011).

*Matrix Completeness and Trimming*

By 'matrix completeness' I refer to the proportion of taxa for which a given locus is represented. In this study, I tested two levels of matrix completeness: 100% complete, for which loci are included only if data is present for every taxon, and 75% complete, for which loci are included if at least 75% of the taxa are represented by data (Table 3.2). At 75% completeness, I found that my matrices pushed computational tractability in terms of memory and time required to run analyses on my supercomputing cluster. An independent avian UCE study found 25% missing data to be the point of diminishing returns in terms of resulting statistical support (Hosner et al. 2016). With these two categories of completeness, four trimming treatments (described below) and the four taxon samplings detailed in Table 3.1, I had 32 major concatenated matrices to analyze. Lastly, matrices containing only those loci that intersected with my different functional annotation categories were created for the complete matrices of taxon sets 1 and 4.

In this study, I analyze untrimmed UCE alignments along with those trimmed with algorithms available in the PHYLUCE and Aliscore software (v2.0). Resulting datasets using the untrimmed alignments varied from 2,289 – 4,243 loci, and were ~2.4 – 4.1 M bp in length (Table 3.2).

I ran the PHYLUCE trimming algorithm with default settings. The alignment was viewed in 20 bp sliding windows, removing individual columns with less than 65% of taxa present. Additionally, whole windows were removed if greater than 20%

divergence was observed between the consensus and any row of the alignment. Lastly, any loci shorter than 100 bp after trimming were excluded. Resulting datasets using the PHYLUCE alignments varied from 2,289 – 4,241 loci, and were ~ 1.3 – 2.3 Mbp in length.

The Aliscore algorithm does not implement any user-defined cutoffs for data exclusion, and as such is considered a relatively objective trimming method. With Aliscore, columns were removed that contained sites with a signal that could not be differentiated from random noise (Kück & Meusemann, 2010; Misof & Misof, 2009). I ran Aliscore with the default parameters (sliding window of 6 bp, 4 x N random pairs checked), as well as with treatment of gaps as ambiguous characters and subsequent removal (`-N`). In both cases, alignments were masked using Aliscore, and the masking trimmed using Alicut. Resulting datasets using the Aliscore alignments varied from 2,289 – 4,242 loci, and were ~1.0 – 1.9 Mbp in length.

*Data Characterization*

I characterized various aspects of my individual loci (and subsequent concatenated datasets) in order to understand how best to design analyses and interpret results. I calculated GC content, GC variance (variance in GC content among taxa in an alignment), parsimony informative sites, and presence of block-shift alignment errors. I then looked for evidence of potential bias in number of loci, number of nucleotides, or GC content for individual taxa.

GC variance was calculated using a custom Perl script `GC_variance.pl` (DOI:10.5281/zenodo.582627) as the mean squared difference of an individual taxon's GC content from the average GC content for a given locus. Block-shift

alignment errors were defined as a string of mismatched base pairs present in only one taxon in a gap-free region of an alignment, and considered to represent alignment error. Some may actually represent small inversions, but those are essentially equivalent to alignment errors for the analyses conducted herein. These were identified using the `findAndMaskBlockshifts.py` script in the PhYloGenOmic Tools package (Zwickl et al. 2014).

To test for methodological bias on tree inference, I used the quartet summary approach implemented in PhYloGenOmic Tools and described by Zwickl et al. (2014). This method queries two sets of individual locus bootstrap trees for the resolution of a given quartet (rooted triplet). Each set of trees is derived from the same data and loci, but under different treatments (e.g., trimming algorithm). PhYloGenOmic Tools then plots cumulative support and flux diagrams that facilitate visualization of the effects of the treatments on the topology of the quartet. A high degree of shift from one triplet resolution to another between the two treatments indicates methodological bias, whereas high flux without obvious directionality indicates general sensitivity to treatment effects and random gene tree error. I examined 13 sets of quartets selected to test the effect of trimming treatment, matrix completeness, and taxon sampling (interchanging the individual species representing each major lineage), on the position of oilbirds, potoos and frogmouths, including one outgroup (Supplementary Fig. S1).

*Model Selection and Topology Searches*

Model selection was conducted in PAUP* (ver. 4.0a151) (Swofford 2003). The most appropriate model was selected according to the corrected Akaike

information criterion (AICc) for all of my concatenated datasets, as well as for individual subsets when partitioning was implemented (Akaike 1973; Hurvich & Tsai, 1989). In addition, all 32 major matrices were RY coded by purine/pyrimidine and analyzed under the GTRGAMMAI model in RAxML.

Maximum likelihood (ML) searches were conducted on concatenated matrices using GARLI (ver. 2.1) (Zwickl 2006) and/or RAxML (ver. 8.2.9; Stamatakis 2014). RAxML is a faster likelihood program, but implements a limited number of evolutionary models. Because of this, all subsequent analyses for which GTR+I+G is the most appropriate model were run with RAxML, while analyses requiring a different model, and all partitioned analyses, were run with GARLI. In this and other studies, I have found the topological results of analysis between GARLI and RAxML to be identical.

Tree searches in RAxML consisted of 20 replicates, with bootstraps run using the bootstopping criterion (option – 'autoMRE'). Trees searches for individual UCE loci used the rapid bootstrapping algorithm, which conducts both a tree search and bootstrapping (`-f a`). Support for individual locus trees was assessed with 100 bootstrap replicates, with the exception of the trees used in the PhYloGenOmic Tools quartet summary analysis, which were run for 200 replicates. Rogue taxon searches were conducted in RAxML, using the bootstrap files produced from the bootstopping criterion search as input ('MR_DROP').

GARLI runs were conducted using 100 independent runs of two search replicates each ('searchreps 2'). The 'treedist' function of PAUP* (ver. 4.0a151) was used to ensure that the same topology was found for the best replicate in each of the

100 runs (if not, searches were re-run with increased searchreps). This method assesses the thoroughness of the search of tree space, and is described in White et al. (2016). One hundred non-parametric bootstrap replicates were conducted, with one search replicate each.

For this study, I elected to use ML analysis and forego Bayesian runs due to the computational demands (both in resources and time) of the study, and difficulty of accurately assessing MCMC convergence in Bayesian analyses. The number of permutations of the dataset analyzed in this study would be impractical to repeat in a Bayesian framework, and the topological results are likely to be similar with the two methods. Instead, I chose to analyze my data utilizing a method that implements the coalescent algorithm—a fundamentally different analytical approach to either ML or Bayesian analysis.

For my 32 major matrices, coalescent-based species tree analyses were run using SVDquartets (Chifman & Kubatko, 2014; 2015) as implemented in PAUP* (ver. 4.0a151). SVDquartets implements the coalescent algorithm, without using a Bayesian framework (Chifman & Kubatko, 2014). It thus uses all of the information in a given datamatrix, but in a time- and computationally-efficient manner. Other popular approaches to species tree analysis either employ summary statistics, which do not encapsulate all variation in the data, or Bayesian frameworks, which are demanding to run on large datasets such as ours (see discussion in Chifman & Kubatko, 2014). Another popular approach, Accurate Species Tree ALgorithm (ASTRAL), does not implement the coalescent algorithm, and is instead a gene tree reconciliation approach (Mirarab et al. 2014). SVDquartets analyses were run

evaluating all possible quartets and implementing 100 non-parametric bootstrap replicates.

In all cases except for the RAxML rapid bootstrap runs, bootstrap values were plotted on the optimal topology using the SumTrees program in the python library DendroPy (ver. 4.2.0) (Sukumaran & Holder, 2010). Taxon jackknifing was carried out by creating datasets in which each has a different (sole) taxon removed.

*Other Topological Analyses*

To ask whether the likelihood scores of the ML trees found in my searches were significantly better than other plausible trees, I used the approximately unbiased (AU) test of Shimodaira (2002). I constructed a separate set of plausible trees for each of my taxon samplings using constraints (Supplementary Fig. S2) that incorporated *a priori* information about the phylogeny of *Strisores*. First, all major lineages of *Strisores* were constrained to be monophyletic. Next, I constrained owlet-nightjars (Aegothelidae) to be sister to the *Apodiformes*. Support for this grouping has been shown in multiple independent analyses, including molecular and morphological data, and is widely accepted (Mary 2002; Ericson et al. 2006; Hackett et al. 2008; Braun & Huddleston, 2009; Mayr 2010). For datasets 3 and 4, which include most or all of the potoo species, I constrained the topology within Nyctibiidae to that of Chapter 2 (White et al. 2017), which resolved the potoo phylogeny unambiguously. Lastly, I constrained the topology of my outgroups, making a sister clade of *Gallus gallus* and *Anseranas semipalmata*. In every case, these constraints resulted in a candidate set of 105 topologies. The AU tests were run in PAUP* using resampling of estimated log likelihoods to generate 1,000 bootstraps.

Phylosort was used to query pools of rooted individual locus trees for given taxon groups (Moustafa & Bhattacharya, 2008). Indel number and length were calculated using the simple coding algorithm implemented in 2matrix (Salinas & Little, 2014) and the output was processed using a custom Perl script `Indel_stats.pl` (DOI:10.5281/zenodo.438047). I used the 'DescribeTree' function in PAUP* to score indel characters on the matrix generated by 2matrix for three candidate topologies. A custom Perl script was used to process the output (`PAUP_Indel_parser.pl`; DOI:10.5281/zenodo.832357). Missing, ambiguous or uninformative indels were excluded, and only indels with a consistency index of 1 and unambiguous character state change (i.e. "==>" rather than "-->") were used.

For comparison with my data, I downloaded the published data of Prum et al. (2015; doi:10.5281/zenodo.28343; trimmed, concatenated matrix 'Concatenated.phy') and Reddy et al. (2017; doi:10.5061/dryad.6536v; 'Reddy_sup_fileS5_ALLtaxset.nex'). Any taxa not within *Strisores* were pruned from the aligned matrix, except for the three outgroup species used in this study, which were present in both datasets. Where both published genome data and new sequences generated by that study were available in the Reddy et al. (2017) data, both were kept. The Prum et al. (2015) taxon name *Eurostopodus macrotis* was changed to *Lyncornis macrotis* to reflect the taxonomic update of Han et al. (2010). After the removal of any columns that were solely gaps, this resulted in a 71,075 bp matrix with 24 taxa for the Reddy et al. data, and a 394,684 bp matrix with 16 taxa for the Prum et al. data. Data were RY-coded, and analyzed in RAxML under GTR+I+G as

above, with the exception that 200 non-parametric bootstraps were run. These are hereafter referred to as the *Strisores* matrices from each respective publication.

## Results

### *Characterization of UCE probes and datasets*

Of my 5,472 probe sequences, all but one mapped to the chicken genome (galGal5 assembly; overall alignment rate 99.98%) and 5,222 mapped uniquely (Table 3.3). In the human genome (hg38 assembly), 1,571 sequences did not map (overall alignment rate 71.29%), and 3,810 mapped uniquely. Within the chicken and human genomes, UCE probes were found to intersect 5' UTRs twice as often as simulated datasets, and 3' UTRs, exons and CDS three times as often as simulated datasets (Table 3.3). All three results were statistically significant ($P > 0.05$), potentially indicating an important functional role for UCEs, either as encoding functional elements themselves (exons, CDS), or in the regulation of expression (UTRs). UCE probes were found to intersect introns significantly more often than simulated datasets, but barely so (1.1X), suggesting that this result may not be biologically significant. Additionally, UCE probes were found to intersect CpG islands significantly less often than the simulated datasets (0.5X), possibly reflecting the high AT content of most UCE loci.

The UCE probes themselves have an average GC content of 37.8%, while overall GC content of the datasets was about 39% (Table 3.2). The GC content of *Strisores* matrices for the Prum et al. and Reddy et al. datasets were 46.5% and 44.5%, respectively. The genome-wide GC content of *Gallus gallus, Columba livia, Anstrostomus carolinensis* (a caprimulgid) and *Chaetura pelagica* (a swift) range

between 40.8% - 42.8% (average 41.6%). Thus, the UCE datasets presented here are closer to the genome-wide GC content of these taxa than either the Prum et al. or Reddy et al. datasets, which are biased toward exons and introns, respectively.

GC content showed great variation among loci within all UCE datasets (24.9-71.2%; Fig. 3.3a), but overall GC content did not vary significantly between datasets (38.8-39.1%; Table 3.2), nor did it appear to be affected by trimming algorithm or matrix completeness (Fig. 3.3b). In contrast, GC variance among taxa was reduced by trimming, and appears slightly lower in datasets trimmed using Aliscore than those using PHYLUCE (Fig. 3.3c).

GC content varied little among taxa across all sites of concatenated matrices, but showed greater differentiation when only considering parsimony informative sites (Fig. 3.4). At informative sites, GC content appears elevated in all of the potoos except for *Nyctibius jamaicensis*, relative to the other taxa. Oilbird shares this anomalously high GC content with potoos.

There was little variation (7-8.8%) in the percentage of parsimony informative sites across datasets (Table 3.2), but considerably more across individual loci (0-40%; Fig. 3.5). I found evidence of block-shift alignment errors in the datasets analyzed, but they constituted a relatively small proportion of the overall alignments. Calculated as the number of individual bases masked (replaced with `N`), they totaled 2% of bases in D1CP, 9% in D1CU, and 4% in D4CA. Trees were calculated for block-shift masked (alignment errors replaced with `N`s) versions of datasets D1CP, D1CU and D4CA. Relative to the concatenated ML topology, the result was the same for D1CU and D4CA while D1CP varied slightly.

Looking at representation of loci for a given taxon across incomplete datasets, I found that *Batrachostomus septimus* and *Nyctibius bracteatus* had, on average, 400-500 fewer loci than other taxa (Supplementary Table S2). When considering representation at a nucleotide level (calculated as the percentage of total alignment length for which a given taxon has nucleotides present), I observe reduced representation for taxa *Aegotheles cristatus* and *Nyctibius bracteatus* for both complete and incomplete matrices (Supplementary Table S3). *Batrachostomus septimus* had reduced representation for incomplete but not complete matrices.

Quartet summary analyses were conducted using the loci of taxon sets 1 and 4 to examine the effects of trimming treatments on topology. Trimming, or not, did not appear to bias preference of topology in any of the 13 quartets tested. In all quartets tested that included one potoo, oilbird, and two outgroups, I saw flux from every topology to the unresolved topology, indicating stochastic error, rather than methodological error (Supplementary Fig. S1).

*Phylogeny*

Preliminary analyses of taxon set 1 revealed uncertainty with placing the oilbird with respect to potoos within the topology. Because of this, I added all known species of potoos to my analyses in a total of four different taxon samplings to test the effect that including different taxa would have on the topology (Table 3.1). Potoos are an unusual and ancient major lineage of the *Strisores*, and in particular, I wanted to test the effect of including or excluding the rufous potoo (*Nyctibius bracteatus*). It is the oldest and most divergent potoo species (Mariaux & Braun, 1996; Brumfield et al. 1997; White et al. 2017), and coincidentally was missing data for a larger number of

loci than all but one other taxon (Table 3.6). The oilbird is also highly divergent, but it is a monotypic lineage, so no additional taxon sampling was possible for it.

Phylogenetic trees found by my ML analyses resulted in only three unique topologies, which vary solely in the position of oilbird relative to potoos (Fig. 3.6). For clarity, I refer to the topology in which oilbird and potoos are sister lineages as topology A, the topology in which oilbird is placed basal to the potoos+frogmouths+owlet-nightjars+*Apodiformes* as topology B, and the topology in which potoos are placed basal to oilbird+frogmouths+owlet-nightjars+*Apodiformes* as topology C. All other relationships in the tree were very highly supported (99-100% bootstrap support) in ML analysis for all matrices of all taxon sets. The most appropriate model selected for all datasets and subsets was either a general-time reversible model, or the transversional model, both with estimated proportion of invariant sites and gamma-distributed rate variation among sites (GTR+I+G or TVM+I+G, respectively; Table 3.4).

ML searches yielded topology A on 18 of 32 data sets analyzed with bootstrap support for the critical node ranging from 50-100% (Table 3.4). However, four complete matrices yielded topology B and two incomplete matrices yielded topology C with generally lower bootstrap support ranging from 49-78%. The AU test was only significant in four instances, all of which favored topology A (Table 3.4). RAxML did not identify any rogue taxa for any of the datasets used in this study. This is not surprising as the overall topology is stable, and the oilbird and potoos do not appear in random positions throughout the *Strisores* tree, but only move in reference to each other.

To test the potential effect of the relatively high GC content observed in both oilbird and potoos, I ran ML analyses with nucleotides recoded as either purine or pyrimidine (RY). These analyses yielded only topologies B or C, with bootstrap support for the critical node ranging from 37-95%, in sharp contrast to the overwhelming predominance of topology A from the nucleotide analyses (Table 3.4). Average bootstrap support for the critical nodes across the 32 datasets was nearly identical between the RY and nucleotide analyses. SVDquartets analysis of the 32 major matrices resulted in five unique topologies (Fig. 3.7; Table 3.4). Interestingly, the only topology shared with the ML analyses was topology A, and of the remaining four unique topologies, three represent re-rootings of topology A.

The wide range of GC content and GC variance among loci (Fig. 3.3) means that loci with extreme values for these parameters will not be well fit by ML models that assume a single ratio of nucleotide composition. I addressed this problem in several ways. Partitioned analyses were designed based on the results of my data characterizations. They included partitioning schemes that separated GC content into three categories, GC content below 34%, GC content 34-42% and GC content above 42%, as well as datasets that removed the top 25% of loci with highest GC variance. ML analyses of datasets partitioned into three categories of GC content were run for all matrices which did not initially recover topology A—D1CP, D1IP, D2IP, D4CP, D4CA and D4CAN, as well as D1CA and D1CU. This three-partition scheme recovered topology A in all cases except for D4CA, which yielded topology B (data not shown). Further partitioning dataset D4CA into eight GC categories still resulted in recovery of topology B. Analyses of datasets for which loci with the highest

amount of GC variance were removed resulted mostly in topology A, and in most cases recovered the same topology as the unpartitioned ML analysis (Table 3.4).

ML analysis of my functional annotation category matrices derived from taxon sets 1 and 4 revealed conflict similar to that in my concatenated ML analyses, without clear preference for any one topology, or consistent results for datasets 1 and 4 (Table 3.5). The majority of analyses resulted in a tree not present in the concatenated ML or SVD analyses Some of these categories contained relatively few loci (e.g., 5'UTRs—15 loci), but others were quite large (e.g., introns—469 loci; Table 3.3).

*Other Topological Analyses*

For a given matrix, the majority of taxon-jackknifed datasets resulted in the same topology as the concatenated ML analysis (Fig. 3.8). D4CU was the only dataset that recovered the same topology regardless of which taxon was dropped. Most changes in topology happen when dropping *Nyctibius grandis* or an earlier branching taxon. There were no differences between the Aliscore trimmed datasets for taxon set 1, and few for taxon set 4. Any trimming caused the jackknifed datasets of taxon set 4 to prefer topology B, but this pattern was not true for taxon set 1, the majority of which preferred topology A, except for the PHYLUCE-trimmed datasets. All nodes in all trees resulting from jackknifed datasets were 100% supported when oilbird was dropped.

I searched individual locus gene trees for presence of monophyletic groups resembling my three major ML topologies (Table 3.6). The taxon groups searched were: topology A (oilbird, potoos), topology B (potoos, other ingroups) and topology

C (oilbird, other ingroups), where the clade "other ingroups" included swifts, hummingbirds, owlet-nightjars and frogmouths. For reference, I also tabulated the number of single locus gene trees that supported a well-established node, the monophyly of potoos. The majority of single locus trees supported topology A.

I also looked for instances where a given locus was present in two datasets that resulted in conflicting topologies. I found a total of 6 loci which were shared between datasets 1 and 4 that supported topology A in one, and topology B in the other. Excluding these six loci from matrices D1CP and D4CA did not change the phylogenetic result, where D4CAN switched to topology A. This is a small number of loci and seems to indicate, as I saw with the quartet summary analyses, that trimming algorithm is not biasing one topology over any other.

The gap coding algorithm identified no indels supporting topology B in either dataset. There were 5 indels in dataset 1 and 4 indels in dataset 4 supporting topology A, as well as 3 indels in dataset 1 and 2 indels in dataset 4 supporting topology C. I inspected each of these indels manually, disregarding putative indels in regions of ambiguous alignment; e.g. where homopolymer runs created variable sizes of gaps without clear taxonomic definition. I judged that there were 4 indels supporting topology A and 2 supporting topology C, and that these were present in both datasets 1 and 4 (Supplementary Fig. S3).

## Discussion

### *Characterization of UCE Data*

In the original description of UCEs, Bejerano et al. (2004a) suggested that these elements are involved in RNA processing (where they overlapped known

mRNAs), or regulation of transcription or development (where they were intronic or intergenic). Further analysis indicated that UCEs are "a heterogeneous set of clusters of a variety of classes," (Bejerano et al. 2004b). It is clear that UCEs are not one type of biological element that serve one function, but their unusual degree of conservation across evolutionarily divergent taxa indicates that they serve important purposes, such as regulation of development (Pennisi 2017). UCEs appear to be enriched in UTRs, which are useful phylogenetic markers (Bonilla et al. 2010; Harshman et al. 2003). My efforts to identify UCEs that overlap a known functional element revealed that 66% of UCEs lie outside of annotated elements in the chicken genome (30% in human). This makes UCEs a challenge for phylogenetics, as many commonly used models of sequence evolution were designed based on knowledge of functional DNA sequence. Development of models that more closely represent the evolutionary constraints on non-coding data are needed to take full advantage of the phylogenetic information in UCEs.

In this study, I produced a dataset of 4,243 UCE loci for resolution of the higher-level relationships in *Strisores*. Concatenated matrices derived from this dataset range from 1-4 Mbp in length, representing a sizeable increase in data from previous estimates (the full Reddy et al. matrix is 137,463 bp in length and the full Prum et al. matrix is 394,684 bp). These matrices are heterogeneous in that they contain both coding and non-coding sequence, where previous matrices contained mostly one or the other.

*Phylogenomic Analysis of UCEs*

The UCE data firmly resolve the position of all major lineages in the *Strisores* phylogeny aside from the position of oilbird with respect to potoos. It is clear from the analyses presented here that that polytomy is not resolved unambiguously with the data and methods at hand. Nevertheless, I believe that the best currently available estimate may come from ML analysis of the RY-coded dataset D4IA. This analysis prefers topology B, with the lineage leading to oilbird branching before a clade including potoos+frogmouths+owlet-nightjars+*Apodiformes* (Fig. 3.9). I believe dataset D4IA is likely to yield the best estimate, as it is one of my largest matrices, in terms of loci, length and informative sites. The phylogram presented in Figure 3.9 also clearly illustrates why resolving the deep phylogeny of the *Strisores* has been difficult—all internodes in this region of the tree are very short. This feature of the tree suggests that there is little information in the UCE data that would allow one to prefer one topology over another. Because the UCE dataset samples the genome broadly and, one hopes, in a representative fashion, the short internodes may also imply that there was a rapid radiation at the origin of *Strisores*, and relatively little molecular divergence accumulated as the major lineages of *Strisores* split.

I identified no bias consistent with trimming method chosen in any of my analyses, nor was trimming identified as a source of bias by my quartet summary analyses. Trimming methods were introduced to remove potential sources of noise present at the end of alignments, which generally represent areas of lower sequencing coverage and greater sequencing error. Due to variation in sequence length across taxa, these regions are also more difficult to align, resulting in the ragged ends

evident in UCE data. Additionally, some trimming algorithms can remove regions of potential misalignment or, as in Aliscore, regions that only contribute random noise to the analysis. For these reasons, I prefer trimming over not, and the more sophisticated and objective trimming method implemented in Aliscore over that in PHYLUCE.

The majority of my nucleotide ML and coalescent analyses preferred the topology A (Table 3.4), even when those loci showing the greatest GC variance among taxa were dropped. However, I believe that the similarity in GC content between potoos and oilbird could be driving the predominance of topology A. Great variation in GC content is not accounted for in many models of sequence evolution, and so I implemented RY coding to reduce this signal and circumvent potential phylogenetic bias.

The greatest evidence for preferring topology B over the others is presented in the RY coding results (Table 3.4). None of my RY-coded datasets recovered topology A, indicating potential bias due to the similarly high GC content of oilbird and potoos (Figure 3.4). Further support is present in the results of taxon jackknifing for dataset 4 (Fig. 3.8b). Applying any trimming to the datasets of taxon set 4 caused a shift to the majority of matrices preferring topology B. I do not believe topology C to be the preferred tree as it was the minority result in the RY-coded analyses (only 10 out of 32 matrices resulted in topology C), and did not occur in any of the jackknifing analyses for either taxon sets 1 or 4.

I acknowledge evidence conflicting with preference for topology B in the indel data, for which I found no instances of indel events supporting topology B, and in the presence of monophyletic groups present in my single locus trees (Table 3.6),

all of which showed a predominance of support for topology A. However, the latter was true even when the nucleotide matrix resulted in a topology other than A (e.g. see D1CP in Tables 3.4 and 3.6). Gene tree error is unsurprising in loci such as UCEs, which vary greatly in information content, and which individually have relatively few informative sites relative to non-ultraconserved sequences (Fig. 3.5).

These analyses demonstrate the importance of accounting for GC content in phylogenomic analyses. Analyses of the RY-coded *Strisores* matrices of Prum et al. (2015) and Reddy et al. (2017) also resulted in topological differences from their nucleotide trees (Supplementary Fig. S4). However, neither of these results provides strong evidence supporting any of my three major topologies. RY-coded analysis of the Prum et al. *Strisores* dataset resulted in topology C, though with extremely low bootstrap support (3%; bootstrap majority-rule consensus resulted in the sister topology). This result could potentially be due to taxon sampling issues as that dataset includes only one potoo, frogmouth and owlet-nightjar (Supplementary Fig. S4a). RY-coded analysis of the Reddy et al. *Strisores* matrix resulted in a clade including frogmouths, oilbird and potoos, which was not seen in any of my analyses (Supplementary Fig. S4b). I attempted to account for GC content variation through partitioning, and GC variance through dropping the most variant loci, but neither produced a consistent answer across matrices (see Table 3.4 D1CP, D3IP, D3CAN, D4CP).

There is evidence that AT-rich genes are more reliable phylogenetic markers than GC-rich genes (Romiguier et al. 2013; 2016; Bossert et al. 2017). GC rich regions can be generated by GC biased gene conversion which is the process by

which G/C alleles are preferentially incorporated during double-stranded break repair of chromosomes during recombination (Galtier et al. 2001). As a consequence, regions of high recombination tend to be GC-rich, and regions of the genome with increased GC content do not necessarily reflect evolutionary history, and therefore can provide errant phylogenetic signal (Eyre-Walker 1993; Galtier & Duret, 2007; Romiguier et al. 2010). Genome-wide variation in recombination rate can produce spatial variation in GC content, a phenomenon present in birds (Figuet et al. 2015), and obscure phylogenetic analysis through incomplete lineage sorting (Kubatko & Degnan, 2007; Degnan & Rosenberg, 2009; Liu et al. 2009). Further, base composition heterogeneity across taxa (as evidenced through GC variation) can bias phylogenetic analyses through long-branch attraction (Bergsten 2005) or model misspecification (Romiguier et al. 2016).

Incomplete lineage sorting (ILS) could contribute to the discordance I saw with these data, and is likely present in my dataset to some extent. My ratios of monophyletic groups found in the single gene trees (Table 3.6) appear to fit expectations of ILS—that is I have one predominant topology, A, that appears at higher frequency than two minority topologies, B and C, which appear at equal frequency. However, it is important to note that these searches were not constrained by topology within the monophyletic group, and so will not perfectly reflect the number of locus trees identical to my three main ML topologies. Implementation of the coalescent model to account for ILS yielded a majority of topology A trees (see Table 3.4 SVDquartets), as well as topologies not found in any of my ML analyses (Fig. 3.7). I have no evidence that ILS is an overwhelming signal or major biasing

factor in these analyses, and the results presented demonstrate that my analyses do respond to changes in accounting for GC content.

Similarly, I do not see evidence that long-branch attraction (LBA) is a strong force in these analyses. Oilbird and potoos are not anomalously long branches in my datasets; in fact they are among the shorter terminals in the *Strisores* tree. (Fig. 3.9). Moreover, the position of oilbird does not move around the tree to other long branches when taxon sampling is changed. The effect of GC content similarity in biasing my topology is a strong factor, however, and not entirely unrelated to LBA, but a more specific cause than encompassed by the term LBA.

It is most likely that model misspecification is at the heart of the analytical issues seen here. I used the best methods available to choose models of sequence evolution that fit my data, whether concatenated or partitioned analyses, but always selected two of the most parameter-rich models of sequence evolution (GTR+I+G and TVM+I+G) which differ in only one parameter. This indicates complex forces acting on my sequences. Phylogenomic analyses offer the advantage of more data to resolve short internodes, but face increasing difficulty in applying appropriate models of molecular evolution as data increases (Philippe et al. 2011). My nucleotide matrices represent a much greater amount of information than my RY-coded matrices, but may contain a bias which prevents accurate resolution of the position of oilbird vs potoos. My results indicate that readily available models of molecular evolution are insufficient to encapsulate all phenomena present in genome-scale matrices, and this problem may be at the root of current issues in phylogenomic analysis.

Through the numerous analyses presented here, I was able to uncover hidden conflict in the data. Had I simply made one matrix with this data, and conducted one phylogenetic analysis, I would have produced a tree with apparently high statistical support and accepted that result as a rigorous basis to my downstream evolutionary analyses. Observing high apparent statistical support in a phylogenomic analysis is not a reliable metric of accuracy when model misspecification is a problem (Kumar et al. 2012; Hahn & Nakhleh, 2015). Only by using multiple analytical approaches I were able to detect the conflict present in my dataset.

*Evolutionary History of Nocturnality in Strisores*

With a best estimate of phylogeny in hand, I can hypothesize scenarios by which nocturnality arose in the *Strisores* (Fig. 3.10a). In debating among these potential scenarios, one can take into consideration adaptations to night vision in the group, as the inability to see in low light conditions would be the greatest hurdle facing a bird transitioning from a diurnal to a nocturnal lifestyle. Little is known about nocturnal vision in birds, though some study has been conducted in *Strisores*. The eyes of two lineages of *Strisores* have been characterized morphologically, and each displays remarkable adaptations to increase light sensitivity (Fig. 3.10b). Caprimulgids (nightjars and nighthawks) have a tapetum lucidum—a mirror-like structure at the back of the eye that reflects light back through retinal cells, providing a second chance for the absorption of incoming photons (Nicol & Arnott, 1974). Tapeta lucida have arisen independently in many nocturnal vertebrates. Their presence can be observed as a brilliant reflection from an animal's eyes in a spotlight at night. However, the biochemical composition of tapeta is quite varied (Walls 1963;

Martin et al. 2004), reflecting their independent evolutionary origin. Like caprimulgids, potoos display a brilliant eye shine, indicating the presence of a tapetum, but the biochemical composition of their tapetum has not been studied, and may have arisen independently. Indeed, the topology of my current best estimate of the tree suggests an independent transition (Fig. 3.9a). Frogmouths and owlet-nightjars do not have a brilliant eye shine, so must not have tapeta, but their visual adaptations have not been studied further. A different adaptation to nocturnality is seen in the oilbird, which has a unique retinal structure with three layers of photoreceptor cells, and the highest photoreceptor cell density of any vertebrate studied (Martin et al. 2004; Rojas et al. 2004).

Our topology of the *Strisores* phylogeny allows for multiple scenarios for the origination of nocturnality in this group—too many to depict here. Instead, for the sake of discussion I illustrate two extreme scenarios; 1) that nocturnality evolved in the ancestor of the *Strisores,* and there was a reversal to diurnality in the swifts and hummingbirds (Fig. 3.10c), or 2) that the *Strisores* originated from a diurnal ancestor, and there were multiple independent adaptations to nocturnality within this clade (Fig. 3.10d). Given the known variety of adaptations to night vision among *Strisores* described above, it is plausible to hypothesize that there were multiple transitions to nocturnality within the *Strisores*, with distinct adaptations to low light vision in each case. Parsimony might prefer the single origin scenario, as was suggested by (Mayr 2010), however, parsimony only holds if the transitions considered are given equal likelihood of occurring. It is difficult to believe that a regeneration of diurnal vision from a nocturnal state is as likely as a loss of diurnal vision.

Perhaps of greater evolutionary importance, having both a topology and the knowledge of adaptations to night vision in hand, I can begin to infer the deeper evolutionary origin of nocturnality in *Strisores*, and likely in all of birds. The non-random clustering of nocturnality in *Strisores* is indicative of a deep homology or predisposition to a nocturnal lifestyle. This predisposition is likely of a genetic basis, such as the gene network underlying appendages in vertebrates, arthropods and other bilaterians (Shubin et al. 2009). The presence of tapeta in other vertebrate groups suggests the possibility of a deep, shared, evolutionary origin of this trait, and genetic analyses can potentially uncover if the same adaptations led to its origin in potoos and caprimulgids, or if, as indicated by my tree, they differ. Studying the vision of these birds in more detail, such as by focusing on the phototransduction cascade, will help to unravel the evolutionary history of nocturnality in these birds, all birds, and to shed light on its origin across life.

**Acknowledgements**

I thank Edward L. Braun, Derrick J. Zwickl and Marek Borowiec for productive discussion of methodological issues. I acknowledge the genetic repositories listed in Supplementary Table S1, and especially thank the intrepid field collectors and diligent curatorial staff whose work makes studies such as this possible. Computations in this paper were run on the Smithsonian Institution's High Performance Computing Cluster, Hydra, as well as the Cyberinfrastructure for Phylogenetic Research (CIPRES) Science Gateway (Miller, Pfeiffer, & Schwartz, 2010). Laboratory work was conducted at and with the support of the US National Museum of Natural History. *Funding* – NDW received Smithsonian Institution

support through the Predoctoral Fellowship Program, the Frontiers in Phylogenetics Program, the Scholarly Studies Program and the Consortium for Understanding and Sustaining a Biodiverse Planet.

## Tables

Table 3.1. The four taxon sets used in this study vary in representation of the potoos (Nyctibiidae) and the presence or absence of the rufous potoo (*Nyctibius bracteatus*). Shaded box = taxon present.

| Family | Species | Common Name | Taxon Set | | | |
|---|---|---|---|---|---|---|
| | | | 1 | 2 | 3 | 4 |
| Trochilidae | *Colibri coruscans* | Sparkling violetear | ■ | ■ | ■ | ■ |
| | *Topaza pella* | Crimson topaz | ■ | ■ | ■ | ■ |
| Apodidae | *Aerodramus vanikorensis* | Uniform swiftlet | ■ | ■ | ■ | ■ |
| | *Streptoprocne zonaris* | White-collared swift | ■ | ■ | ■ | ■ |
| Hemiprocnidae | *Hemiprocne mystacea* | Moustached treeswift | ■ | ■ | ■ | ■ |
| Aegothelidae | *Aegotheles cristatus* | Australian owlet-nightjar | ■ | ■ | ■ | ■ |
| | *Aegotheles insignis* | Feline owlet-nightjar | ■ | ■ | ■ | ■ |
| Caprimulgidae | *Hydropsalis rufiventris* | Rufous-bellied nighthawk | ■ | ■ | ■ | ■ |
| | *Lyncornis macrotis* | Great-eared nightjar | ■ | ■ | ■ | ■ |
| | *Eurostopodus mystacalis* | White-throated nightjar | ■ | ■ | ■ | ■ |
| Nyctibiidae | *Nyctibius aethereus* | Long-tailed potoo | | ■ | ■ | ■ |
| | *Nyctibius bracteatus* | Rufous potoo | ■ | | | ■ |
| | *Nyctibius grandis* | Great potoo | ■ | ■ | ■ | ■ |
| | *Nyctibius griseus* | Common potoo | | | ■ | ■ |
| | *Nyctibius jamaicensis* | Northern potoo | | | ■ | ■ |
| | *Nyctibius leucopterus* | White-winged potoo | | | ■ | ■ |
| | *Nyctibius maculosus* | Andean potoo | | | ■ | ■ |
| Podargidae | *Batrachostomus septimus* | Philippine frogmouth | ■ | ■ | ■ | ■ |
| | *Podargus strigoides* | Tawny frogmouth | ■ | ■ | ■ | ■ |
| Steatornithidae | *Steatornis caripensis* | Oilbird | ■ | ■ | ■ | ■ |
| Outgroups | *Columba livia* | Rock dove | ■ | ■ | ■ | ■ |
| | *Anseranas semipalmata* | Magpie goose | ■ | ■ | ■ | ■ |
| | *Gallus gallus* | Red junglefowl | ■ | ■ | ■ | ■ |

Table 3.2. Definition and characteristics of the 32 data matrices used in this study. % = % complete. PI = parsimony informative.

| Taxon Set | Matrix | # Loci | % | Trimming | Length (bp) | % GC | # (%) PI Sites |
|---|---|---|---|---|---|---|---|
| 1 | D1CA | 2,465 | 100 | Aliscore | 1,201,015 | 39.1 | 87,392 (7.3) |
| | D1IA | 4,195 | 75 | Aliscore | 1,881,164 | 38.9 | 131,519 (7.0) |
| | D1CAN | 2,465 | 100 | Aliscore (-N) | 1,147,997 | 39 | 83,864 (7.3) |
| | D1IAN | 4,195 | 75 | Aliscore (-N) | 1,802,357 | 38.9 | 127,590 (7.1) |
| | D1CP | 2,464 | 100 | PHYLUCE | 1,388,516 | 39 | 107,944 (7.8) |
| | D1IP | 4,195 | 75 | PHYLUCE | 2,208,804 | 38.8 | 166,533 (7.5) |
| | D1CU | 2,465 | 100 | Untrimmed | 2,515,454 | 39 | 194,669 (7.7) |
| | D1IU | 4,196 | 75 | Untrimmed | 4,068,682 | 38.9 | 301,319 (7.4) |
| 2 | D2CA | 2,588 | 100 | Aliscore | 1,269,297 | 39 | 94,471 (7.4) |
| | D2IA | 4,234 | 75 | Aliscore | 1,927,078 | 38.9 | 139,286 (7.2) |
| | D2CAN | 2,588 | 100 | Aliscore (-N) | 1,214,001 | 38.9 | 91,084 (7.5) |
| | D2IAN | 4,234 | 75 | Aliscore (-N) | 1,851,320 | 38.9 | 135,395 (7.3) |
| | D2CP | 2,587 | 100 | PHYLUCE | 1,465,994 | 38.8 | 116,795 (8.0) |
| | D2IP | 4,233 | 75 | PHYLUCE | 2,248,723 | 38.8 | 174,140 (7.7) |
| | D2CU | 2,588 | 100 | Untrimmed | 2,633,594 | 38.9 | 206,659 (7.8) |
| | D2IU | 4,235 | 75 | Untrimmed | 4,098,197 | 38.9 | 308,435 (7.5) |
| 3 | D3CA | 2,435 | 100 | Aliscore | 1,219,747 | 38.9 | 97,113 (8.0) |
| | D3IA | 4,242 | 75 | Aliscore | 1,973,815 | 38.9 | 152,885 (7.7) |
| | D3CAN | 2,435 | 100 | Aliscore (-N) | 1,173,606 | 38.9 | 95,433 (8.1) |
| | D3IAN | 4,242 | 75 | Aliscore (-N) | 1,896,143 | 38.8 | 150,313 (7.9) |
| | D3CP | 2,435 | 100 | PHYLUCE | 1,407,808 | 38.8 | 121,358 (8.6) |
| | D3IP | 4,241 | 75 | PHYLUCE | 2,264,506 | 38.8 | 188,352 (8.3) |
| | D3CU | 2,435 | 100 | Untrimmed | 2,493,709 | 38.9 | 216,598 (8.7) |
| | D3IU | 4,243 | 75 | Untrimmed | 4,137,329 | 38.9 | 345,202 (8.3) |
| 4 | D4CA | 2,289 | 100 | Aliscore | 1,156,873 | 39 | 92,668 (8.0) |
| | D4IA | 4,196 | 75 | Aliscore | 1,951,742 | 38.9 | 152,284 (7.8) |
| | D4CAN | 2,289 | 100 | Aliscore (-N) | 1,099,980 | 39 | 89,041 (8.1) |
| | D4IAN | 4,196 | 75 | Aliscore (-N) | 1,865,976 | 38.9 | 148,868 (8.0) |
| | D4CP | 2,289 | 100 | PHYLUCE | 1,298,686 | 38.9 | 110,129 (8.5) |
| | D4IP | 4,194 | 75 | PHYLUCE | 2,220,004 | 38.8 | 184,467 (8.3) |
| | D4CU | 2,289 | 100 | Untrimmed | 2,353,369 | 39 | 206,876 (8.8) |
| | D4IU | 4,196 | 75 | Untrimmed | 4,110,801 | 38.9 | 349,239 (8.5) |

Table 3.3. Results of efforts to discover where UCEs intersect functional annotations. Simulated results represent averages of 100 individual simulated datasets. Note that the UTR, CDS and exon annotation files may overlap. * = significant result from the binomial test ($P < 0.05$).

| | Chicken | | | Human | | |
|---|---|---|---|---|---|---|
| | **UCE** | **Simulated** | | **UCE** | **Simulated** | |
| **# Probes** | 5,472 | 5,472 | | 5,472 | 5,472 | |
| **Unique Map to Genome** | 5,222 | n/a | | 3,810 | n/a | |
| **Intersected an annotation** | 1,788 | 1,542 | | 2,672 | 3,344 | |
| **Intersections** | | | **Fold Change** | | | **Fold Change** |
| **CpG Islands** | 51* | 96 | 0.5X | 30* | 46 | 0.7X |
| **3` UTRs** | 104* | 33 | 3.1X | 265* | 76 | 3.5X |
| **5` UTRs** | 22* | 10 | 2.2X | 95* | 41 | 2.3X |
| **CDS** | 250* | 92 | 2.7X | 336* | 104 | 3.2X |
| **Exons** | 348* | 128 | 2.7X | 590* | 194 | 3.0X |
| **Introns** | 948* | 854 | 1.1X | 2,072* | 2,323 | 0.9X |

Table 3.4. Topological results and support values from phylogenetic analyses of the 32 unpartitioned datasets (Table 3.2). The three alternative topologies found in ML analyses are color-coded and labeled as in Fig. 3.6. `BS` indicates bootstrap support values for nodes labeled with an asterisk in Fig. 3.6. RY-coded analyses had the datasets recoded to purine/pyrimidine to reduce GC artifacts. “GCvar” analyses had 25% of loci with the highest GC variance removed from the datasets. * = ML analyses run in GARLI under TVM+I+G; all other ML analyses run in RAxML under GTR+I+G. [A] = significant AU test results. SVDquartets topologies correspond to those presented in Fig. 3.7. SVD-2, -3 and -4 topologies are color coded blue to highlight that they differ from topology A only in the placement of the outgroup root.

| Matrix | ML | BS | RY Coded | BS | SVDquartets | BS | GCvar low 75% | BS |
|---|---|---|---|---|---|---|---|---|
| D1CP | B | 55 | B | 92 | SVD-1 | | A* | 66 |
| D1IP | C | 55 | C | 81 | SVD-1 | | C | 65 |
| D1CU | A* | 52 | B | 87 | SVD-1 | | A* | 74 |
| D1IU | A | 76 | C | 72 | SVD-2 | 74 | A | 74 |
| D1CA | A | 50 | B | 90 | A | 52 | A | 78 |
| D1CAN | A* | 81 | B | 80 | SVD-1 | | A* | 65 |
| D1IA | A | 66 | B | 58 | SVD-3 | 44 | A | 52 |
| D1IAN | A* | 95 | C | 41 | SVD-3 | 69 | A* | 89 |
| D2CP | A | 54 | B | 87 | SVD-1 | | A | 44 |
| D2IP | C | 53 | C | 76 | SVD-3 | 70 | C | 56 |
| D2CU | A* | 94 | B | 90 | SVD-1 | | A* | 60 |
| D2IU | A | 90 | C | 64 | SVD-4 | 73 | A | 84 |
| D2CA | A* | 50 | B | 67 | SVD-1 | | A | 64 |
| D2CAN | A | 88 | B | 81 | SVD-1 | | A* | 60 |
| D2IA | A | 86 | B | 37 | SVD-4 | 67 | A | 92 |
| D2IAN | A | 92 | B | 51 | SVD-4 | 80 | A* | 80 |
| D3CP | A | 68 | B | 87 | SVD-1 | | A | 64 |
| D3IP | A | 69 | B | 58 | SVD-1 | | C | 60 |
| D3CU | A*,A | 99 | C | 37 | SVD-1 | | A | 96 |
| D3IU | A[A] | 100 | C | 59 | A | 93 | A | 92 |
| D3CA | A | 86 | B | 87 | SVD-3 | 75 | A* | 73 |
| D3CAN | A | 52 | B | 90 | SVD-3 | 78 | B | 53 |
| D3IA | A | 90 | B | 57 | SVD-3 | 93 | A* | 88 |
| D3IAN | A | 86 | B | 82 | SVD-3 | 93 | A* | 76 |
| D4CP | B | 49 | B | 90 | SVD-1 | | A | 72 |
| D4IP | A | 72 | C | 75 | A | 90 | A | 47 |
| D4CU | A*,A | 98 | C | 76 | SVD-1 | | A* | 100 |
| D4IU | A[A] | 100 | C | 67 | A | 82 | A | 98 |
| D4CA | B | 78 | B | 95 | SVD-1 | | B | 45 |
| D4CAN | B | 56 | B | 91 | SVD-1 | | A | 64 |
| D4IA | A | 96 | B | 81 | SVD-3 | 88 | A | 82 |
| D4IAN | A | 78 | B | 74 | SVD-3 | 85 | A* | 94 |

Table 3.5. Topologies and support values resulting from phylogenetic analyses of functional annotation subsets of loci from datasets 1 and 4. The three alternative topologies found in ML analyses are color-coded as in Fig. 3.6. Analyses that resulted in a topology other than those presented in Fig. 3.6 or Fig. 3.7 are shaded grey (those that recovered oilbirds and potoos sister are shaded blue). `BS` indicates bootstrap support values for nodes labeled with an asterisk in Fig. 3.6. * = analysis run in GARLI under TVM+I+G.

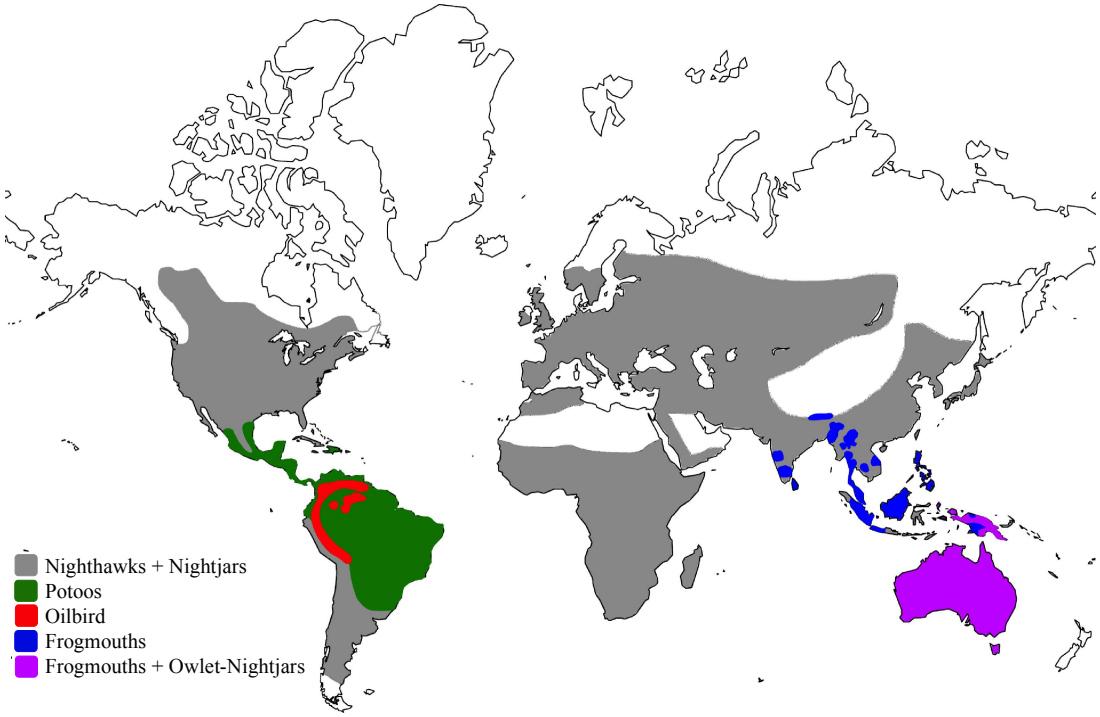
| Matrix | CpG | BS | 3` UTRs | BS | 5` UTRs | BS |
|---|---|---|---|---|---|---|
| D1CP | | | * | | | |
| D1CU | | | B* | 62 | * | |
| D1CA | | | * | | | |
| D4CP | | | * | | * | |
| D4CU | | | A* | 41 | | |
| D4CA | | | * | | * | |
| Matrix | CDS | BS | Exon | BS | Intron | BS |
| D1CP | C | 47 | A | 38 | A* | 33 |
| D1CU | | | | | A | 72 |
| D1CA | | | | | A* | 44 |
| D4CP | | | A | 40 | | |
| D4CU | | | | | A* | 52 |
| D4CA | C | 28 | C | 27 | B | 50 |

Table 3.6. Single locus tree support for groups representing my three main ML topologies—A (oilbird, potoos), B (potoos, other ingroups), C (oilbird, other ingroups), where the clade "other ingroups" is defined as swifts+hummingbirds+owlet-nightjars+frogmouths. Numbers represent individual UCE locus trees for which a given monophyletic group was present, in both nucleotide and RY-coded form. Columns representing the three major ML topologies are shaded from low values (white) to high values (dark green), separately for nucleotide and RY-coded results. The number of single locus tree that recover monophyly of potoos is included for reference.

| | | Nucleotide | | | | RY-coded | | | |
|---|---|---|---|---|---|---|---|---|---|
| Matrix | # Loci | A | B | C | Potoos | A | B | C | Potoos |
| D1CP | 2,464 | 201 | 85 | 72 | 1,812 | 132 | 32 | 22 | 1,181 |
| D1CU | 2,465 | 199 | 80 | 75 | 1,818 | 99 | 39 | 40 | 1,238 |
| D1IP | 4,195 | 464 | 124 | 118 | 3,169 | 309 | 58 | 56 | 2,299 |
| D1IU | 4,196 | 438 | 137 | 115 | 3,171 | 265 | 60 | 58 | 2,349 |
| D1CA | 2,465 | 188 | 82 | 56 | 1,691 | 95 | 20 | 24 | 1,072 |
| D1CAN | 2,465 | 192 | 63 | 44 | 1,678 | 107 | 19 | 11 | 1,047 |
| D1IA | 4,195 | 406 | 115 | 110 | 2,984 | 267 | 49 | 36 | 2,123 |
| D1IAN | 4,195 | 419 | 103 | 94 | 2,997 | 279 | 35 | 35 | 2,105 |
| D4CP | 2,289 | 176 | 70 | 71 | 1,679 | 109 | 30 | 22 | 1,013 |
| D4CU | 2,289 | 178 | 67 | 78 | 1,742 | 96 | 32 | 27 | 1,057 |
| D4IP | 4,194 | 374 | 118 | 120 | 2,956 | 218 | 48 | 51 | 1,782 |
| D4IU | 4,196 | 355 | 130 | 117 | 3,043 | 185 | 65 | 47 | 1,862 |
| D4CA | 2,289 | 173 | 66 | 70 | 1,572 | 104 | 25 | 12 | 892 |
| D4CAN | 2,289 | 173 | 51 | 59 | 1,575 | 90 | 22 | 18 | 879 |
| D4IA | 4,196 | 348 | 104 | 110 | 2,715 | 200 | 40 | 26 | 1,557 |
| D4IAN | 4,196 | 330 | 93 | 107 | 2,751 | 192 | 37 | 47 | 1,542 |

## Figures

Figure 3.1. Range map of the major *Strisores* lineages. Nighthawks and nightjars overlap frogmouth and owlet-nightjar distribution, with the exception that they do not occur on Tasmania. Range maps based on data from the Handbook of the Birds of the World.


Nighthawks + Nightjars
Potoos
Oilbird
Frogmouths
Frogmouths + Owlet-Nightjars

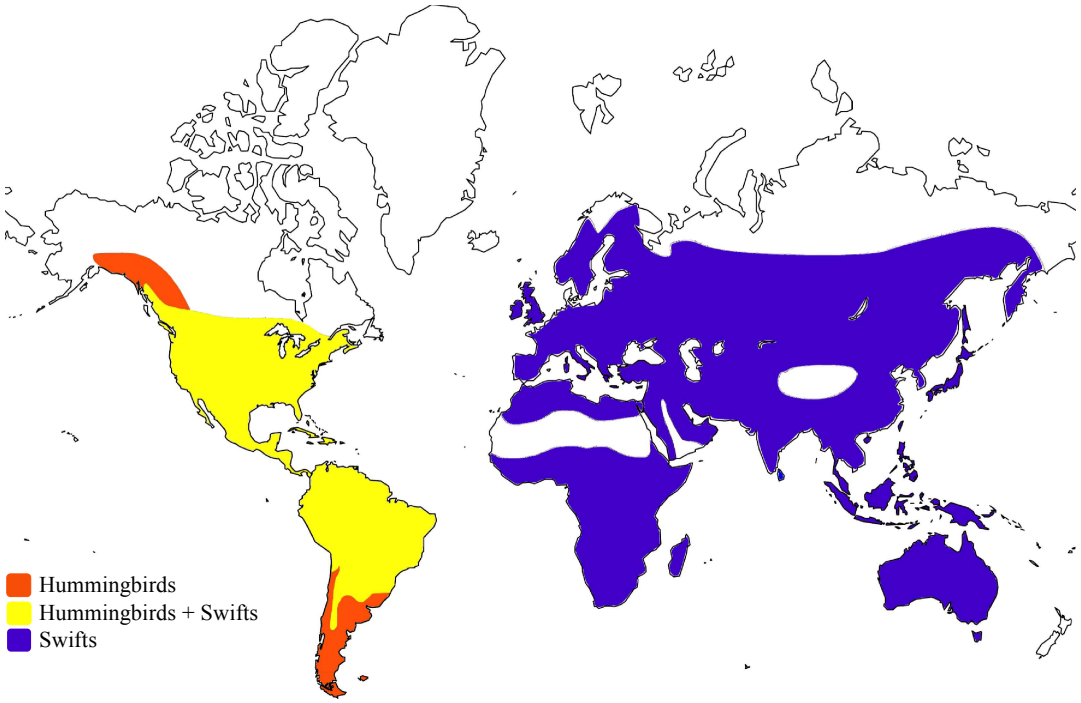

Hummingbirds
Hummingbirds + Swifts
Swifts

Figure 3.2. Previous molecular estimates of the *Strisores* phylogeny. Support values are bootstraps, with the exception of Ericson et al., which are posterior probabilities. Ericson et al. collapsed all nodes with PP < 0.95. Braun & Huddleston topology shown is from their combined analysis. Prum et al. and Reddy et al. support values are from the ML trees available in their supplementary material.

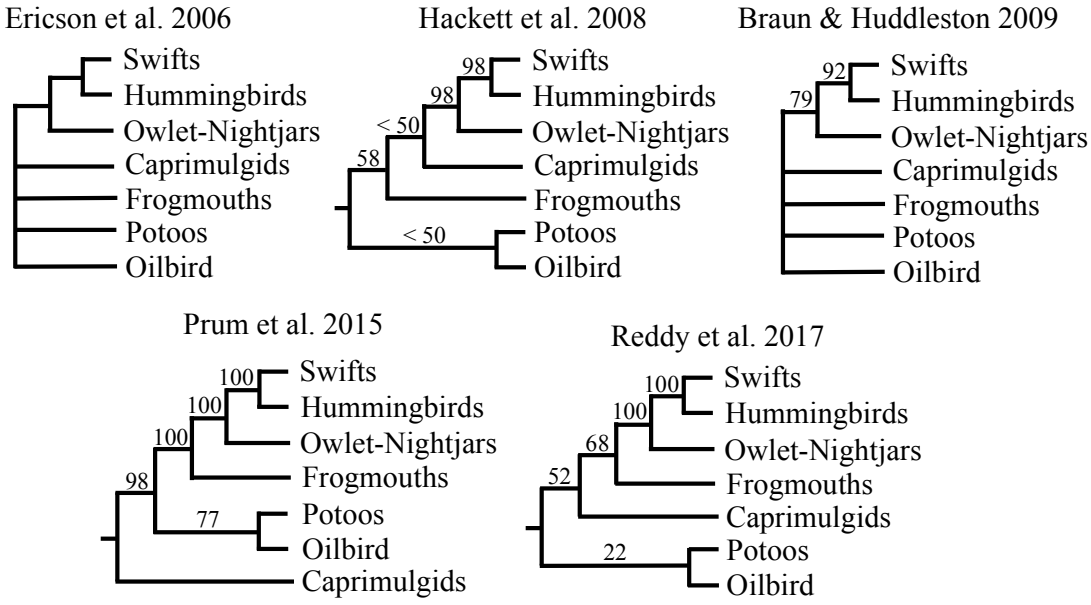


Figure 3.3. Variation in GC content among loci across all untrimmed datasets (a) and all taxon set 1 matrices (b). GC variance among taxa within individual locus alignments of taxon sets 1 and 4 matrices (c). Small dots are individual data points summarized by the underlying box and whisker plot. Median is represented by the line bisecting the box, and separates the first and third quartiles. Whiskers extend to values no further than 1.5X the interquartile range from the median. Large dots represent outliers. All other matrices showed similar patterns of variation.

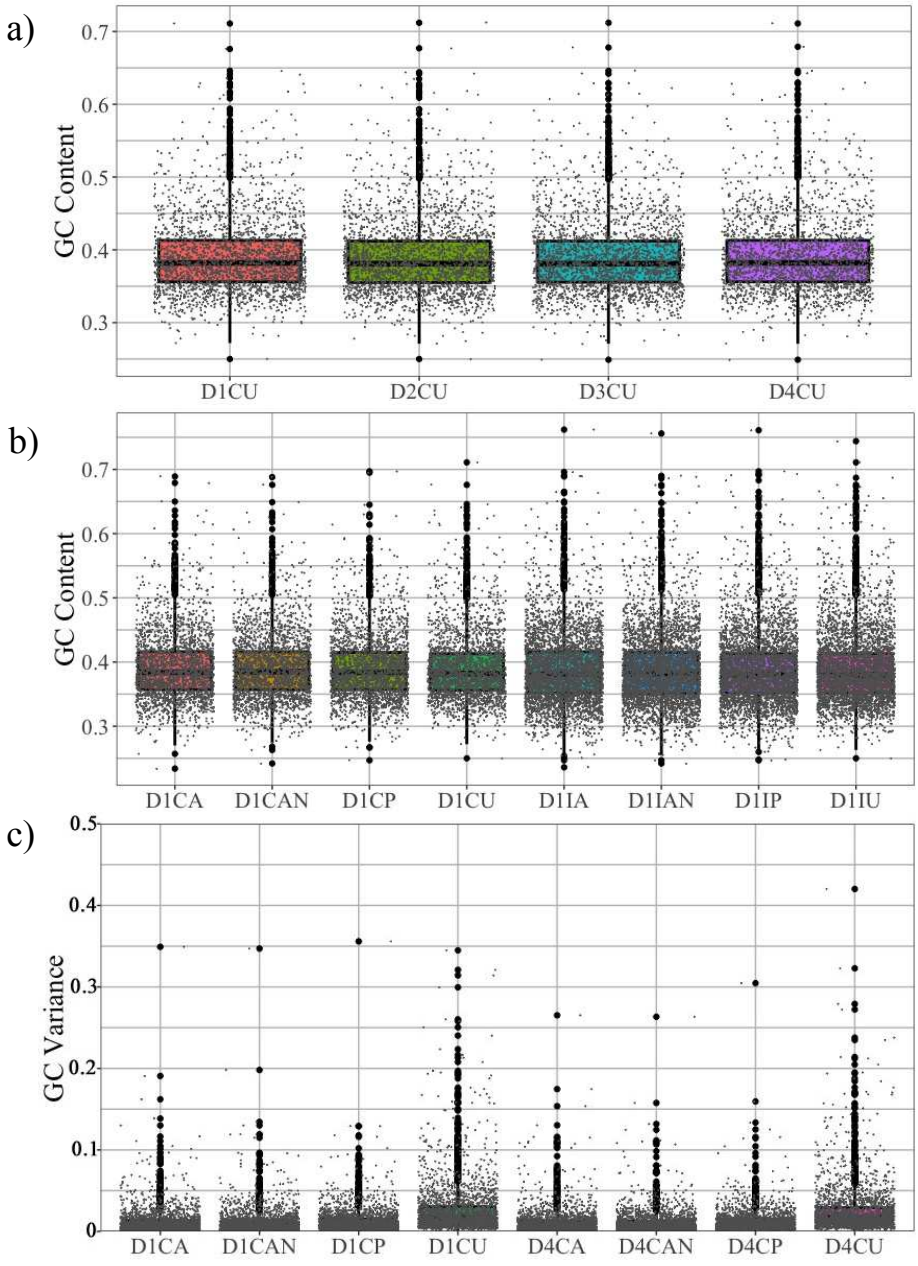

a)
GC Content
0.7
0.6
0.5
0.4
0.3
D1CU
D2CU
D3CU
D4CU
b)
GC Content
0.7
0.6
0.5
0.4
0.3
D1CA
D1CAN
D1CP
D1CU
D1IA
D1IAN
D1IP
D1IU
c)
GC Variance
0.5
0.4
0.3
0.2
0.1
0
D1CA
D1CAN
D1CP
D1CU
D4CA
D4CAN
D4CP
D4CU

Figure 3.4. Variation in GC content for individual taxa across all matrices in which they appear (see Table 3.1). The left portion of the graph shows GC content over full alignments, the right over parsimony informative sites only. Dashed lines demarcate major *Strisores* lineages. Note the break in the x-axis. Dots are individual data points summarized by the underlying box and whisker plot. Median is represented by the line bisecting the box, and separates the first and third quartiles. Whiskers extend to values no further than 1.5X the interquartile range from the median.

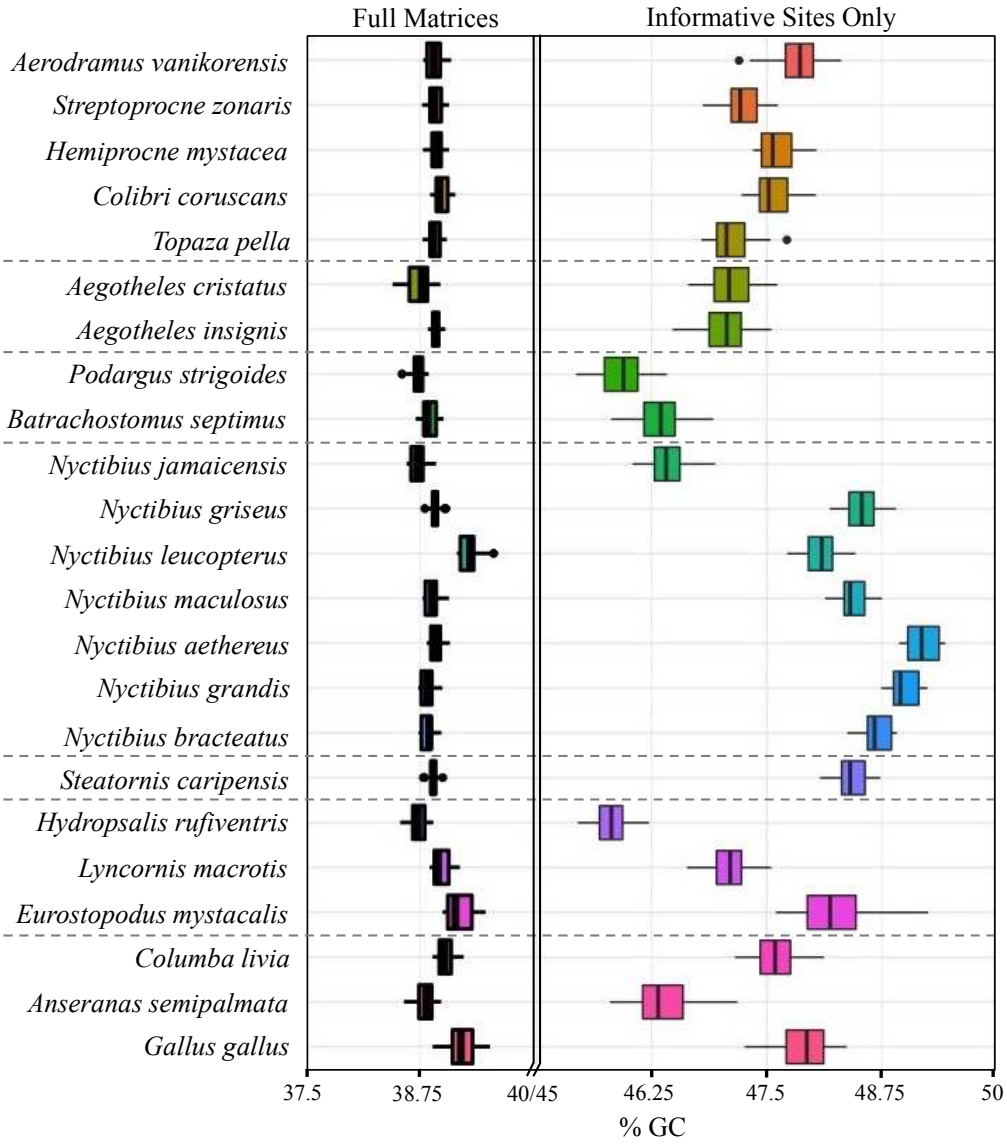


Figure 3.5. Proportion of parsimony informative (PI) sites in the loci of dataset 1 matrices.

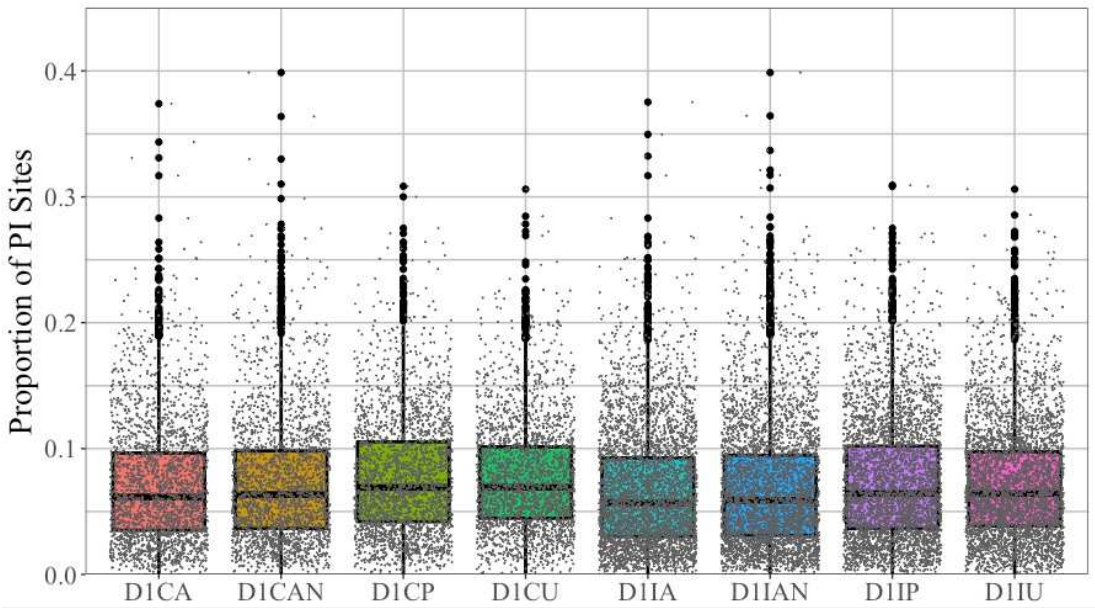


Figure 3.6. The three alternative tree topologies for the major lineages of *Strisores* found in maximum likelihood (ML) analyses of this study. Each topology is color-coded to correspond to Tables 3.4 and 3.5. Asterisks mark nodes for which relevant bootstrap support is given in other figures and tables.

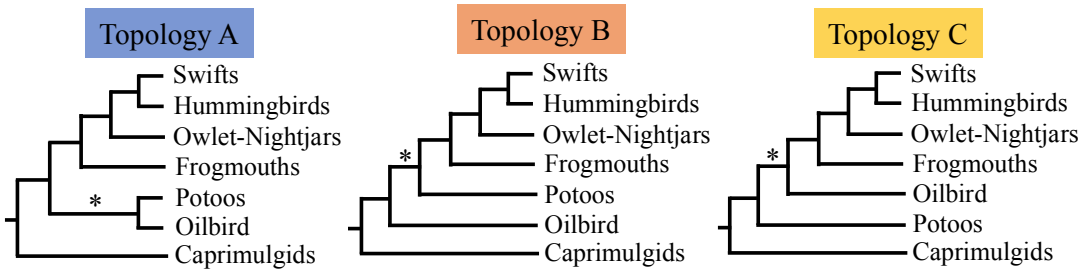


Figure 3.7. SVDquartet analysis of the 32 major matrices resulted in these four topologies as well as topology A (see Figure 3.6). Note that SVD-2, -3 and -4 have the same ingroup topology as topology A, and differ only in the placement of the outgroup root.

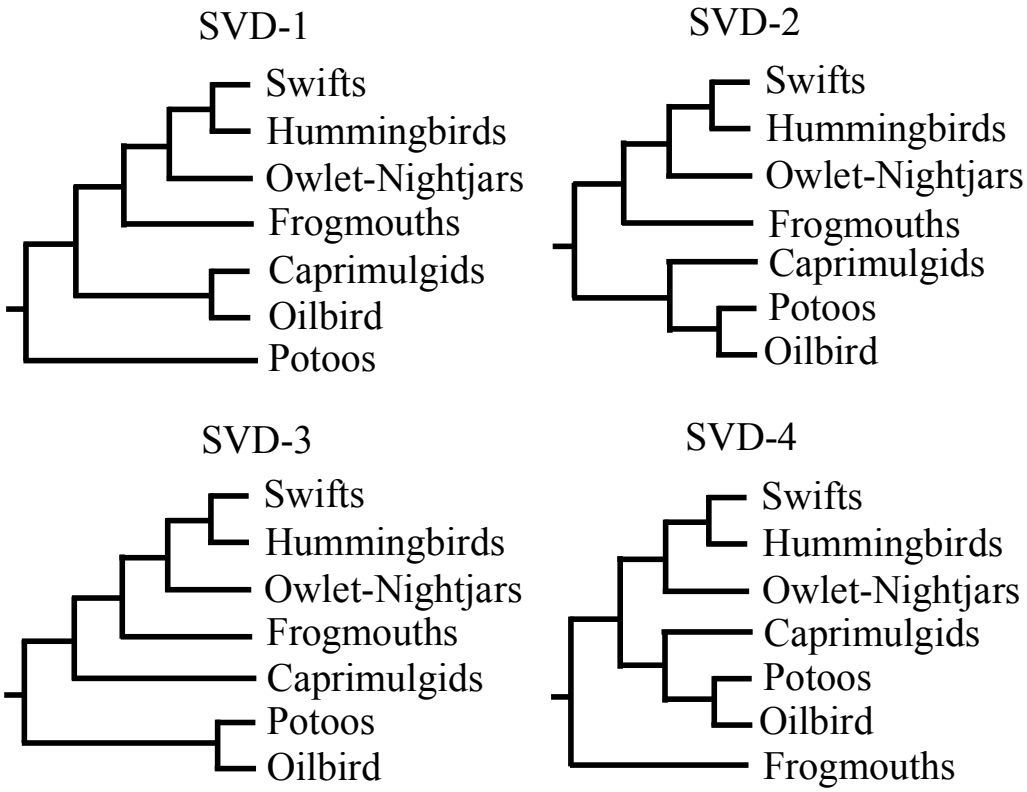
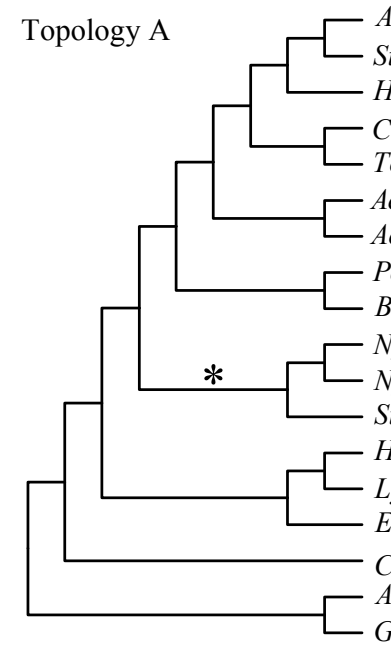
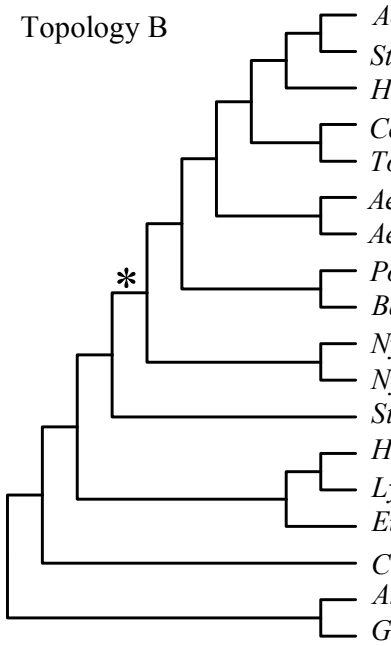


Figure 3.8. Topologies and support values resulting from taxon jackknifing of datasets for taxon sets 1 (a) and 4 (b). Jackknifed datasets resulted in either topology A or B; numbers are bootstrap support values at the node marked by the asterisk when that respective taxon was dropped. All nodes in all matrices had 100% bootstrap support when *Steatornis caripensis* was dropped.

a)

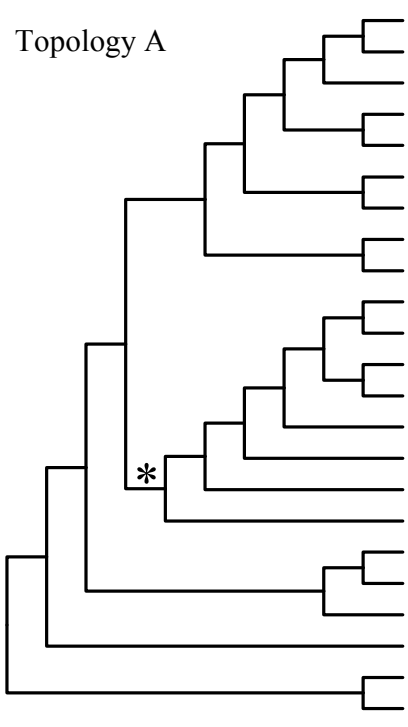



| | D1CU | D1CP | D1CA | D1CAN |
|---|---|---|---|---|
| *Aerodramus vanikorensis* | 58 | 49 | 51 | 78 |
| *Streptoprocne zonaris* | 76 | | 70 | 74 |
| *Hemiprocne mystacea* | 52 | | 58 | 72 |
| *Colibri coruscans* | 55 | 46 | 68 | 80 |
| *Topaza pella* | 51 | | 59 | 78 |
| *Aegotheles cristatus* | 62 | 46 | 68 | 80 |
| *Aegotheles insignis* | 62 | | 56 | 88 |
| *Podargus strigoides* | 65 | 62 | 66 | 80 |
| *Batrachostomus septimus* | 59 | | 64 | 72 |
| *Nyctibius grandis* | 76 | | | |
| *Nyctibius bracteatus* | | | 49 | 70 |
| *Steatornis caripensis* | n/a | n/a | n/a | n/a |
| *Hydropsalis rufiventris* | 72 | 50 | 60 | 76 |
| *Lyncornis macrotis* | | | 54 | 76 |
| *Eurostopodus mystacalis* | | | 63 | 80 |
| *Columba livia* | 76 | 52 | 70 | 88 |
| *Anseranas semipalmata* | 59 | 52 | 62 | 80 |
| *Gallus gallus* | 57 | | 66 | 74 |

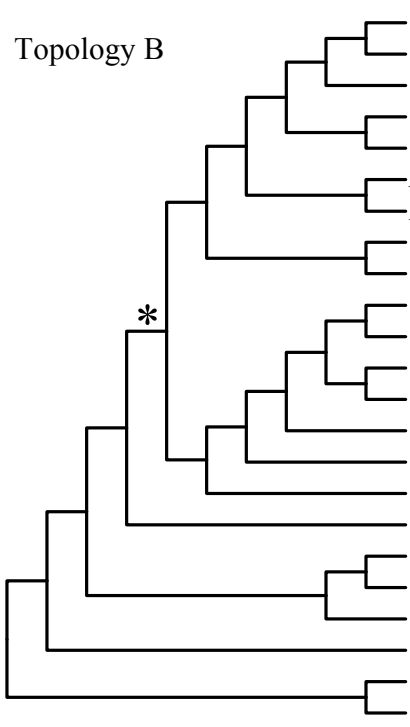



| | | | | |
|---|---|---|---|---|
| *Aerodramus vanikorensis* | | | | |
| *Streptoprocne zonaris* | | 47 | | |
| *Hemiprocne mystacea* | | 64 | | |
| *Colibri coruscans* | | | | |
| *Topaza pella* | | 53 | | |
| *Aegotheles cristatus* | | | | |
| *Aegotheles insignis* | | 45 | | |
| *Podargus strigoides* | | | | |
| *Batrachostomus septimus* | | 60 | | |
| *Nyctibius grandis* | | 72 | 80 | 61 |
| *Nyctibius bracteatus* | 57 | 60 | | |
| *Steatornis caripensis* | n/a | n/a | n/a | n/a |
| *Hydropsalis rufiventris* | | | | |
| *Lyncornis macrotis* | 47 | 62 | | |
| *Eurostopodus mystacalis* | 53 | 58 | | |
| *Columba livia* | | | | |
| *Anseranas semipalmata* | | | | |
| *Gallus gallus* | | 61 | | |

b)




| | D4CU | D4CP | D4CA | D4CAN |
|---|---|---|---|---|
| *Aerodramus vanikorensis* | 96 | | | |
| *Streptoprocne zonaris* | 100 | | | |
| *Hemiprocne mystacea* | 98 | | | |
| *Colibri coruscans* | 100 | 56 | | |
| *Topaza pella* | 98 | | | |
| *Aegotheles cristatus* | 98 | | | |
| *Aegotheles insignis* | 100 | | | |
| *Podargus strigoides* | 100 | 57 | | 68 |
| *Batrachostomus septimus* | 98 | | | |
| *Nyctibius jamaicensis* | 88 | | | |
| *Nyctibius griseus* | 96 | | | |
| *Nyctibius leucopterus* | 100 | | | |
| *Nyctibius maculosus* | 98 | | | |
| *Nyctibius aethereus* | 92 | | | |
| *Nyctibius grandis* | 100 | 60 | | |
| *Nyctibius bracteatus* | 99 | 51 | | |
| *Steatornis caripensis* | n/a | n/a | n/a | n/a |
| *Hydropsalis rufiventris* | 98 | 54 | | 52 |
| *Lyncornis macrotis* | 94 | | | |
| *Eurostopodus mystacalis* | 100 | 48 | | |
| *Columba livia* | 88 | 80 | 61 | 64 |
| *Anseranas semipalmata* | 96 | 46 | | 47 |
| *Gallus gallus* | 98 | | | |




| | | | | |
|---|---|---|---|---|
| *Aerodramus vanikorensis* | | 53 | 82 | 53 |
| *Streptoprocne zonaris* | | 53 | 68 | 59 |
| *Hemiprocne mystacea* | | 53 | 68 | 54 |
| *Colibri coruscans* | | | 68 | 57 |
| *Topaza pella* | | 66 | 68 | 68 |
| *Aegotheles cristatus* | | 53 | 76 | 55 |
| *Aegotheles insignis* | | 53 | 78 | 58 |
| *Podargus strigoides* | | | 63 | |
| *Batrachostomus septimus* | | 59 | 68 | 61 |
| *Nyctibius jamaicensis* | | 58 | 84 | 66 |
| *Nyctibius griseus* | | 54 | 70 | 92 |
| *Nyctibius leucopterus* | | 51 | 78 | 49 |
| *Nyctibius maculosus* | | 53 | 76 | 54 |
| *Nyctibius aethereus* | | 56 | 78 | 63 |
| *Nyctibius grandis* | | | 74 | 50 |
| *Nyctibius bracteatus* | | | 68 | 55 |
| *Steatornis caripensis* | n/a | n/a | n/a | n/a |
| *Hydropsalis rufiventris* | | | 62 | |
| *Lyncornis macrotis* | | 64 | 76 | 65 |
| *Eurostopodus mystacalis* | | | 63 | 52 |
| *Columba livia* | | | | |
| *Anseranas semipalmata* | | | 64 | |
| *Gallus gallus* | | 66 | 78 | 61 |

Figure 3.9. Phylogram depicting best estimate of the *Strisores* phylogeny based on UCE data. Branch lengths are from ML analysis of the concatenated, RY coded D4IA matrix. Bootstrap support values are plotted at their respective nodes. Scale units are substitutions per site.

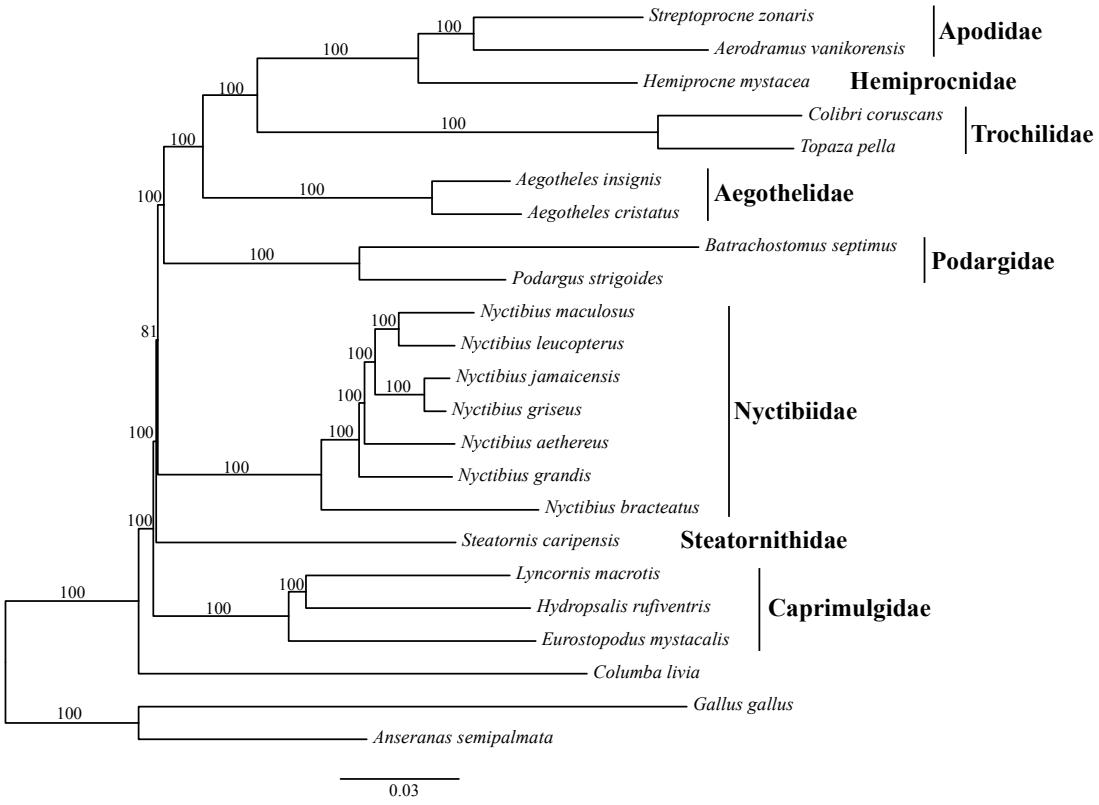
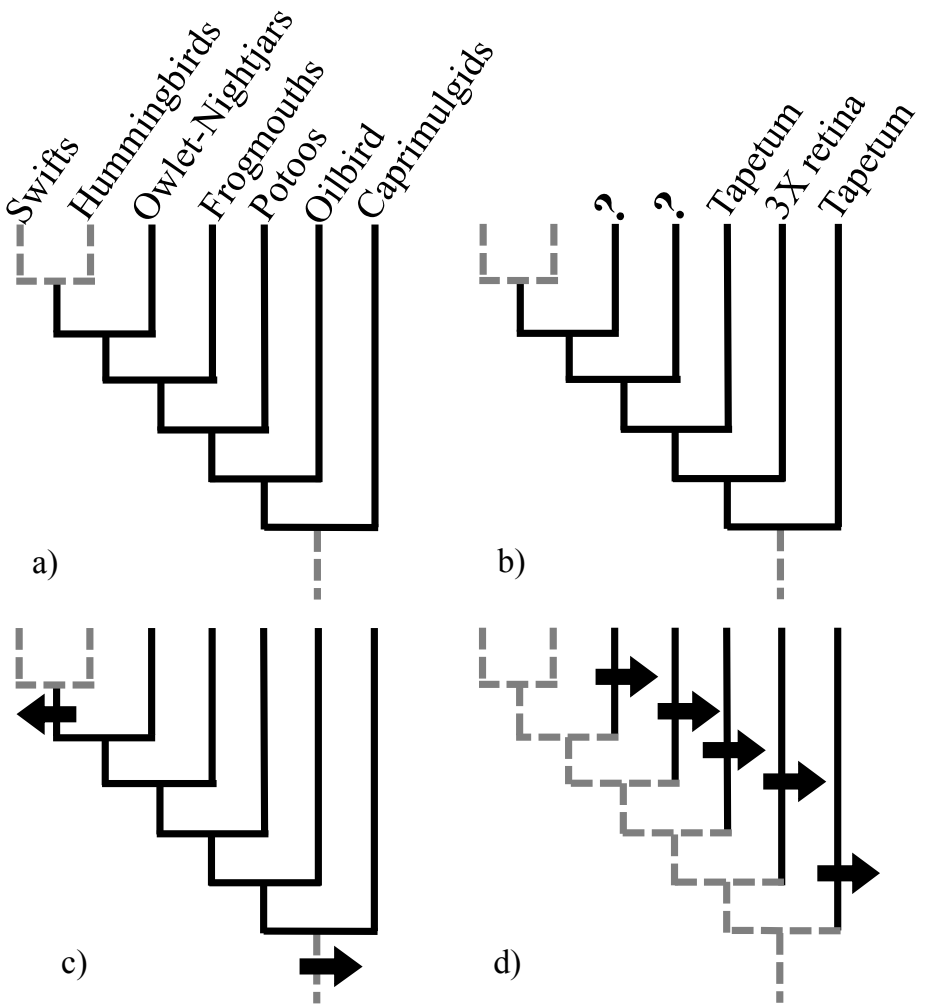


Figure 3.10. Evolutionary scenarios for visual adaptations to nocturnality in *Strisores*. (a) Best estimate of the *Strisores* backbone phylogeny based on UCE data. (b) Visual adaptations to nocturnality are known in the oilbird, potoos and caprimulgids, but have not been studied in the frogmouths or owlet-nightjars. (c, d) The topology of the *Strisores* phylogeny indicates two major alternative hypotheses of evolutionary origin of nocturnality. Transition events indicated with arrows; oriented right=diurnal to nocturnal, oriented left=nocturnal to diurnal. In all trees, taxa are arranged as in a), and diurnal lineages are depicted with grey dashed lines. A similar figure has appeared in Mayr (2010).


Swifts
Hummingbirds
Owlet-Nightjars
Frogmouths
Potoos
Oilbird
Caprimulgids
a)
?
?
Tapetum
3X retina
Tapetum
b)
c)
d)

# Chapter 4: A Novel Tool for Enrichment of Phototransduction Cascade Genes across Aves

## Introduction

Birds (Aves) comprise roughly 10,000 species spanning environments (forest, open grassland, aquatic, polar) and lifestyles (diurnal, nocturnal, crepuscular—active at dawn and dusk, cathemeral—active throughout a 24-hour period). The diversity of birds presents an excellent system with which to investigate the evolutionary history of a wide variety of traits, and the mechanisms that may have allowed for species to transition from one lifestyle to another. Birds are a particularly good study system for investigating the evolution of nocturnality, as there have been multiple transitions within birds between a diurnal and nocturnal, or crepuscular, lifestyle.

The greatest difficulty for visually oriented organisms (such as birds) transitioning between a day-living and night-living lifestyle is the inability to see in low-light conditions. Unraveling the evolutionary history of nocturnal vision will allow inference of the evolution of nocturnality itself. For the purposes of my research, I define a nocturnal bird as one that is doing visually demanding tasks (such as foraging or hunting) primarily in low-light conditions, throughout the year, excluding cathemeral species. Contrary to the organisms that suffer a loss of vision, such as the much-studied *Astyanax* blind cave fish (*Astyanax mexicanus*), the transition of a bird into a nocturnal lifestyle represents an adaptation to enhanced sensitivity to light while maintaining visual acuity. For example, many nocturnal birds hunt insects in flight, a demonstration of their high visual function. In addition, a nocturnal bird's eyes are very different morphologically from those of a diurnal

bird; they are larger and more conically shaped, indicating functional adaptation (Walls 1963).

There are only two species-rich radiations of nocturnal birds, owls (*Strigiformes*) and nightbirds (*Strisores*, previously *Caprimulgiformes*). Little is known about nocturnal vision in birds, though some study has been conducted in *Strisores*, which contains both nocturnal and crepuscular (active at dawn and dusk) members. The eyes of two lineages of *Strisores* have been characterized morphologically, and each displays remarkable adaptations to increase light sensitivity. The oilbird (*Steatornis caripensis*) has a unique retinal structure with three layers of photoreceptor cells, and the highest photoreceptor cell density of any vertebrate studied (Martin et al. 2004; Rojas et al. 2004). The caprimulgids (nightjars and nighthawks; Caprimulgidae) have a tapetum lucidum—a mirror-like structure at the back of the eye that reflects light back through retinal cells, providing a second chance for the absorption of incoming photons (Nicol & Arnott, 1974). Tapeta lucida have arisen independently in many nocturnal vertebrates, and their presence can be observed by the brilliant reflection of their eyes in a spotlight at night, often visible from a great distance. The biochemical or structural composition of tapeta lucida are quite varied, including crystals, lipid droplets, specialized endothelial cells and extracellular fibrils (Walls 1963; Rodieck 1973). Potoos (Nyctibiidae) display a brilliant eye shine, indicating the presence of a tapetum, but the biochemical composition of their tapetum has not been studied, and may have arisen independently of the lipid-based tapetum present in caprimulgids (Martin et al. 2004). Indeed, the topology of my current best estimate of the tree suggests an independent

origin (see Chapter 3). Frogmouths and owlet-nightjars (Aegothelidae) do not have a brilliant eyeshine, so must not have tapeta, but their visual adaptations have not been studied further.

The nightbirds are a particularly interesting group in which to study night vision because the clade subsumes an order of diurnal birds, previously known as the *Apodiformes* (swifts and hummingbirds). By using bird specimens and genetic resources available in museum collections and targeting the molecular basis of vision, I can expand my study to include all instances of nocturnality present on the avian tree of life. Using the definition of nocturnality mentioned above, I have identified eight instances of nocturnality across Aves (Figure 4.1).

Molecular analyses of vision-related genes have a rich history of comparative work across vertebrates. There is also some representation within birds, though that work is incomplete, unevenly dispersed taxonomically and particularly lacking in nocturnal species. In order to elucidate the visual adaptations in birds, I intend to conduct molecular evolutionary analyses of vision by studying the sequence of all genes involved in the phototransduction cascade from both nocturnal species and their diurnal relatives. In order to conduct this work efficiently, I have designed a sequence capture bait set that will enrich a pool of genomic DNA for these targeted genes prior to next generation sequencing.

When light is absorbed by a chromophore, it causes a structural change in the surrounding opsin that triggers enzymatic activity, beginning a biochemical cascade that leads to the release of neurotransmitter from the photoreceptor (Pugh et al. 1999). This is called the phototransduction cascade (PTC). Phototransduction is the

conversion of an optical signal into a neural response. It involves both an activation phase and a recovery or inactivation phase. Acronyms for the genes mentioned in Table 4.1 and the probe set described herein, are given in parentheses.

When an opsin is activated, it activates a G protein (transducin; GNAT, GNB, GNGT). The opsin catalyzes the exchange of GDP for GTP on the alpha subunit of the G protein, which then binds to phosphodiesterase (PDE6). Activated PDE6 hydrolyzes cyclic guanosine monophosphate (cGMP) to GMP, and the decrease in cytoplasmic cGMP causes the cyclic nucleotide gated channel (CNG) to close. The $Na^{+}/Ca^{2+}$ exchanger (SLC24A1) continues to pump $Ca^{2+}$ out of the cell, and the cell hyperpolarizes, generating the signal.

Inactivation of the response begins when the cytoplasmic $Ca^{2+}$ concentration becomes so low that it generates three different responses in the cell. 1) $Ca^{2+}$ dissociates from guanyly cyclase activating protein (GUCA1), allowing it to bind and activate guanyly cyclase, which regenerates cGMP. 2) Dissociation of $Ca^{2+}$ from calmodulin (CALM1) allows it to bind the ion channel and lower its half-activation concentration for cGMP. These two mechanisms cause the ion channels to reopen. 3) Low $Ca^{2+}$ concentration causes recoverin (RCVRN) to dissociate from the G protein-coupled receptor kinase (GRK). GRK phosphorylates the opsin, allowing arrestin (ARR3, SAG) to bind and stopping its activity. The final step in recovery is inactivation of the PDE6-G protein complex. This occurs when GTPase-activating protein (RGS, RGS9) hydrolyzes the G protein alpha subunit GTP to GDP, causing it to dissociate from PDE6.

Pugh et al. (1999) estimate that there are eight distinct molecular mechanisms involved in vertebrate photoreceptor light adaptation, and the authors identify a major protein of the PTC involved with each molecular mechanism. For example, increase in binding affinity of the guanylyl cyclase substrate binding pocket for GMP would increase the rate of cGMP synthesis, causing an extension of the photoreceptor operating range, and an increased sensitivity to photons. This is a potential mechanism by which a retina could adapt to low-light environments and similar mechanisms could be used to alter the response of cones versus rods.

While all components of the PTC are necessary for full generation of a visual response, the opsins can be regarded as the molecular gatekeepers of the reaction. Opsins are transmembrane proteins that surround a vitamin chromophore. Changes in the sequence of an opsin can cause shifts in the polarity of its bond to the chromophore, and thereby alter the wavelengths of light that are optimally absorbed by the chromophore and initiate the PTC. This phenomenon is called "spectral tuning," and allows an organism to adapt to the environment in which it is most visually active (Govardovskii et al. 2000; Wilkie et al. 2000). Spectral tuning sites in vertebrates are well characterized, and an observed amino acid change in an opsin sequence can indicate a shift in visual sensitivity (Takahashi & Ebrey, 2003; Yokoyama et al. 2008; Hunt et al. 2009; Schott et al. 2017).

There are five spectral types of cone opsins present in birds, though any given bird only expresses four (most of the time, see Ödeen et al. 2008). They include short wavelength 1 (SWS1), short wavelength 2 (SWS2), medium wave length (MWS) and long wavelength (LWS). SWS1 comes in two varieties, UV-sensitive (UVS) or

violet-sensitive (VS), a shift that can be caused by as little as one amino acid replacement. Birds also have one rod opsin, rhodopsin, as well as double cones.

It has been noted in a few papers that avian opsins display a striking lack of variation across birds, due to the narrow range of wavelengths to which they are tuned (Imai et al. 1997; Wilkie et al. 2000; Yokoyama et al. 2000; Hart 2001; Ödeen & Håstad, 2013). These remarks are presumptive, when so little of avian diversity has been sampled. In particular, species from unique habitats have not been studied. For example, the tawny owl (*Strigiformes*; *Strix aluco*), is the only nocturnal bird to have had its opsin spectral sensitivities studied, even though these sensitivities could vary significantly in nocturnal birds versus diurnal, or within nocturnal lineages. In addition, there have been at least two attempts to detect an SWS1 opsin from the tawny owl both through PCR amplification and microspectrophotometry, but it is still unknown whether or not the species possesses a functional copy of the gene (Hart 2001; Ödeen 2003; Bowmaker 2008). A related example of such variation in birds is seen in the Humboldt penguin (*Spheniscus humboldti*), which is not nocturnal, but dives to a few hundred meters of ocean water, a depth at which < 10% of photons would be able to penetrate (Thewissen & Nummela, 2008). This penguin has an LWS absorbance of 543 nm, where the other 31 non-penguin birds studied to date range between 555-571 nm (Bowmaker & Martin, 1985; Hart 2001).

The vast majority of attention to avian opsins has been paid to the SWS1 opsin, as it varies between VS and UVS. In 2003, a study indicated that the ancestral avian SWS1 opsin was violet-sensitive, and that ultra-violet sensitivity was gained four times independently (Ödeen 2003; Bowmaker 2008). The authors updated their

taxon sampling in a 2013 publication, and now believe there were 14 such shifts, by at least 4 independent molecular mechanisms (Hart 2001; Bowmaker 2008; Ödeen & Håstad, 2013). The fact that additional taxa could alter the conclusions of the analysis so dramatically is an indication that a broad taxon sampling is needed. Analysis of the other five visual opsins is also likely to give insight into visual adaptation. A handful of MWS and LWS sequences have been published largely due to a study of bowerbirds, a single family of passerines (Coyle et al. 2012). There are thirteen published rhodopsin sequences for nocturnal birds. Those rhodopsin sequences have been used solely as markers in phylogenetic studies, and no analysis of their molecular evolution in relation to vision, or across birds, has been published (Brumfield et al. 2008; Hackett et al. 2008). An effort to use whole genome sequences to identify vision-related genes failed to recover sequences for two opsins in the budgerigar (*Melopsittacus undulates*) and turkey (*Meleagris gallopavo*), for which published microspectrophotometric data exists (Bowmaker et al. 1997, Hart et al. 1999, Borges et al. 2015).

Given the state of the field and the opportunities it presents, I have developed an in-solution sequence capture approach to efficiently enrich all PTC genes from any extant species of bird. Unlike PCR-based methods which have failed to retrieve full opsin sequence in the past (Ödeen 2003), sequence capture is easily scalable to multiple genes, multiple species, and more readily retrieves full gene sequences, including introns and flanking regions. This method is also not reliant on genome assembly methods, which is a rapidly developing, and improving, field. The data generated using this tool will enable more comprehensive molecular evolutionary

analyses of adaptation to nocturnal vision across Aves, filling in taxon sampling where it is most needed. Including all of the PTC genes will allow discovery of instances where a fully functional opsin gene is present, but a pathway it depends upon is shut down. Similar analyses in the Tokay gecko (*Gekko gekko*) revealed that it had achieved a cone-like physiology (responding to bright light environments) in a retina composed entirely of rods by co-opting the cone PTC genes (Zhang et al. 2006). Analysis of collected PTC gene sequences will give new insight into variation in the color vision and visual sensitivity of all bird lineages.

## Methods

### *Bait array design*

A total of 47 genes comprising the PTC pathway, as well as selected non-visual opsins, were chosen as target sequence for the bait array (Table 4.1). Functions for each gene were obtained from the Online Mendelian Inheritance in Man database <www.omim.org>. Sequences were downloaded from Ensembl (Herrero et al. 2016; Yates et al. 2016) and the National Center for Biotechnology Information's Genbank (Benson 2004) between April 19 and May 23 of 2016. All genes were mapped to the chicken genome using the UCSC genome browser (Kent 2002; Kent et al. 2002; Miller et al. 2007) for verification, both with chicken and other vertebrates. Messenger RNA (mRNA) sequences from Genbank were downloaded for those genes used in (Lamb et al. 2016), as they represent verified sequences (Table 4.1). Genomic DNA (gDNA) sequence was included in the target sequence set for genes with short introns (where baits would have overlapped). Ensembl annotations were used to note intron-exon boundaries for all genes in chicken and zebra finch. Individual exons

were mapped to the genomic sequence for each gene in Geneious (Biomatters, Ltd.). Any exons shorter than 120 bp were padded with intronic flanking sequence. Sequence for GNB3 was not found in chicken, and so is represented solely by zebra finch in the bait set. I was not able to find any sequence for GNGT1 or PDE6A in birds.

Bait design from the target sequence set began with generating bait sequences of 120 bp length, overlapping 60 bp so that all targeted gene sequences are covered twice (2X depth). These potential baits were aligned to the genomes of 10 birds representing avian phyletic diversity using BLAST+ (Camacho et al. 2009). These genomes included killdeer (*Charadrius vociferous*), chuck-will's-widow (*Caprimulgus carolinensis*), bald eagle (*Haliaeetus leucocephalus*), brown kiwi (*Apteryx australis*), downy woodpecker (*Picoides pubescens*), African ostrich (*Struthio camelus*), crested ibis (*Nipponia nippon*), rock pigeon (*Columba livia*), chicken and zebra finch. Potential bait sequences were subjected to masking using RepeatMasker <repeatmasker.org>.

An attempt was made to incorporate other genome-enabled taxa into the bait set design. Adding duck, flycatcher and turkey (available from Ensembl) did not significantly increase enrichment in *in silico* tests. Additionally, querying genomes published but not available in Ensembl revealed several instances of predicted proteins that didn't appear to share sequence identity with the chicken or zebra finch sequences, and thus all other taxa were excluded from further analysis. Lastly, taxa other than chicken and zebra finch displayed high variation in exon number for these conserved vision genes.

Baits were designed following the "relaxed filtering" protocol of MYcroarray MYbaits (MYcroarray, Inc.). In such, any potential bait sequences containing more than 10 bp of masked sequence were dropped. Remaining bait sequences were included if they had at most 10 BLAST+ matches with a melting temperature of 62.5-65°C, 4 matches above 65°C, and fewer than 2 passing baits on each flank (to avoid biasing recovery). The remaining 2,500 bait sequences were synthesized in RNA.

*Taxon sampling and data generation*

A total of 33 taxa were selected for this study, including both nocturnal and diurnal species, with a special emphasis on the *Strisores*, and one crocodilian outgroup (Table 4.2). Frozen tissue samples were assembled through loans from major museum collections. DNA was extracted using a phenol-chloroform protocol (Rosel and Block, 1996), and quality was assessed by agarose gel electrophoresis and fluorometric quantification. Samples were sheared to 200-500 bp via sonication, and Illumina libraries were prepared using KAPA Biosystems library preparation kits, attaching a unique single-index adapter to each taxon (KAPA Biosystems, Inc.). Taxa were combined into pools of 8-9 species for hybrid enrichment.

Enrichment followed protocols outlined in the MYcroarray MYbaits protocol (v. 3.02), with the exception that chicken Cot-1 DNA was used instead of human (BLOCK #1; Applied Genetics Laboratories, Inc.), and a custom adapter blocker matching the "TruSeq-style" Illumina adapter sequence was used instead of the provided BLOCK #3 (Integrated DNA Technologies, Inc.). Hybridization was carried out at 66°C for 22 hours.

Enrichment success was verified by qPCR. Five custom primer pairs, each targeting one exon sequence present in the bait set were designed, and used to quantify enrichment of those loci in enriched versus unenriched libraries. Average enrichment over those five loci was 534 X (224-690 X), indicating successful enrichment of targeted loci. Enriched pools of taxa were combined into one pool for sequencing with a paired-end 300 bp run of an Illumina MiSeq.

Read quality control and adapter sequence removal was performed using the illumiprocessor program (based on trimmomatic; Faircloth 2013b; Bolger et al. 2014). Reads were assembled *de novo* via Trinity (v 2.4.0; Grabherr et al. 2011), using the 'Assemblo' wrapper script available in the PHYLUCE software package (v 1.5.0; Faircloth 2015b). Custom Perl scripts were written to process the resulting contigs.

## Results

### *Assembly statistics*

An average of 484,065 reads were recovered per species (226,103 - 1,107,115; Table 4.3). These were assembled into an average of 55,997 contigs (14,177 - 139,689), with a mean length of 326 bp (291 – 409 bp). Mitochondrial genomes were assembled as off-target capture from most taxa, causing off-target contig lengths to have greater maximum values (avg. 14,748 bp) than PTC contigs (avg. 5,545 bp). Average length of off-target contigs (avg. 322 bp) was lower than PTC contigs (avg. 1,359 bp). This is likely due to the lower average read coverage of off-target contigs (avg. 4X) than PTC contigs (avg. 80X). The much higher coverage of PTC contigs than off-target contigs is another indication of successful enrichment.

On average, sequence (whole or partial) from 91% of the genes was recovered for every taxon (43 - 100%; Figure 4.2). Recovery of individual bait sequences was much lower per taxon (avg. 27%). Chicken recovered 75% of the bait sequences, which reflects that 71% of the baits were designed from that species. Low recovery of individual bait sequences across taxa is likely due to redundancy in the bait set, since sequence from two species was used in the design, as well as multiple sequence types (exons, introns, mRNAs, etc.). Almost half of the baits were only recovered in the *Gallus gallus* sample (1,011), and 175 baits were not recovered from any taxon.

Capture success was greatest in the chicken, which had the greatest amount of captured target sequence (and from which most of the bait sequences were designed; Table 4.4). The majority of taxa found baits that were generated from chicken sequence. Zebra finch (a member of *Passeriformes*) DNA was not available for this initial test, but two other members of the *Passeriformes* are present, *Fringilla montifringilla* and *Thamnophilus nigrocinereus*. As expected, both of these taxa show a greater percentage of found baits coming from zebra finch sequence than chicken sequence. Surprisingly, the same is true from *Lyncornis macrotis*, though this was not true for closely related taxa *Hydropsalis rufiventris* and *Eurostopodus mystacalis*.

It is logical to expect that increased sequence coverage would lead to increased recovery of targeted sequences. However, taxa *L. macrotis* and *H. rufiventris* both have roughly twice the sequence coverage as the other taxa in this study, but only one (*L. macrotis*) has higher than average recovery of PTC contigs. In fact, *H. rufiventris* has lower than average percent genes found for any bird taxa. There does not appear to be any difference in capture success between nocturnal and

diurnal taxa, nor are there any genes which appear to be totally missing exclusively from nocturnal taxa.

*Rhodopsin Sequence*

I recovered full rhodopsin coding sequence from 27 (of 32) ingroup species. One taxon (*Aegotheles cristatus*) did not recover exon 1, and three taxa (*Tyto alba*, *Amazona autumnalis* and *Hydropsalis rufiventris*) did not recover exons 2-5. The outgroup, alligator (*Alligator mississippiensis*), did not recover exon 5. Of the 27 taxa that recovered full coding sequence, they were assembled in 2-3 contigs. Very small gaps of missing sequence were noted between contigs that could likely be closed by increasing sequencing coverage, or adding sequence to the bait set that targets intronic sequence in that gapped region. Even incomplete, the recovery of the majority of intronic sequence from most species is a significant improvement over previously published PCR-based methods. With the complete, translated rhodopsin alignment it is possible to identify variation at known spectral tuning sites (Figure 4.3), as well as conduct analyses of gene structure.

## Future Directions

*Methodological Improvements*

First priorities in continuing this work are to streamline the bioinformatic pipeline for processing these data. Particular attention will be paid to improving the assembly process to increase the recovery of whole genes. Gene sequences generated in these analyses will be used in the next iteration of bait array design, to increase recovery of whole genes through incorporation of more diverse sequence.

Alteration of the hybridization reaction conditions is another target for improving sequence recovery. Lowering the hybridization temperature will allow more diverse sequences to bind the baits, and can be done incrementally to test for the optimal temperature to return full targeted sequence without increasing off-target capture. If necessary, other conditions of the reaction can also be altered (such as pH and ionic strength of the buffer), but this will likely not be necessary. I will also test increasing sequencing depth by simply pooling fewer taxa on a MiSeq run and looking for improved sequence recovery.

*Analyses*

Once the bait set is optimized and a dataset of PTC gene sequences is in hand, I will analyze the sequences for evidence of molecular evolution both between nocturnal and diurnal pairs and across nocturnal lineages. Examples of specific adaptations I may identify are:

*Amino acid replacement at spectral tuning sites* − Amino acid replacements at spectral tuning sites will be identified in alignments of opsin genes and can indicate: 1) spectral tuning of this opsin has occurred, or 2) this opsin is not under purifying selection and may be functionally unimportant. The latter scenario fits into the analyses of pseudogenization described later. The first scenario, that spectral tuning has occurred, has a number of potential evolutionary implications. I could see a mutation indicating spectral tuning common to lineages that share an origin of nocturnality, or of an ecological trait (such as use of echolocation). For example, a lineage of crepuscular birds may share an adaptation that spectrally tunes their short

wavelength opsin to the predominantly violet light environment present at dawn and dusk (Endler 1993). These phenomena will be investigated using phylogenetically corrected analyses, such as those implemented in Comparative Analyses of Phylogenetics and Evolution in R (available from < http://caper.r-forge.r-project.org/>).

*Evidence for selection* − I will use branch models in PAML (CODEML) to look for evidence of changes in the intensity of selection (as measured by the ratio of synonymous to non-synonymous substitutions, ω) for each PTC gene (Yang 2007). This analysis will be applied to specific branches of the avian tree, comparing nocturnal branches to diurnal ones, and testing assumptions of whether the ancestor was nocturnal or diurnal. I will also use a site-based model in PAML to determine which codon positions of a given gene may be targets of selection. Analyses of codon positions under selection will identify changes that are adaptive or caused by relaxed purifying selection, and thereby suggest plausible molecular mechanisms involved in nocturnal adaptation. These changes could occur at opsin spectral tuning sites, as a single point mutation could cause a functional change (Ödeen 2003; Ödeen et al. 2012).

*Gene Presence/Absence or Pseudogenization* − Available comparative data from mammals indicates that variation between nocturnal and diurnal species may be observed in the shortest wavelength (SWS) opsin (Meredith et al. 2013). In primates, pseudogenization of the SWS opsin has occurred by multiple mechanisms, including indels, resulting in frame shifts and premature stop codons, as well as deletion of splice sites, point mutations, and nonsense mutations (Tan et al. 2005). In bats, indels

and premature stop codons rendered the SWS opsin nonfunctional in certain lineages (Zhao et al. 2009). Alignments of avian PTC genes will be searched for evidence of pseudogenization by any potential mechanism. A shared mechanism between closely related taxa would be consistent with a single origin of nocturnality, whereas the same mutation shared between disparate taxa would suggest convergent origins. Loci that have some predisposition for a certain mutation, such as the presence of repetitive sequence that leads to slippage of the replication machinery, may also be discovered by these analyses.

*Potential Outcomes*

Functional relaxation and even loss of the short-wavelength opsin gene (mammals only have one) was found concurrent with putatively independent transitions to nocturnality in primates, bats, and a similar transition in the terrestrial ancestor of deep-diving marine mammals (Tan 2005; Zhao et al. 2009; Meredith et al. 2013). For this reason, and the already reported diversity of SWS1 opsins across birds, I might expect to see loss of SWS1 opsins in nocturnal birds. In instances where a UV SWS1 gene is present and functional, I could see a shift in the spectral tuning of the SWS2 to partially cover the VS range, as has been noted in some birds (Hart 2001). A similar shift in the MWS spectrum in the instance of a loss of SWS2 has never been documented, but would serve a similar function and could be hypothesized to occur. LWS opsins, while also showing very little variation across birds, are a logical target for study (Hart 2001). Ambient light at sunset and sunrise is predominantly long-wavelength, and thus would be perceived by the LWS opsins (Endler 1993). Strictly nocturnal organisms may not yet be active at sunset, or may

have ceased activity before sunrise, but crepuscular or cathemeral organisms could be active during this light environment. In those lineages, I would expect a functional LWS opsin to be present, and could see shared spectral tuning coincident with different lifestyles. For example, within *Strisores*, I may see variation between the crepuscular nighthawks, which are very active around dusk and the more fully nocturnal oilbird, which is active during moonrise (Snow 1961; 1962). There is not enough known about the other visual opsins to make strong predictions about how they will vary between nocturnal and diurnal lineages.

**Discussion**

The PTC probe set developed and tested here has have given some first glimpses of into visual system evolution across the living diversity of birds. After additional optimization, these data will allow inference of issues related to adaptation and sexual selection, and will contribute to analysis of the underlying causes of avian diversification. Lythgoe (1979) first suggested that the visual systems of birds likely evolved first and foremost to meet ecological needs−namely foraging and predator avoidance−and other consequences, such as ornate plumage decoration, evolved to take advantage of capabilities that already existed (Hart 2001). Evidence of this was seen in a study of bowerbirds, where no link was found between differential tuning of visual pigments and variation in display color (Coyle et al. 2012). Supporting this idea, work in bats has shown variation in visual pigments coincident with ecological variation. Losses of the SWS opsin in bats was found coincident with the origin of high-duty-cycle echolocation and changes in roosting behavior, including a loss in cave-roosting versus tree-roosting bats, a pattern that warrants investigation in the

cave-dwelling oilbird versus other nightbirds (Zhao et al. 2009). More recently, analyses of expression have indicated that opsin evolution in birds was driven by habitat and sexual selection on plumage (Bloch 2015).

By sequencing and analyzing the full complement of phototransduction genes, I can investigate these different scenarios, and elucidate the evolutionary history of opsin evolution in birds. This work will uncover adaptations that enabled the ecological diversification of birds and contributed to the great biodiversity seen today.

**Acknowledgements**

This work was funded through a National Science Foundation Doctoral Dissertation Improvement Grant awarded to White (award number 1501796).

## Tables

Table 4.1. Genes targeted by the bait array designed in this study. * = the rod gamma G protein subunit appears to be missing in birds. ** = the rod alpha PDE6 subunit appears to be missing in birds. $^{M}$ = mRNA sequence included in bait set. $^{G}$ = gDNA sequence included in bait set. cGMP = cyclic guanosine monophosphate. GPCR = G protein-coupled receptor (GPCR).

| Gene | Function |
|---|---|
| Short wave sensitive opsin 1 (SWS1)$^{M,G}$ | Cone opsin |
| Short wave sensitive opsin 2 (SWS2)$^{M}$ | Cone opsin |
| Medium wave sensitive opsin (MWS)$^{M}$ | Cone opsin |
| Long wave sensitive opsin (LWS)$^{M}$ | Cone opsin |
| Rhodopsin (RHO)$^{M}$ | Rod opsin |
| G protein alpha polypeptide 1 (GNAT1)$^{M}$ | Alpha subunit of transducin in rods (1/3)* |
| G protein alpha polypeptide 2 (GNAT2)$^{M}$ | Alpha subunit of transducin in cones (1/3) |
| G protein beta polypeptide 1 (GNB1) | Beta subunit of transducin in rods (2/3)* |
| G protein beta polypeptide 3 (GNB3 ) | Beta subunit of transducin in cones (2/3) |
| G protein beta polypeptide 4 (GNB4) | Transducin, unknown location |
| G protein gamma polypeptide 2 (GNGT2) | Gamma subunit of transducin in cones (3/3) |
| Phosphodiesterase catalytic subunit B (PDE6B)$^{M}$ | Phosphodiesterase catalytic subunit in rods (1/3)** |
| Phosphodiesterase catalytic subunit C (PDE6C)$^{M}$ | Phosphodiesterase catalytic subunit in cones (1/2) |
| Phosphodiesterase inhibitory subunit G (PDE6G)$^{G}$ | Phosphodiesterase inhibitory subunit in rods (2/3)** |
| Phosphodiesterase inhibitory subunit H (PDE6H) | Phosphodiesterase inhibitory subunit in cones (2/2) |
| Cyclic nucleotide gated | Alpha subunit of the photoreceptor cyclic nucleotide-gated |

| | |
|---|---|
| channel alpha 1 (CNGA1) | channel in rods (1/2) |
| Cyclic nucleotide gated channel alpha 3 (CNGA3) | Alpha subunit of the photoreceptor cyclic nucleotide-gated channel in cones (1/2) |
| Cyclic nucleotide gated channel beta 1 (CNGB1)[G] | Beta subunit of the photoreceptor cyclic nucleotide-gated channel in rods (2/2) |
| Cyclic nucleotide gated channel beta 3 (CNGB3) | Beta subunit of the photoreceptor cyclic nucleotide-gated channel in cones (2/2) |
| Sodium/potassium/calcium exchanger member 1 (SLC24A1) | $Na^+/Ca^{2+}$ exchanger in rods |
| Sodium/potassium/calcium exchanger member 2 (SLC24A2) | $Na^+/Ca^{2+}$ exchanger in cones |
| Guanylate cyclase activator 1A (GUCA1A) | Activates guanylyl cyclase; stimulates synthesis of cGMP |
| Guanylate cyclase activator 1B (GUCA1B) | Activates guanylyl cyclase; stimulates synthesis of cGMP |
| Guanylate cyclase activator 1C (GUCA1C) | Activates guanylyl cyclase; stimulates synthesis of cGMP |
| Calmodulin 1 (CALM1) | Lowers the cyclic nucleotide-gated channel's half-activation concentration for cGMP |
| Recoverin (RCVRN) | Inhibits GRK during light response |
| G protein-coupled receptor kinase 1 (GRK1)[M] | Opsin phosphorylation in rods |
| G protein-coupled receptor kinase 7 (GRK7) | Opsin phosphorylation in cones |
| G protein-coupled receptor kinase 2 (GRK2)[G] | Unknown |
| Arrestin (ARR3)[G] | Cone arrestin |
| S-antigen (SAG) | Rod arrestin |
| Regulator of G-protein signaling 9 (RGS9) | GTPase, deactivates PDE6 |
| Regulator of G-protein signaling 9 binding protein (RGS9BP) | Activates RGS9 |
| Teleost multiple tissue opsin 2 (TMT2) | Nonvisual opsin |
| Vertebrate ancient opsin (VA)[M] | Nonvisual opsin |
| Pineal Opsin (OPNP)[M] | Nonvisual opsin |
| Encephalopsin (OPN3) | Nonvisual opsin |
| Melanopsin (OPN4) | Nonvisual opsin |
| Melanopsin-like photopigment (OPN4-1) | Unknown |
| Neuropsin (OPN5) | Member of the rhodopsin family of GPCRs |
| Opsin 5-like 1 (OPN5L1) | Unknown |
| Opsin 5-like 2 (OPN5L2) | Unknown |

| | |
|---|---|
| Retinal G protein coupled receptor (RGR) | Functions in the light-dependent pathway of the rod visual cycle |
| Retinal pigment epithelium-derived rhodopsin homolog (RRH) | Encodes a visual pigment-like GPCR localized in the retinal pigment epithelium |
| Gustducin (GNAT3)[M] | Gustducin's role in taste transduction is analogous to that of transducin in light transduction; expressed in photoreceptors in lizard parietal eye |
| Nuclear receptor subfamily 2 group E member 3 (NR2E3) | Retinal nuclear receptor; a ligand-dependent transcription factor |
| Phosducin (PDC) | Phosphoprotein expressed in photoreceptors and in pinealocytes of the pineal gland |

Table 4.2. Catalog numbers for tissues used in this study. Taxa shaded in grey are nocturnal. Collection details for all specimens used in this study. Abbreviations: ABTC = Australian Biological Tissue Collection (South Australian Museum, Adelaide); ANSP = Academy of Natural Sciences, Philadelphia; ANWC = Australian National Wildlife Collection, Canberra; FMNH = Field Museum of Natural History, Chicago; KUNHM = Univ. of Kansas Natural History Museum, Lawrence; LSUMZ = Louisiana State Univ. Museum of Zoology, Baton Rouge; WLK = preparator L. K. Wang; USNM = US National Museum of Natural History, Washington, DC. * = previously known as ANSP 4467.

| Order | Common Name | Species | Catalog No. |
|---|---|---|---|
| *Apterygiformes* | Southern Brown Kiwi | *Apteryx australis* | LSUMZ B8606 |
| *Cathartiformes* | Turkey Vulture | *Cathartes aura* | LSUMZ B17242 |
| *Charadriiformes* | Grey Pratincole | *Glareola cinerea* | FMNH 429393 |
| *Charadriiformes* | American Oystercatcher | *Haematopus palliatus* | LSUMZ B26383 |
| *Columbiformes* | Pink-necked Green Pigeon | *Treron vernans* | LSUMZ B47229 |
| *Eurypygiformes* | Sunbittern | *Eurypyga helias* | LSUMZ B1980 |
| *Galliformes* | Red junglefowl | *Gallus gallus* | LSUMZ B36208 |
| *Gaviiformes* | Common Loon | *Gavia immer* | LSUMZ B7923 |
| *Gruiformes* | Virginia Rail | *Rallus limicola* | FMNH 380034 |
| *Passeriformes* | Brambling | *Fringilla montifringilla* | USNM B10165 |
| *Passeriformes* | Blackish-Grey Antshrike | *Thamnophilus nigrocinereus* | LSUMZ B20234 |
| *Pelecaniformes* | Shoebill | *Belaeniceps rex* | LSUMZ B13372 |
| *Pelecaniformes* | Boat-billed Heron | *Cochlearius cochlearius* | USNM 621016 |
| *Psittaciformes* | Red-lored Amazon | *Amazona autumnalis* | USNM 606742 |
| *Strigiformes* | Oriental Bay Owl | *Phodilus badius* | AMNH PRS1614 |
| *Strigiformes* | Burrowing Owl | *Speotyto cunicularia* | FMNH 396871 |
| *Strigiformes* | Barn Owl | *Tyto alba* | LSUMZ B19295 |
| *Strisores* | Australian owlet-nightjar | *Aegotheles cristatus* | ABTC 24643 |
| *Strisores* | Feline owlet-nightjar | *Aegotheles insignis* | KUNHM 5081 |
| *Strisores* | Uniform swiftlet | *Aerodramus vanikorensis* | USNM B04039 |
| *Strisores* | Philippine frogmouth | *Batrachostomus septimus* | FMNH 429205 |
| *Strisores* | Sparkling violetear | *Colibri coruscans* | LSUMZ B5574 |
| *Strisores* | White-throated nightjar | *Eurostopodus mystacalis* | ANWC B40813 |
| *Strisores* | Moustached treeswift | *Hemiprocne mystacea* | WLK 385 |
| *Strisores* | Rufous-bellied nighthawk | *Lurocalis rufiventris* | ANSP 19154* |
| *Strisores* | Great-eared nightjar | *Lyncornis macrotis* | USNM B03732 |
| *Strisores* | Great potoo | *Nyctibius grandis* | LSUMZ B15415 |
| *Strisores* | Tawny frogmouth | *Podargus strigoides* | USNM B06435 |
| *Strisores* | Oilbird | *Steatornis caripensis* | LSUMZ B32579 |
| *Strisores* | White-collared swift | *Streptoprocne zonaris* | USNM B13064 |
| *Strisores* | Crimson topaz | *Topaza pella* | USNM B11959 |
| *Tinamiformes* | Elegant Crested Tinamou | *Eudromia elegans* | LSUMZ B5893 |
| Outgroup | American Alligator | *Alligator mississippiensis* | LSUMZ-HSC |

Table 4.3. Assembly statistics. Nocturnal taxa are shaded in grey. Each column is shaded from low values (white) to high values (dark green). Minimum (-), maximum (+), and average (/) values are displayed. PTC = phototransduction cascade.

| Taxon | # Reads | # Contigs | # PTC Contigs | Off-Target Contig Length | | | PTC Contig Length | | | Off-Target Coverage | | | PTC Coverage | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | - | + | / | - | + | / | - | + | / | - | + | / |
| *Tyto alba* | 465,848 | 58,913 | 151 | 224 | 16,056 | 315 | 262 | 5,220 | 1,397 | 0 | 857 | 3 | 2 | 275 | 84 |
| *Phodilus badius* | 439,287 | 54,930 | 156 | 224 | 10,795 | 321 | 254 | 5,098 | 1,402 | 0 | 1,082 | 4 | 3 | 373 | 105 |
| *Speotyto cunicularia* | 313,998 | 36,869 | 156 | 224 | 13,370 | 317 | 242 | 6,097 | 1,272 | 0 | 953 | 4 | 2 | 968 | 79 |
| *Cathartes aura* | 412,088 | 36,770 | 179 | 224 | 17,125 | 340 | 288 | 5,829 | 1,422 | 0 | 1,984 | 5 | 2 | 188 | 66 |
| *Fringilla montifringilla* | 672,907 | 97,287 | 192 | 224 | 16,840 | 303 | 242 | 6,550 | 1,386 | 0 | 628 | 3 | 2 | 196 | 57 |
| *Thamnophilus nigrocinereus* | 397,539 | 53,188 | 157 | 224 | 10,795 | 305 | 311 | 5,926 | 1,206 | 0 | 645 | 3 | 4 | 186 | 69 |
| *Amazona autumnalis* | 328,609 | 37,647 | 145 | 224 | 18,232 | 328 | 347 | 5,510 | 1,555 | 0 | 314 | 4 | 9 | 306 | 95 |
| *Belaeniceps rex* | 256,325 | 24,264 | 166 | 224 | 10,932 | 328 | 371 | 8,342 | 1,374 | 0 | 929 | 4 | 4 | 167 | 60 |
| *Cochlearius cochlearius* | 470,662 | 46,871 | 226 | 224 | 10,795 | 325 | 245 | 5,552 | 995 | 0 | 723 | 4 | 0 | 230 | 41 |
| *Gavia immer* | 502,420 | 79,848 | 178 | 224 | 24,294 | 301 | 386 | 5,563 | 1,308 | 0 | 1,867 | 3 | 5 | 140 | 51 |
| *Eurypyga helias* | 642,325 | 80,157 | 166 | 224 | 16,870 | 289 | 251 | 4,691 | 1,182 | 0 | 1,131 | 3 | 5 | 127 | 43 |
| *Streptoprocne zonaris* | 293,195 | 34,423 | 148 | 224 | 12,469 | 318 | 278 | 5,096 | 1,227 | 0 | 365 | 4 | 2 | 238 | 67 |
| *Aerodramus vanikorensis* | 623,642 | 59,577 | 127 | 224 | 15,493 | 347 | 312 | 5,110 | 1,535 | 0 | 541 | 5 | 3 | 396 | 154 |
| *Hemiprocne mystacea* | 595,129 | 57,379 | 145 | 224 | 17,720 | 326 | 294 | 6,746 | 1,407 | 0 | 6,489 | 5 | 3 | 371 | 107 |
| *Colibri coruscans* | 789,285 | 81,360 | 132 | 224 | 18,386 | 330 | 325 | 7,559 | 1,705 | 0 | 2,050 | 4 | 8 | 326 | 116 |
| *Topaza pella* | 472,874 | 54,301 | 146 | 224 | 15,470 | 314 | 229 | 4,141 | 1,290 | 0 | 424 | 4 | 2 | 215 | 73 |
| *Aegotheles cristatus* | 395,920 | 35,690 | 143 | 224 | 16,801 | 347 | 303 | 5,362 | 1,638 | 0 | 500 | 4 | 5 | 328 | 96 |
| *Aegotheles insignis* | 469,154 | 54,750 | 145 | 224 | 10,795 | 319 | 380 | 4,543 | 1,369 | 0 | 426 | 4 | 3 | 359 | 110 |
| *Batrachostomus septimus* | 250,422 | 14,177 | 151 | 224 | 10,795 | 395 | 364 | 7,047 | 1,653 | 0 | 257 | 4 | 1 | 437 | 59 |
| *Podargus strigoides* | 438,580 | 51,622 | 163 | 224 | 13,996 | 331 | 279 | 5,125 | 1,446 | 0 | 533 | 4 | 5 | 417 | 119 |
| *Nyctibius grandis* | 410,098 | 39,591 | 170 | 224 | 15,788 | 334 | 310 | 4,681 | 1,397 | 0 | 1,013 | 4 | 4 | 300 | 82 |
| *Steatornis caripensis* | 341,735 | 48,036 | 157 | 224 | 12,083 | 310 | 261 | 5,629 | 1,252 | 0 | 863 | 3 | 1 | 200 | 77 |
| *Lyncornis macrotis* | 1,066,134 | 139,689 | 347 | 224 | 17,831 | 321 | 253 | 5,896 | 1,410 | 0 | 421 | 4 | 0 | 392 | 85 |
| *Hydropsalis rufiventris* | 1,107,115 | 95,972 | 158 | 224 | 16,421 | 342 | 274 | 4,849 | 1,464 | 0 | 2,421 | 5 | 4 | 648 | 192 |
| *Eurostopodus mystacalis* | 611,414 | 89,301 | 179 | 224 | 15,394 | 302 | 275 | 4,144 | 1,189 | 0 | 2,942 | 3 | 0 | 165 | 50 |
| *Treron vernans* | 226,103 | 22,702 | 163 | 224 | 10,795 | 322 | 240 | 4,801 | 1,334 | 0 | 424 | 4 | 0 | 159 | 46 |
| *Glareola cinerea* | 316,701 | 46,089 | 178 | 224 | 15,151 | 295 | 229 | 5,318 | 1,041 | 0 | 332 | 3 | 2 | 117 | 36 |
| *Haematopus palliatus* | 277,218 | 38,068 | 193 | 224 | 15,653 | 296 | 248 | 6,601 | 1,080 | 0 | 190 | 3 | 1 | 99 | 38 |
| *Rallus limicola* | 570,072 | 56,968 | 165 | 224 | 15,793 | 330 | 285 | 4,834 | 1,411 | 0 | 1,344 | 4 | 2 | 257 | 85 |
| *Gallus gallus* | 602,588 | 93,161 | 255 | 224 | 21,350 | 295 | 233 | 7,406 | 1,472 | 0 | 177 | 3 | 0 | 270 | 81 |
| *Apteryx australis* | 589,007 | 61,336 | 125 | 224 | 10,795 | 331 | 291 | 4,985 | 1,320 | 0 | 2,295 | 4 | 11 | 262 | 97 |
| *Eudromia elegans* | 367,435 | 36,036 | 110 | 224 | 10,795 | 338 | 224 | 5,366 | 1,638 | 0 | 1,366 | 4 | 0 | 245 | 68 |
| *Alligator mississippiensis* | 258,310 | 30,916 | 35 | 224 | 10,795 | 325 | 230 | 3,354 | 1,066 | 0 | 155 | 3 | 2 | 84 | 38 |

Table 4.4. Recovery statistics of PTC baits and contigs. Percentage of baits recovered that were designed from chicken (Ggal) and zebra finch (Tgut) sequence is reported, both in the percentage of baits captured for each taxon (% Found) and percentage of possible baits (% Total). Percent targeted sequence recovery was calculated by mapping targeted sequences (not baits) to assembled contigs. Where multiple contigs are available for a target sequence, or contigs map to both chicken and zebra finch sequence, percentages can exceed 100. Nocturnal taxa are shaded in grey. Each column is shaded from low values (white) to high values (dark green). Ggal = chicken, Tgut = zebra finch.

| Taxon | % Reads | % Found baits Ggal | % Found baits Tgut | % Total Ggal Baits | % Total Tgut Baits | Length of captured sequence | % Targeted Sequence Recovered |
|---|---|---|---|---|---|---|---|
| *Tyto alba* | 3% | 63% | 37% | 22% | 34% | 77,002 | 57% |
| *Phodilus badius* | 3% | 63% | 37% | 23% | 34% | 79,151 | 59% |
| *Speotyto cunicularia* | 2% | 62% | 38% | 22% | 34% | 75,576 | 56% |
| *Cathartes aura* | 3% | 62% | 38% | 27% | 41% | 93,289 | 69% |
| *Fringilla montifringilla* | 4% | 33% | 67% | 16% | 82% | 105,493 | 78% |
| *Thamnophilus nigrocinereus* | 2% | 48% | 52% | 18% | 47% | 77,629 | 58% |
| *Amazona autumnalis* | 2% | 60% | 40% | 20% | 33% | 71,296 | 53% |
| *Belaeniceps rex* | 2% | 63% | 37% | 27% | 40% | 91,729 | 68% |
| *Cochlearius cochlearius* | 3% | 64% | 36% | 26% | 38% | 88,032 | 65% |
| *Gavia immer* | 3% | 63% | 37% | 27% | 39% | 90,773 | 67% |
| *Eurypyga helias* | 4% | 62% | 38% | 23% | 35% | 79,276 | 59% |
| *Streptoprocne zonaris* | 2% | 61% | 39% | 20% | 32% | 71,031 | 53% |
| *Aerodramus vanikorensis* | 4% | 61% | 39% | 20% | 31% | 68,516 | 51% |
| *Hemiprocne mystacea* | 4% | 61% | 39% | 21% | 34% | 74,226 | 55% |
| *Colibri coruscans* | 5% | 62% | 38% | 19% | 29% | 65,159 | 48% |
| *Topaza pella* | 3% | 61% | 39% | 19% | 31% | 67,421 | 50% |
| *Aegotheles cristatus* | 2% | 63% | 37% | 21% | 31% | 71,092 | 53% |
| *Aegotheles insignis* | 3% | 62% | 38% | 21% | 32% | 71,938 | 53% |
| *Batrachostomus septimus* | 2% | 63% | 37% | 23% | 33% | 76,781 | 57% |
| *Podargus strigoides* | 3% | 65% | 35% | 24% | 33% | 79,752 | 59% |
| *Nyctibius grandis* | 3% | 61% | 39% | 23% | 36% | 79,990 | 59% |
| *Steatornis caripensis* | 2% | 63% | 37% | 22% | 34% | 76,658 | 57% |
| *Lyncornis macrotis* | 7% | 43% | 57% | 25% | 83% | 124,491 | 92% |
| *Hydropsalis rufiventris* | 7% | 61% | 39% | 22% | 35% | 76,429 | 57% |
| *Eurostopodus mystacalis* | 4% | 62% | 38% | 23% | 36% | 80,962 | 60% |
| *Treron vernans* | 1% | 61% | 39% | 23% | 38% | 81,943 | 61% |
| *Glareola cinerea* | 2% | 62% | 38% | 24% | 38% | 83,970 | 62% |
| *Haematopus palliatus* | 2% | 63% | 37% | 27% | 40% | 92,367 | 69% |
| *Rallus limicola* | 4% | 63% | 37% | 22% | 32% | 74,228 | 55% |
| *Gallus gallus* | 4% | 90% | 10% | 94% | 26% | 223,890 | 166% |
| *Apteryx australis* | 4% | 67% | 33% | 18% | 23% | 58,841 | 44% |
| *Eudromia elegans* | 2% | 65% | 35% | 14% | 20% | 47,632 | 35% |
| *Alligator mississippiensis* | 2% | 66% | 34% | 4% | 5% | 13,518 | 10% |

**Figures**

Figure 4.1. Major nocturnal lineages (black lines) on the avian tree of life. Topology is a maximum likelihood phylogeny based on ultraconserved elements (E. Braun, B. Faircloth & N. White et al., *unpublished*) and rooted on crocodilians (not shown).

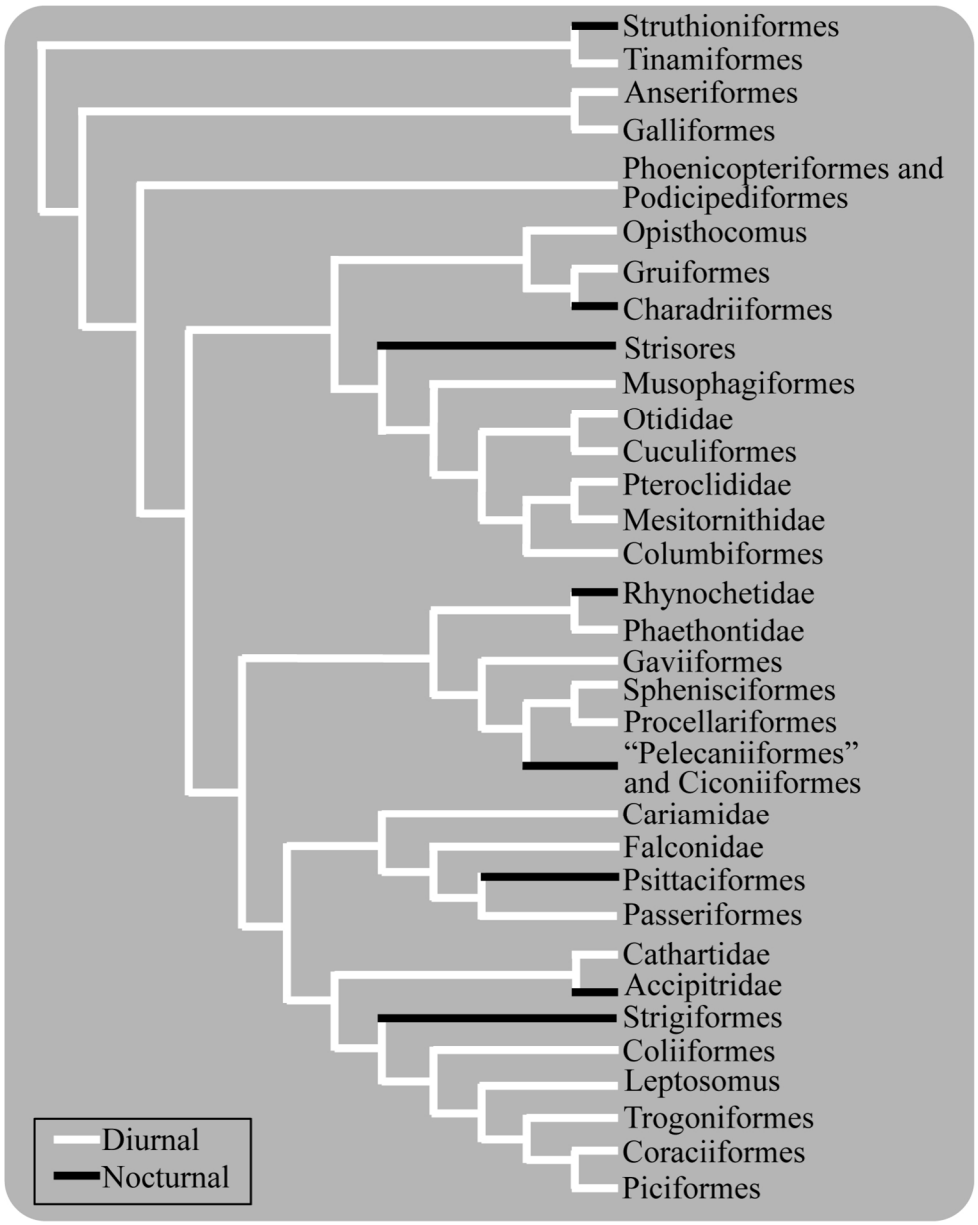
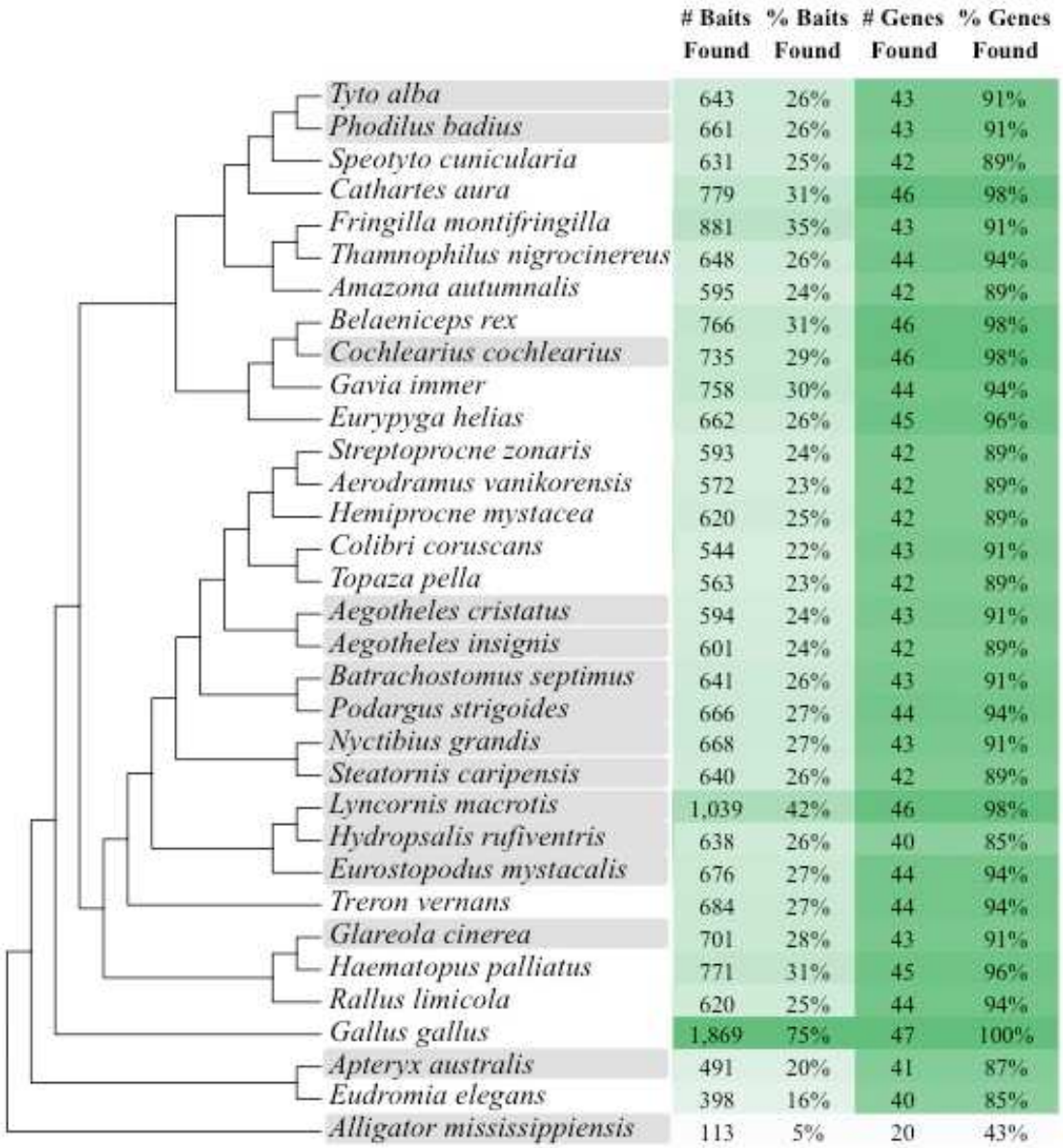
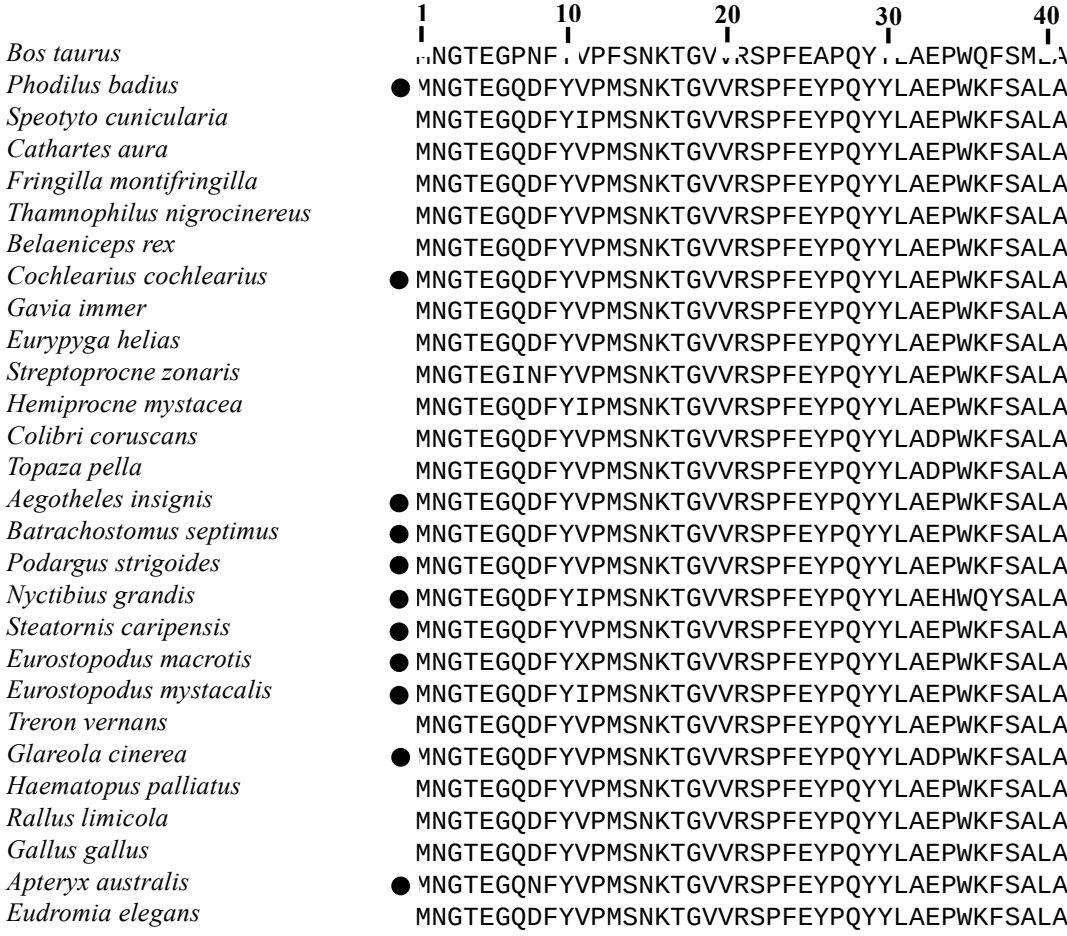


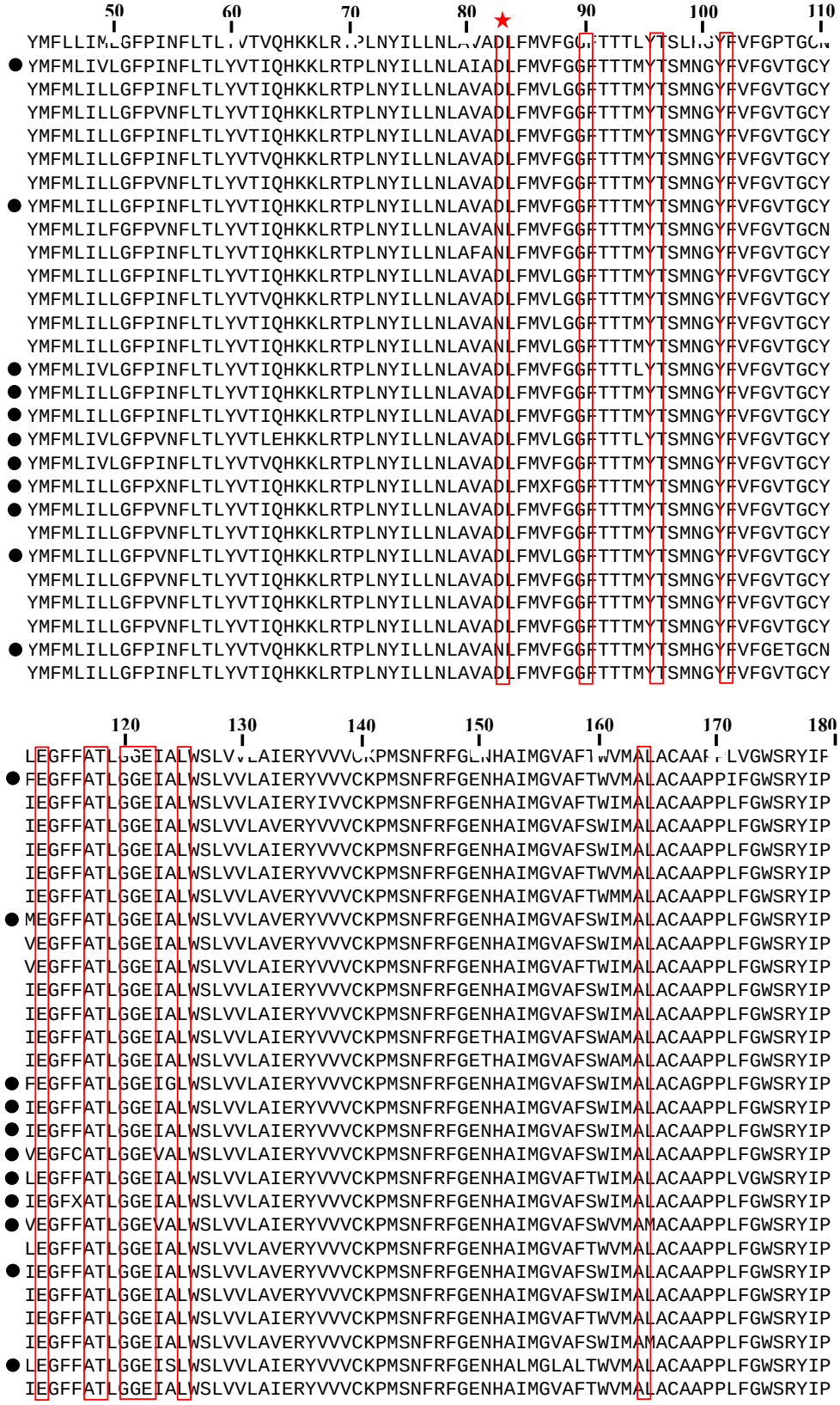
Figure 4.2. Number and percentage of baits and genes found for each taxon. Nocturnal taxa are shaded in grey. Each column is shaded from low values (white) to high values (dark green). Phylogeny is extracted from a maximum likelihood phylogeny based on ultraconserved elements (E. Braun, B. Faircloth & N. White et al., *unpublished*).

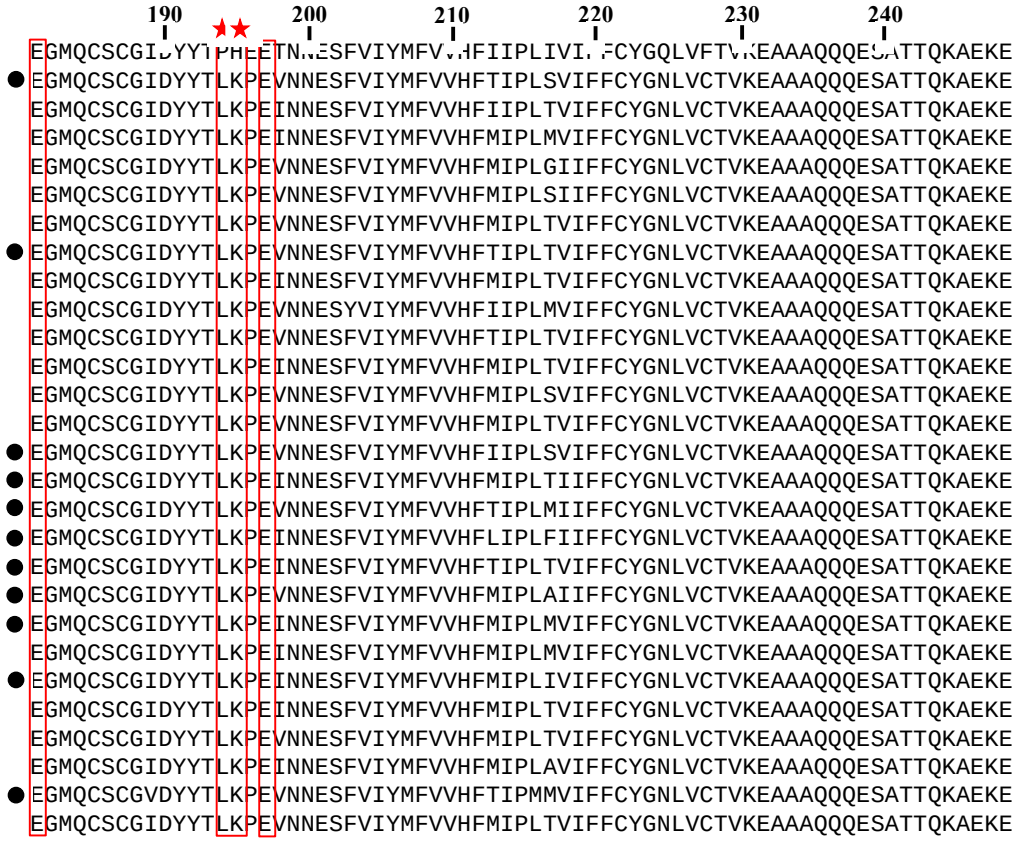



| | # Baits Found | % Baits Found | # Genes Found | % Genes Found |
|---|---|---|---|---|
| *Tyto alba* | 643 | 26% | 43 | 91% |
| *Phodilus badius* | 661 | 26% | 43 | 91% |
| *Speotyto cunicularia* | 631 | 25% | 42 | 89% |
| *Cathartes aura* | 779 | 31% | 46 | 98% |
| *Fringilla montifringilla* | 881 | 35% | 43 | 91% |
| *Thamnophilus nigrocinereus* | 648 | 26% | 44 | 94% |
| *Amazona autumnalis* | 595 | 24% | 42 | 89% |
| *Belaeniceps rex* | 766 | 31% | 46 | 98% |
| *Cochlearius cochlearius* | 735 | 29% | 46 | 98% |
| *Gavia immer* | 758 | 30% | 44 | 94% |
| *Eurypyga helias* | 662 | 26% | 45 | 96% |
| *Streptoprocne zonaris* | 593 | 24% | 42 | 89% |
| *Aerodramus vanikorensis* | 572 | 23% | 42 | 89% |
| *Hemiprocne mystacea* | 620 | 25% | 42 | 89% |
| *Colibri coruscans* | 544 | 22% | 43 | 91% |
| *Topaza pella* | 563 | 23% | 42 | 89% |
| *Aegotheles cristatus* | 594 | 24% | 43 | 91% |
| *Aegotheles insignis* | 601 | 24% | 42 | 89% |
| *Batrachostomus septimus* | 641 | 26% | 43 | 91% |
| *Podargus strigoides* | 666 | 27% | 44 | 94% |
| *Nyctibius grandis* | 668 | 27% | 43 | 91% |
| *Steatornis caripensis* | 640 | 26% | 42 | 89% |
| *Lyncornis macrotis* | 1,039 | 42% | 46 | 98% |
| *Hydropsalis rufiventris* | 638 | 26% | 40 | 85% |
| *Eurostopodus mystacalis* | 676 | 27% | 44 | 94% |
| *Treron vernans* | 684 | 27% | 44 | 94% |
| *Glareola cinerea* | 701 | 28% | 43 | 91% |
| *Haematopus palliatus* | 771 | 31% | 45 | 96% |
| *Rallus limicola* | 620 | 25% | 44 | 94% |
| *Gallus gallus* | 1,869 | 75% | 47 | 100% |
| *Apteryx australis* | 491 | 20% | 41 | 87% |
| *Eudromia elegans* | 398 | 16% | 40 | 85% |
| *Alligator mississippiensis* | 113 | 5% | 20 | 43% |

Figure 4.3. Translated alignment of captured rhodopsin sequence. The full coding sequence was recovered from 27 taxa, and aligned with cow (*Bos taurus*) rhodopsin for standard numbering of sites. Known spectral tuning sites are outlined in red, and sites displaying variation are indicated with a star above the column. Nocturnal taxa are indicated with a black dot at the beginning of their row. Numbers indicate base pair position in the alignment.




50
60
70
80
90
100
110
120
130
140
150
160
170
180


190
200
210
220
230
240
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFIIPLIVIFFCYGQLVFTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPLSVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFIIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLMVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLGIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLSIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESYVIYMFVVHFIIPLMVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLSVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFIIPLSVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLTIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFTIPLMIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFLIPLFIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFTIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLAIIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLMVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLMVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLIVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEINNESFVIYMFVVHFMIPLAVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGVDYYTLKPEVNNESFVIYMFVVHFTIPMMVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE
EGMQCSCGIDYYTLKPEVNNESFVIYMFVVHFMIPLTVIFFCYGNLVCTVKEAAAQQQESATTQKAEKE


250
260
270
280
290
300
310


320
330
340
350
TTLCCGKNPLGDDEAS---TTVSKTETSQVAPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSSSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDDDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSASKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA
TTLCCGKNPLGDEDTSAGKTETS--STSQVSPA
TTLCCGKNPLGDEDTSAGKTETSSVSTSQVSPA

# Supplementary Material (Ch. 3)

## Tables

Table S1. Collection details for all specimens used in this study. Localities appear as given in their respective collection databases. Abbreviations: ABTC = Australian Biological Tissue Collection (South Australian Museum, Adelaide); ANSP = Academy of Natural Sciences, Philadelphia; ANWC = Australian National Wildlife Collection, Canberra; FMNH = Field Museum of Natural History, Chicago; KUNHM = Univ. of Kansas Natural History Museum, Lawrence; LSUMZ = Louisiana State Univ. Museum of Zoology, Baton Rouge; WLK = preparator L. K. Wang; USNM = US National Museum of Natural History, Washington, DC; UWBM = Univ. of Washington, Burke Museum, Seattle; SAMA = South Australian Museum, Adelaide. * = previously known as ANSP 4467.

| Species | Tissue No. | Voucher No. | Locality | Collector |
|---|---|---|---|---|
| *Columba livia* | FMNH 428785 | FMNH 428785 | USA: Wisconson, Brown County, Green Bay | Bay Beach Wildlife Sanctuary |
| *Gallus gallus* | LSUMZ B36208 | LSUMZ 177895 | USA; Hawaii; Kauai Co.; Kauai District, Pioneer HiBred | USDA; prep. By Donna L. Dittmann |
| *Colibri coruscans* | LSUMZ B5574 | LSUMZ 116592 | PERU; Depto. San Martìn; 15 km by trail NE Jirillo on trail to Balsapuerto | T. J. Davis |
| *Lurocalis rufiventris* | ANSP 19154* | ANSP 185137 | ECUADOR; Zamora Chinchipe; Panguri, 12km NE San Francisco del Vergel; 4 deg 37 min S, 78 deg 58 min W, 1575m | T. J. Davis |
| *Anseranas semipalmata* | USNM B02954 | USNM 621019 | Captive | --- |
| *Streptoprocne zonaris* | USNM B13064 | USNM 626065 | GUYANA; Wiwitau Mountain, East Rupununi Savannah; 02 52 N, 059 16 W | --- |
| *Aegotheles cristatus* | ABTC 24643 | SAMA B39188 | AUSTRALIA; South Australia, 15 km from Lagoon Witt, Mabel Creek | --- |
| *Nyctibius griseus* | ANSP 18217 (B3238) | ANSP 183090 | ECUADOR; Prov. Sucumbios, Imuya Cocha; 0°34'S, 75°17'W, 200 m | F. Sornoza |
| *Eurostopodus mystacalis* | ANWC B40813 | ANWC B40813 | AUSTRALIA; Australian Capital Territory, Canberra; ~35°21'S, 149°058'E | --- |
| *Batrachostomus septimus* | FMNH 429205 | FMNH 429205 | PHILIPPINES; Luzon County, Kalinga; 17.4417N, 121.0708E | --- |
| *Nyctibius jamaicensis* | KUNH M 2116 | KUNHM 92957 | MEXICO; Silvituc, 24 km S, Mexico, Campeche; 18.233, -90.200 | --- |
| *Aegotheles insignis* | KUNH M 5081 | KUNHM 95997 | PAPUA NEW GUINEA; Simbu, 12.5 km ENE Haia, Aedo Camp., Crater Mt. Wildlife Management Area; 06 41'40 S, 145 06' 24 E | A. Mack |
| *Nyctibius aethereus* | LSUMZ B10877 | LSUMZ 156210 | PERU; Depto. Ucayali, SE slope of Cerro Tahuayo | A. S. Meyer |
| *Nyctibius grandis* | LSUMZ B15415 | LSUMZ 150545 | BOLIVIA; Depto. Santa Cruz; Prov. Velasco, Pre-Parque Nacional: "Noel Kempff Mercado," 30 km E of Aserradero Moira | J. M. Bates |
| *Nyctibius leucopterus* | LSUMZ B20315 | LSUMZ 165693 | BRAZIL; Amazonas, Munic. Manaus, km 34 ZF-3, Faz. Esteio, ~80 km N of Manaus | M. Cohn-Haft |
| *Nyctibius maculosus* | LSUMZ B271 | LSUMZ 97586 | PERU; Depto. Cajamarca, Lucuma on the Sapalache-Carmen Trail | M. J. Braun |
| *Steatornis caripensis* | LSUMZ B32579 | LSUMZ 169580 | PERU; Depto. Cajamarca; Quebrada Las Palmas, approximately 13 km WSW Chontali | R. C. Faucett |
| *Nyctibius bracteatus* | LSUMZ B4509 | LSUMZ 114641 | PERU; Depto. Loreto, lower Rio Napo region, E bank of Rio Yanayacu, ~90 km N of Iquitos | S. W. Cardiff |
| *Lyncornis macrotis* | USNM B03732 | USNM 607328 | PHILIPPINES; Luzon, Cagayan, Barrio Via, Sitio Hot Springs, Baggo Mun.; 1750 N, 12201 E | R. W. Dickerman and party |
| *Aerodramus vanikorensis* | USNM B04039 | USNM 608672 | PAPUA NEW GUINEA; New Ireland, New Ireland Province; 04 33 S, 152 57 E | J. P. Angle & B. M. Beehler |
| *Podargus strigoides* | USNM B06435 | USNM 612702 | AUSTRALIA; Queensland, Maryborough, Hervey Bay | F. Smith & T. Smith |
| *Topaza pella* | USNM | USNM | GUYANA; 1 39 N, 58 37 W; 225 meters; | --- |

| | | | | |
|---|---|---|---|---|
| | B11959 | 625397 | West Bank Upper Essequibo River, Gunn'S Landing | |
| *Hemiprocne mystacea* | WLK 385 | UWBM 68087 | SOLOMON ISLANDS; New Georgia Island, Arara; 8°29.5'S, 157°38.8'E (-8.492; 157.647) | C. E. Filardi |

Table S2. For each taxon, number of sequencing reads obtained, contig assembly length stats (`All Contigs Assembled` and `UCE Contigs Assembled`) and number of UCE loci averaged across incomplete datasets. Each column is shaded from low values (white) to high values (dark green).

| | | All Contigs Assembled | | | | UCE Contigs Assembled | | | | UCE Loci in Datasets | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Taxa** | **# Reads** | **#** | **Min** | **Max** | **Mean** | **#** | **Min** | **Max** | **Mean** | **Min** | **Max** | **Mean** |
| *Colibri coruscans* | 2,094,596 | 7,065 | 129 | 14,728 | 467 | 4,471 | 150 | 1,265 | 579 | 4,106 | 4,151 | 4,127 |
| *Topaza pella* | 2,059,428 | 5,382 | 137 | 12,347 | 525 | 4,271 | 179 | 1,474 | 567 | 4,065 | 4,098 | 4,083 |
| *Aerodramus vanikorensis* | 1,883,193 | 5,601 | 133 | 14,836 | 558 | 4,311 | 120 | 1,679 | 619 | 4,088 | 4,122 | 4,107 |
| *Streptoprocne zonaris* | 1,387,513 | 4,975 | 133 | 14,888 | 511 | 4,257 | 169 | 1,222 | 534 | 4,082 | 4,116 | 4,101 |
| *Hemiprocne mystacea* | 1,150,087 | 6,719 | 113 | 4,506 | 453 | 4,200 | 82 | 1,282 | 559 | 3,981 | 4,014 | 3,999 |
| *Aegotheles cristatus* | 2,995,105 | 7,406 | 137 | 16,907 | 554 | 4,418 | 136 | 1,741 | 705 | 4,059 | 4,106 | 4,081 |
| *Aegotheles insignis* | 1,399,898 | 4,933 | 133 | 15,633 | 442 | 4,179 | 171 | 1,034 | 460 | 4,009 | 4,030 | 4,020 |
| *Hydropsalis rufiventris* | 1,643,354 | 5,113 | 133 | 7,905 | 538 | 4,193 | 178 | 1,215 | 570 | 4,027 | 4,059 | 4,043 |
| *Lyncornis macrotis* | 1,505,696 | 5,485 | 133 | 15,154 | 501 | 4,265 | 133 | 1,582 | 549 | 4,048 | 4,080 | 4,065 |
| *Eurostopodus mystacalis* | 1,314,674 | 7,364 | 117 | 9,978 | 461 | 4,375 | 146 | 1,497 | 573 | 4,059 | 4,095 | 4,076 |
| *Nyctibius aethereus* | 1,320,945 | 7,455 | 121 | 15,316 | 432 | 4,369 | 123 | 1,575 | 555 | 4,067 | 4,105 | 4,089 |
| *Nyctibius bracteatus* | 2,198,584 | 4,951 | 137 | 15,694 | 432 | 3,698 | 137 | 1,281 | 473 | 3,614 | 3,617 | 3,616 |
| *Nyctibius grandis* | 1,305,354 | 7,940 | 117 | 8,423 | 407 | 4,287 | 84 | 1,446 | 553 | 4,042 | 4,078 | 4,059 |
| *Nyctibius griseus* | 1,342,176 | 8,618 | 117 | 13,724 | 384 | 4,362 | 135 | 1,332 | 524 | 4,089 | 4,119 | 4,104 |
| *Nyctibius jamaicensis* | 1,868,947 | 8,032 | 121 | 14,093 | 456 | 4,466 | 95 | 1,452 | 607 | 4,125 | 4,165 | 4,146 |
| *Nyctibius leucopterus* | 1,246,547 | 7,283 | 121 | 15,671 | 430 | 4,384 | 121 | 1,622 | 547 | 4,073 | 4,114 | 4,094 |
| *Nyctibius maculosus* | 1,484,550 | 6,880 | 125 | 15,679 | 432 | 4,321 | 131 | 1,432 | 535 | 4,064 | 4,103 | 4,084 |
| *Batrachostomus septimus* | 600,814 | 7,880 | 101 | 15,050 | 483 | 3,600 | 100 | 1,584 | 691 | 3,504 | 3,516 | 3,510 |
| *Podargus strigoides* | 1,344,558 | 5,014 | 133 | 14,033 | 506 | 4,193 | 171 | 1,681 | 534 | 4,012 | 4,046 | 4,030 |
| *Steatornis caripensis* | 3,507,074 | 6,729 | 129 | 15,668 | 491 | 4,461 | 129 | 1,686 | 590 | 4,137 | 4,182 | 4,160 |
| *Columba livia* | 5,497,426 | 8,838 | 141 | 15,650 | 433 | 4,583 | 163 | 1,606 | 603 | 4,115 | 4,165 | 4,139 |
| *Anseranas semipalmata* | 6,338,255 | 7,269 | 133 | 10,466 | 550 | 4,682 | 127 | 1,579 | 681 | 4,117 | 4,166 | 4,140 |
| *Gallus gallus* | 2,337,215 | 7,430 | 129 | 16,783 | 472 | 4,641 | 114 | 1,520 | 584 | 4,150 | 4,200 | 4,173 |

Table S3. Representation of each taxon on the nucleotide level. Calculated as the percentage of total alignment length for which a given taxon has nucleotides present (non-gap characters), averaged over all complete (a) or incomplete (b) datasets for each taxon set. Minimum (Min), maximum (Max) and average (Avg) values are given. Each column is shaded from low values (white) to high values (dark green).

a) **Complete**

| | **Taxon Set 1** | | | **Taxon Set 2** | | | **Taxon Set 3** | | | **Taxon Set 4** | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Taxon** | Min | Max | Avg | Min | Max | Avg | Min | Max | Avg | Min | Max | Avg |
| *Colibri coruscans* | 0.62 | 0.95 | 0.85 | 0.62 | 0.95 | 0.85 | 0.62 | 0.94 | 0.84 | 0.62 | 0.94 | 0.84 |
| *Topaza pella* | 0.61 | 0.95 | 0.85 | 0.60 | 0.95 | 0.84 | 0.60 | 0.95 | 0.84 | 0.60 | 0.95 | 0.84 |
| *Aerodramus vanikorensis* | 0.67 | 0.97 | 0.87 | 0.66 | 0.96 | 0.87 | 0.66 | 0.96 | 0.87 | 0.66 | 0.96 | 0.87 |
| *Streptoprocne zonaris* | 0.57 | 0.95 | 0.83 | 0.56 | 0.94 | 0.83 | 0.56 | 0.94 | 0.82 | 0.56 | 0.94 | 0.83 |
| *Hemiprocne mystacea* | 0.61 | 0.95 | 0.84 | 0.60 | 0.94 | 0.84 | 0.60 | 0.94 | 0.83 | 0.60 | 0.94 | 0.84 |
| *Aegotheles cristatus* | 0.76 | 0.96 | 0.89 | 0.76 | 0.96 | 0.89 | 0.75 | 0.96 | 0.89 | 0.75 | 0.96 | 0.89 |
| *Aegotheles insignis* | 0.49 | 0.89 | 0.76 | 0.48 | 0.88 | 0.76 | 0.48 | 0.87 | 0.75 | 0.48 | 0.88 | 0.75 |
| *Hydropsalis rufiventris* | 0.61 | 0.96 | 0.85 | 0.61 | 0.95 | 0.85 | 0.60 | 0.95 | 0.84 | 0.61 | 0.95 | 0.85 |
| *Lyncornis macrotis* | 0.59 | 0.95 | 0.84 | 0.58 | 0.94 | 0.84 | 0.58 | 0.94 | 0.83 | 0.58 | 0.94 | 0.83 |
| *Eurostopodus mystacalis* | 0.62 | 0.95 | 0.85 | 0.62 | 0.95 | 0.84 | 0.61 | 0.94 | 0.84 | 0.61 | 0.95 | 0.84 |
| *Nyctibius aethereus* | | | | 0.60 | 0.94 | 0.84 | 0.60 | 0.94 | 0.83 | 0.60 | 0.94 | 0.83 |
| *Nyctibius bracteatus* | 0.48 | 0.87 | 0.75 | | | | | | | 0.48 | 0.87 | 0.74 |
| *Nyctibius grandis* | 0.60 | 0.93 | 0.83 | 0.59 | 0.93 | 0.83 | 0.59 | 0.93 | 0.82 | 0.59 | 0.93 | 0.83 |
| *Nyctibius griseus* | | | | | | | 0.56 | 0.91 | 0.80 | 0.56 | 0.92 | 0.81 |
| *Nyctibius jamaicensis* | | | | | | | 0.65 | 0.96 | 0.87 | 0.65 | 0.96 | 0.87 |
| *Nyctibius leucopterus* | | | | | | | 0.59 | 0.94 | 0.83 | 0.59 | 0.94 | 0.83 |
| *Nyctibius maculosus* | | | | | | | 0.57 | 0.93 | 0.82 | 0.57 | 0.94 | 0.82 |
| *Batrachostomus septimus* | 0.71 | 0.91 | 0.84 | 0.71 | 0.91 | 0.83 | 0.70 | 0.90 | 0.83 | 0.71 | 0.91 | 0.83 |
| *Podargus strigoides* | 0.57 | 0.94 | 0.83 | 0.56 | 0.93 | 0.82 | 0.56 | 0.93 | 0.82 | 0.56 | 0.93 | 0.82 |
| *Steatornis caripensis* | 0.64 | 0.96 | 0.86 | 0.63 | 0.96 | 0.86 | 0.63 | 0.95 | 0.85 | 0.63 | 0.96 | 0.85 |
| *Columba livia* | 0.65 | 0.96 | 0.87 | 0.65 | 0.96 | 0.86 | 0.64 | 0.96 | 0.86 | 0.64 | 0.96 | 0.86 |
| *Anseranas semipalmata* | 0.73 | 0.97 | 0.89 | 0.73 | 0.97 | 0.89 | 0.73 | 0.97 | 0.89 | 0.72 | 0.97 | 0.89 |
| *Gallus gallus* | 0.63 | 0.96 | 0.85 | 0.63 | 0.96 | 0.85 | 0.62 | 0.95 | 0.84 | 0.62 | 0.95 | 0.85 |

b) **Incomplete**

| | Taxon Set 1 | | | Taxon Set 2 | | | Taxon Set 3 | | | Taxon Set 4 | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Taxon** | Min | Max | Avg | Min | Max | Avg | Min | Max | Avg | Min | Max | Avg |
| *Colibri coruscans* | 0.61 | 0.93 | 0.83 | 0.61 | 0.93 | 0.83 | 0.60 | 0.93 | 0.82 | 0.60 | 0.93 | 0.83 |
| *Topaza pella* | 0.58 | 0.93 | 0.82 | 0.58 | 0.92 | 0.82 | 0.57 | 0.92 | 0.81 | 0.57 | 0.92 | 0.82 |
| *Aerodramus vanikorensis* | 0.64 | 0.94 | 0.85 | 0.64 | 0.94 | 0.85 | 0.63 | 0.94 | 0.84 | 0.63 | 0.94 | 0.84 |
| *Streptoprocne zonaris* | 0.55 | 0.92 | 0.81 | 0.54 | 0.92 | 0.81 | 0.54 | 0.91 | 0.80 | 0.54 | 0.92 | 0.80 |
| *Hemiprocne mystacea* | 0.56 | 0.90 | 0.80 | 0.56 | 0.89 | 0.79 | 0.55 | 0.89 | 0.79 | 0.56 | 0.89 | 0.79 |
| *Aegotheles cristatus* | 0.73 | 0.93 | 0.87 | 0.73 | 0.93 | 0.86 | 0.72 | 0.93 | 0.86 | 0.72 | 0.93 | 0.86 |
| *Aegotheles insignis* | 0.46 | 0.86 | 0.73 | 0.46 | 0.85 | 0.73 | 0.45 | 0.84 | 0.72 | 0.46 | 0.85 | 0.73 |
| *Hydropsalis rufiventris* | 0.58 | 0.92 | 0.82 | 0.57 | 0.91 | 0.81 | 0.57 | 0.91 | 0.81 | 0.57 | 0.92 | 0.81 |
| *Lyncornis macrotis* | 0.56 | 0.92 | 0.81 | 0.56 | 0.91 | 0.81 | 0.55 | 0.91 | 0.80 | 0.55 | 0.91 | 0.80 |
| *Eurostopodus mystacalis* | 0.59 | 0.92 | 0.82 | 0.59 | 0.92 | 0.82 | 0.59 | 0.91 | 0.81 | 0.59 | 0.91 | 0.81 |
| *Nyctibius aethereus* | | | | 0.57 | 0.91 | 0.80 | 0.57 | 0.91 | 0.80 | 0.57 | 0.91 | 0.81 |
| *Nyctibius bracteatus* | 0.42 | 0.78 | 0.66 | | | | | | | 0.42 | 0.77 | 0.66 |
| *Nyctibius grandis* | 0.56 | 0.90 | 0.80 | 0.56 | 0.90 | 0.79 | 0.56 | 0.89 | 0.79 | 0.56 | 0.90 | 0.79 |
| *Nyctibius griseus* | | | | | | | 0.54 | 0.89 | 0.78 | 0.54 | 0.89 | 0.78 |
| *Nyctibius jamaicensis* | | | | | | | 0.63 | 0.94 | 0.84 | 0.63 | 0.94 | 0.85 |
| *Nyctibius leucopterus* | | | | | | | 0.56 | 0.91 | 0.80 | 0.56 | 0.91 | 0.80 |
| *Nyctibius maculosus* | | | | | | | 0.54 | 0.90 | 0.79 | 0.54 | 0.91 | 0.79 |
| *Batrachostomus septimus* | 0.60 | 0.79 | 0.72 | 0.60 | 0.79 | 0.72 | 0.59 | 0.78 | 0.71 | 0.59 | 0.79 | 0.72 |
| *Podargus strigoides* | 0.54 | 0.91 | 0.79 | 0.53 | 0.90 | 0.79 | 0.53 | 0.90 | 0.78 | 0.53 | 0.90 | 0.79 |
| *Steatornis caripensis* | 0.62 | 0.95 | 0.85 | 0.62 | 0.94 | 0.84 | 0.61 | 0.94 | 0.84 | 0.61 | 0.94 | 0.84 |
| *Columba livia* | 0.64 | 0.95 | 0.85 | 0.64 | 0.94 | 0.85 | 0.63 | 0.94 | 0.84 | 0.63 | 0.94 | 0.84 |
| *Anseranas semipalmata* | 0.73 | 0.95 | 0.88 | 0.73 | 0.95 | 0.88 | 0.72 | 0.95 | 0.87 | 0.72 | 0.95 | 0.87 |
| *Gallus gallus* | 0.62 | 0.95 | 0.85 | 0.62 | 0.95 | 0.84 | 0.62 | 0.94 | 0.84 | 0.62 | 0.94 | 0.84 |

**Figures**

Figure S1. Taxon quartets used in quartet summary analyses (a), and cumulative plots (left, right) and flux diagram (center) representing results of quartet summary analyses (b). Cumulative plots depict presence of each possible resolution of quartet 1, which includes *Nyctibius bracteatus* (N)*, Steatornis caripensis* (S)*, Colibri coruscans* (C) and *Gallus gallus* (G). There is one plot present for each treatment (in this case, Untrimmed (left), PHYLUCE-trimmed (right); taxon set 4). M, 1 and 2 represent the three possible rooted triplet topologies given in newick format on the left of the figure. U is the unresolved topology. Majority rule (50%) consensus of a set of bootstraps for each locus are summarized, and the numbers of loci that support each quartet resolution are plotted in the cumulative count graphs. In the flux diagram, arrows indicate the directionality of change in loci that switch their support from one resolution to the other between treatments. Numbers indicate the quantity of loci switching and arrow width is proportional to those numbers. The areas of the circles are proportional to the number of loci supporting each treatment.

a)

| Quartet | Taxa |
|---|---|
| 1 | Nyctibius bracteatus, Steatornis caripensis, Colibri coruscans, Gallus gallus |
| 2 | Nyctibius grandis, Steatornis caripensis, Colibri coruscans, Gallus gallus |
| 3 | Hydropsalis rufiventris, Steatornis caripensis, Colibri coruscans, Gallus gallus |
| 4 | Podargus strigoides, Aegotheles cristatus, Colibri coruscans, Gallus gallus |
| 5 | Podargus strigoides, Aegotheles cristatus, Nyctibius grandis, Gallus gallus |
| 6 | Batrachostomus septimus, Aegotheles cristatus, Colibri coruscans, Gallus gallus |
| 7 | Batrachostomus septimus, Aegotheles cristatus, Nyctibius grandis, Gallus gallus |
| 8 | Nyctibius grandis, Nyctibius bracteatus, Steatornis caripensis, Gallus gallus |
| 9 | Hydropsalis rufiventris, Lyncornis macrotis, Eurostopodus mystacalis, Gallus gallus |
| 10 | Podargus strigoides, Aegotheles cristatus, Nyctibius grandis, Colibri coruscans |
| 11 | Podargus strigoides, Aegotheles cristatus, Nyctibius bracteatus, Colibri coruscans |
| 12 | Batrachostomus septimus, Aegotheles cristatus, Nyctibius grandis, Colibri coruscans |
| 13 | Batrachostomus septimus, Aegotheles cristatus, Nyctibius bracteatus, Colibri coruscans |

b)

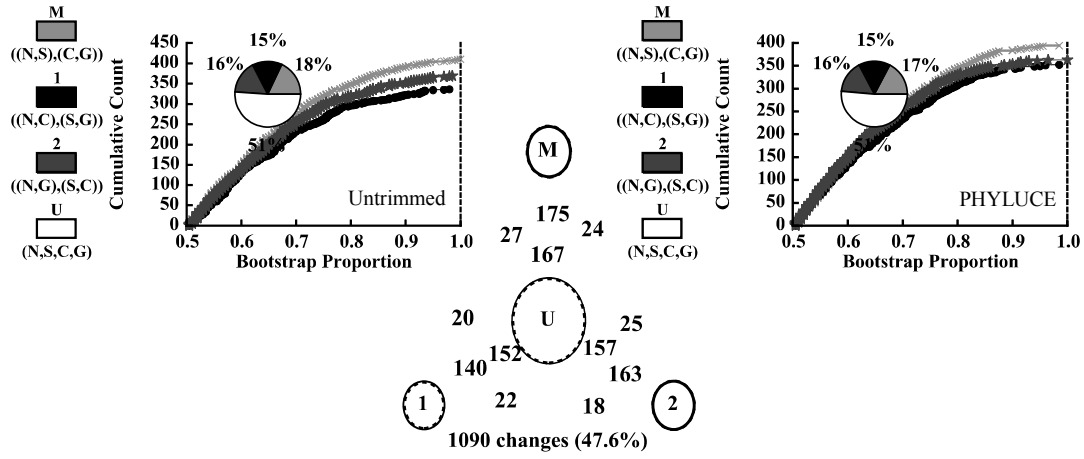


Figure S2. Constraint trees used in AU tests for each dataset.

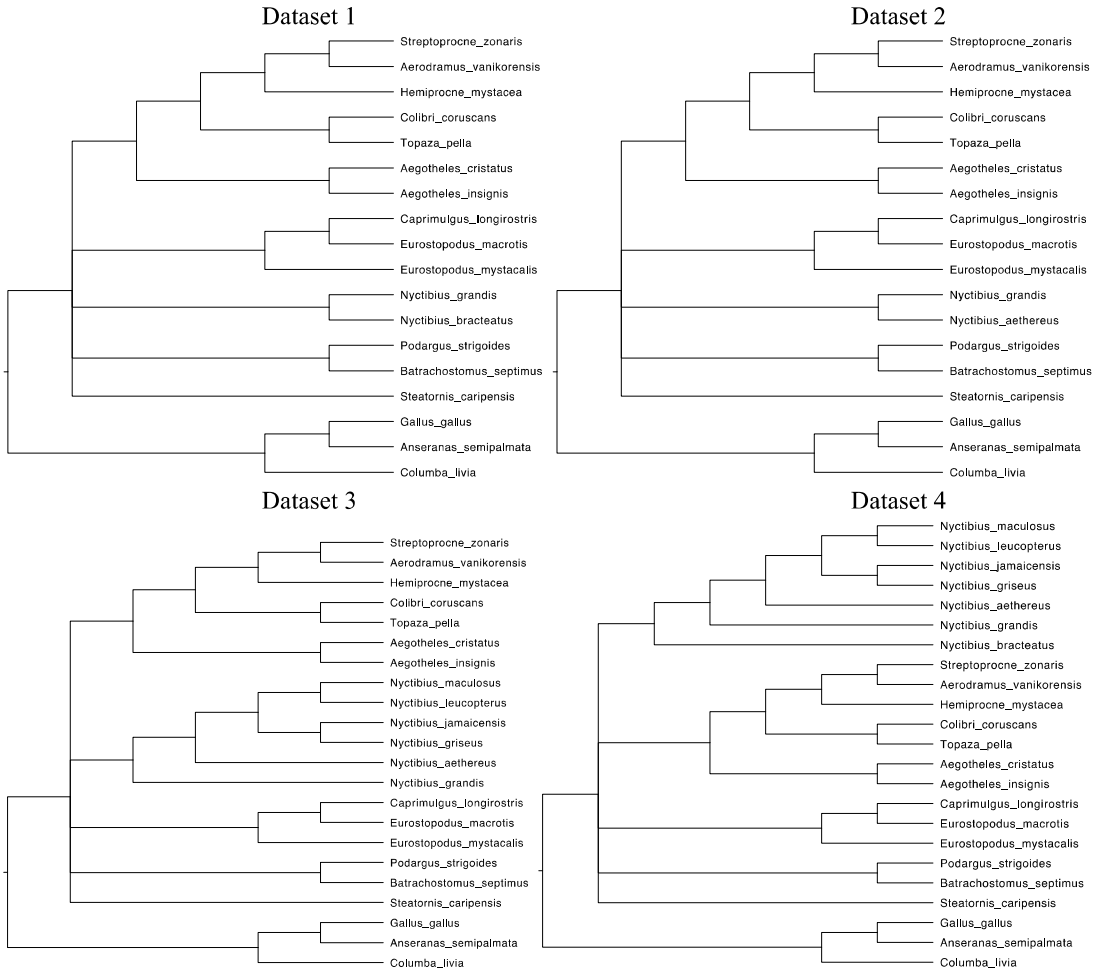
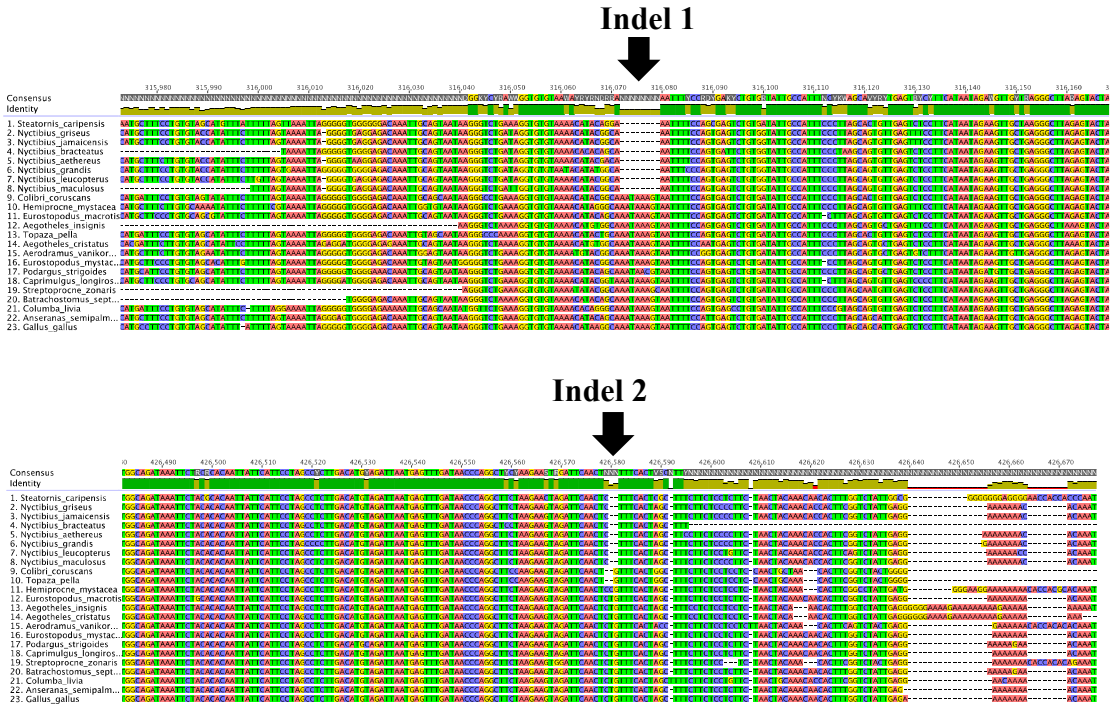
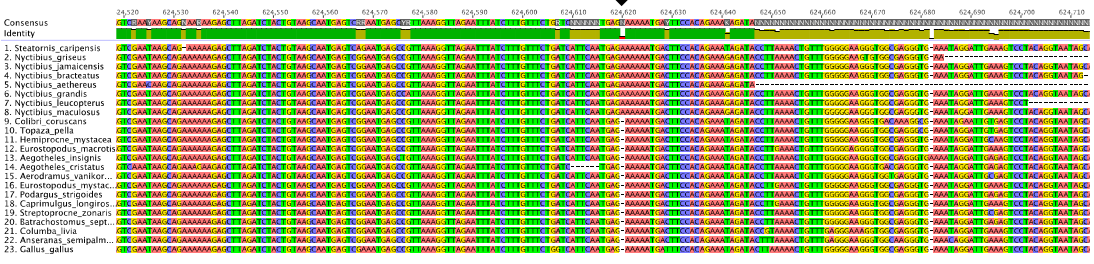


Figure S3. Description of informative indels used in this study, and images of alignments where they occur. All indels used were present in both dataset 1 and 4 matrices. Indels are depicted in the D4CU matrix. Images exported from Geneious (v10.0.5; Biomatters, Ltd.). No indels were identified supporting topology B.

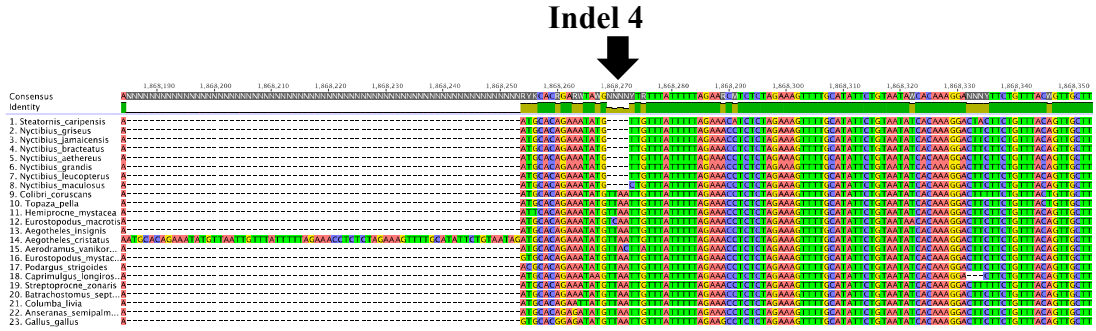
| Indel | Supports Topology A | Supports Topology C |
|---|---|---|
| 1 | 8 bp deletion | |
| 2 | 2 bp deletion; overlaps with 2bp deletion in hummers by 1 bp | |
| 3 | 1 bp insertion of an `A` in a homopolymer run | |
| 4 | 4 bp deletion | |
| 5 | | 7 bp deletion |
| 6 | | 2 bp duplication of `AT` in an AT rich region |



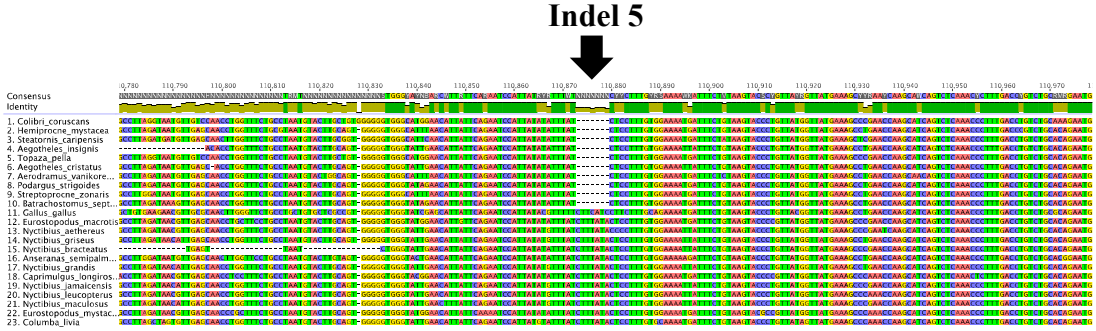
## Indel 3



## Indel 4



## Indel 5



## Indel 6

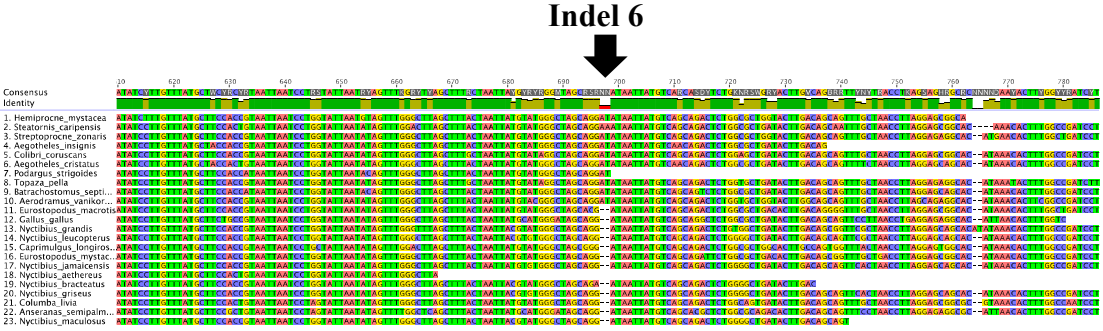
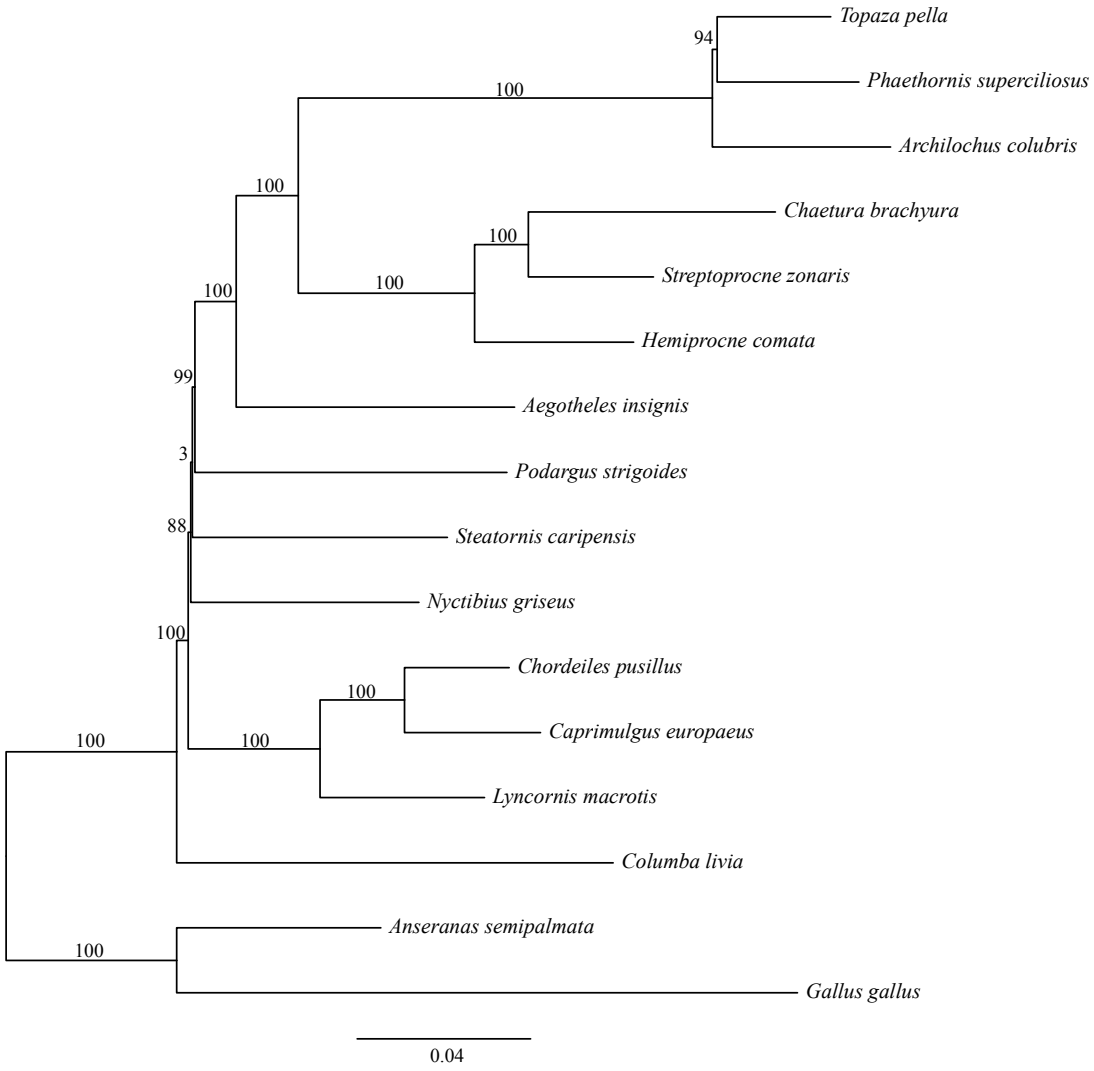
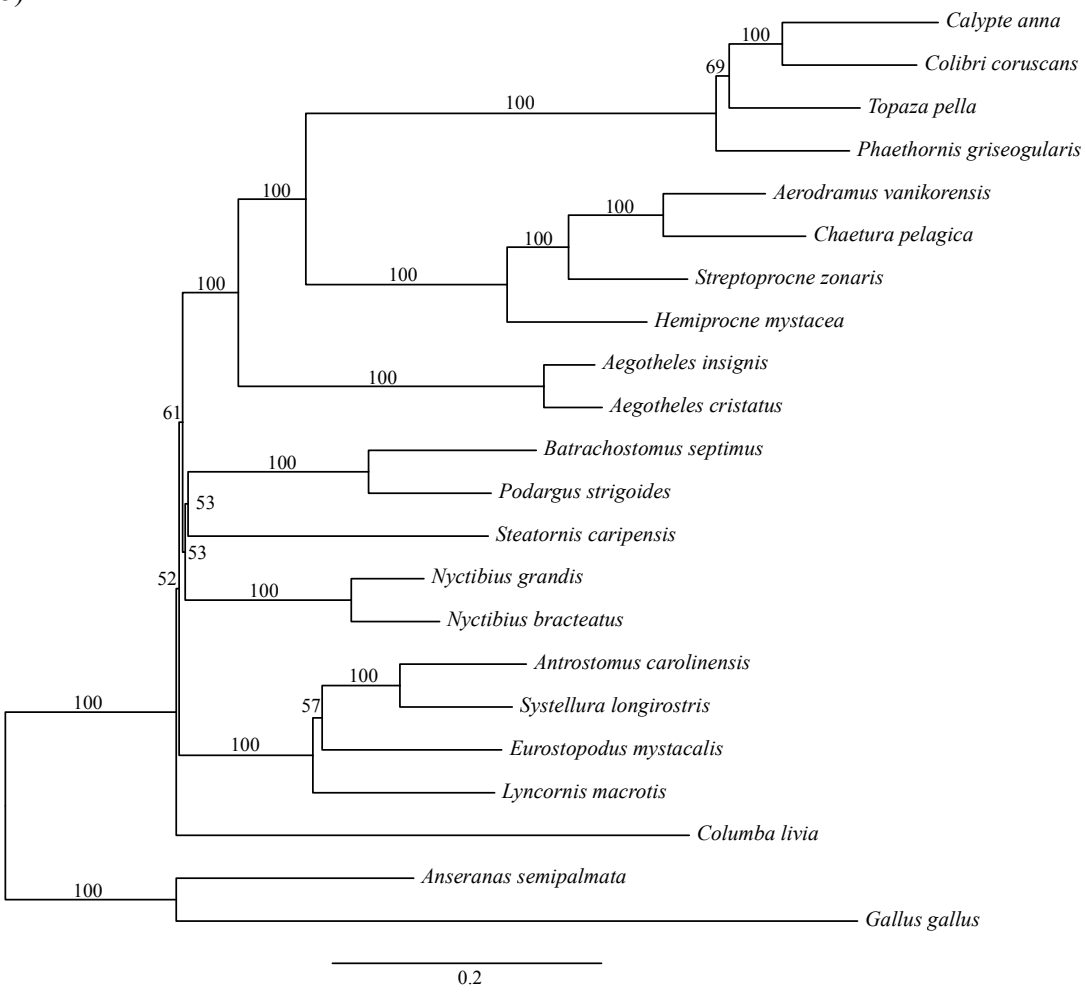


Figure S4. Maximum likelihood estimates of RY-coded datasets from (a) Prum et al. (2015) and (b) Reddy et al. (2017). Bootstrap support values are plotted at their respective nodes. Scale units are substitutions per site. Tip labels have been collapsed in (b) where both new sequence and published genome data were in the data matrix.

a)



b)


Calypte anna
Colibri coruscans
Topaza pella
Phaethornis griseogularis
Aerodramus vanikorensis
Chaetura pelagica
Streptoprocne zonaris
Hemiprocne mystacea
Aegotheles insignis
Aegotheles cristatus
Batrachostomus septimus
Podargus strigoides
Steatornis caripensis
Nyctibius grandis
Nyctibius bracteatus
Antrostomus carolinensis
Systellura longirostris
Eurostopodus mystacalis
Lyncornis macrotis
Columba livia
Anseranas semipalmata
Gallus gallus
100
69
100
100
100
100
100
100
100
100
61
100
53
53
52
100
100
57
100
100
100
0.2

# Literature Cited

Akaike, H. (1973). Information Theory and Extension of the Maximum Likelihood Principle (B. N. Petrov & F. Csaki, Eds.) In *Proceedings of the Second International Symposium on Information Theory* (pp. 267-281). Akademiai Kaido, Budapest.

Baird, S. F., Cassin, J., & Lawrence, G. N. (1858). General Report Upon the Zoology of the Several Pacific Railroad Routes. Part II: Birds (Vol. 9). Washington: Beverly Tucker.

Barrowclough, G. F., Groth, J. G., & Mertz, L. A. (2006). The RAG-1 Exon in the Avian Order Caprimulgiformes: Phylogeny, Heterozygosity, and Base Composition. *Molecular Phylogenetics and Evolution*, *41*(1), 238-248.

Bejerano, G., Haussler, D., & Blanchette, M. (2004a). Into the Heart of Darkness: Large-scale Clustering of Human Non-Coding DNA. *Bioinformatics*, *20*(Suppl 1), i40-i48.

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S., & Haussler, D. (2004b). Ultraconserved Elements in the Human Genome. *Science*, *304*(5675), 1321-1325.

Benson, D. A. (2004). GenBank. *Nucleic Acids Research*, *33* (Database issue), D34-D38.

Bergsten, J. (2005). A Review of Long-Branch Attraction. *Cladistics*, *21*(2), 163-193.

Bloch, N. I. (2015). Evolution of Opsin Expression in Birds Driven by Sexual Selection and Habitat. *Proceedings of the Royal Society B, 282*(1798), 20142321.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a Flexible Trimmer

for Illumina Sequence Data. *Bioinformatics, 30*(15), 2114-2120.

Bonilla, A. J., Braun, E. L., & Kimball, R. T. (2010). Comparative Molecular Evolution and Phylogenetic Utility of 3′-UTRs and Introns in Galliformes. *Molecular Phylogenetics and Evolution*, *56*(2), 536-542.

Borges, R., Khan, I., Johnson, W. E., Gilbert, M. T. P., Zhang, G., Jarvis, E. D., O'Brien, S. J., & Antunes, A. (2015). Gene Loss, Adaptive Evolution and the Co-evolution of Plumage Coloration Genes with Opsins in Birds. *BMC Genomics, 16,* 751.

Borrero, J. I., H. (1974). Notes on the Structure of the Upper Eyelid of Potoos (Nyctibius). *The Condor*, (76), 210-240.

Bossert, S., Murray, E. A., Blaimer, B. B., & Danforth, B. N. (2017). The Impact of GC Bias on Phylogenetic Accuracy using Targeted Enrichment Phylogenomic Data. *Molecular Phylogenetics and Evolution*, *111*, 149-157.

Bowmaker, J. K. (2008). Evolution of Vertebrate Visual Pigments. *Vision Research*, *48*(20), 2022-2041.

Bowmaker, J. K., & Martin, G. R. (1978). Visual Pigments and Colour Vision in a Nocturnal Bird, Strix aluco (tawny owl). *Vision Research*, *18*(9), 1125-1130.

Bowmaker, J. K., Heath, L. A., Wilkie, S. E., & Hunt, D. M. (1997). Visual Pigments and Oil Droplets from Six Classes of Photoreceptor in the Retinas of Birds. *Vision Research, 37*(16), 2183-2194.

Branstetter, M. G., Danforth, B. N., Pitts, J. P., Faircloth, B. C., Ward, P. S., Buffington, M. L., Gates, M. W., Kula, R. R., & Brady, S. G. (2017). Phylogenomic Insights into the Evolution of Stinging Wasps and the Origins of

Ants and Bees. *Current Biology*, *27*(7), 1019-1025.

Braun, M. J., & Huddleston, C. (2009). A Molecular Phylogenetic Survey of Caprimulgiform Nightbirds illustrates the Utility of Non-coding Sequences. *Molecular Phylogenetics and Evolution*, *53*(3), 948-960.

Brumfield, R. T., Swofford, D. L., & Braun, M. J. (1997). Evolutionary Relationships among the Potoos (Nyctibiidae) Based on Isozymes. *Ornithological Monographs*, *48*, 129-145.

Brumfield, R., Liu, L., Lum, D., & Edwards, S. (2008). Comparison of Species Tree Methods for Reconstructing the Phylogeny of Bearded Manakins (Aves: Pipridae, Manacus) from Multilocus Sequence Data. *Systematic Biology*, *57*(5), 719-731.

Bull, J. J., Huelsenbeck, J. P., Cunningham, C. W., Swofford, D. L., & Waddell, P. J. (1993). Partitioning and Combining Data in Phylogenetic Analysis. *Systematic Biology*, *42*(3), 384-397.

Bühler, P. (1970). Schädelmorphologie und Kiefermechanik der Caprimulgidae (Aves). *Zeitschrift Für Morphologie Der Tiere, 66,* 337-399.

Calle-Mustienes, E., Feijóo, C. G., Manzanares, M., Tena, J. J., Rodríguez-Seguel, E., Letizia, A., Allend, M. L., & Gómez-Skarmeta, J. L. (2005). A Functional Survey of the Enhancer Activity of Conserved Non-Coding Sequences from Vertebrate *Iroquois* Cluster Gene Deserts. *Genome Research, 15*, 1601-1072.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: Architecture and Applications. *BMC Bioinformatics*, *10*(1), 421.

Capella-Gutierrez, S., Silla-Martinez, J. M., & Gabaldon, T. (2009). TrimAl: a Tool

for Automated Alignment Trimming in Large-scale Phylogenetic Analyses. *Bioinformatics*, *25*(15), 1972-1973.

Chantler, P. (2017). Swifts (Apodidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retreived May 22, 2017, from www.hbw.com/node/52266

Chifman, J., & Kubatko, L. (2014). Quartet Inference from SNP Data under the Coalescent Model. *Bioinformatics*, *30*(23), 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.

Cleere, N. (1998). *Nightjars: a Guide to the Nightjars, Nighthawks, and their Relatives* (1st ed.). New Haven, CT: Yale University Press.

Cleere, N., Kratter, A. W., Steadman, D. W., Braun, M. J., Huddleston, C. J., Filardi, C. E., & Dutson, G. (2007). A New Genus of Frogmouth (Podargidae) from the Solomon Islands-Results from a Taxonomic Review of Podargus ocellatus inexpectatus Hartert 1901. *Ibis*, *149*(2), 271-286.

Cleere, N. (2010). *Nightjars, Potoos, Frogmouths, Oilbird and Owlet-nightjars of the World* (1st ed.). Princeton, NJ: Princeton University Press.

Cleere, N. (2017). Nightjars (Caprimulgidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52265

Cohn-Haft, M. (1999). Family Nyctibiidae (Potoos). In J. del Hoyo, A. Elliott, & J.

Sargatal (Eds.), *Barn-owls to Hummingbirds* (1st ed., Vol. 5, Handbook of the Birds of the World, pp. 288-301). Barcelona: Lynx Edicions.

Cohn-Haft, M. (2016). Potoos (Nyctibiidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved June 6, 2016, from www.hbw.com/node/52264

Costa, T. V. V. (2014). Osteology and Phylogeny of the Avian Order Caprimulgiformes, with Special Emphasis on Nyctibiidae and Caprimulgidae (Doctoral dissertation, Universidade de São Paulo, pp. 1-465). São Paulo.

Coyle, B. J., Hart, N. S., Carleton, K. L., & Borgia, G. (2012). Limited Variation in Visual Sensitivity among Bowerbird Species Suggests that there is No Link Between Spectral Tuning and Variation in Display Colouration. *Journal of Experimental Biology*, *215*(7), 1090-1105.

Cracraft, J. (1981). Toward a Phylogenetic Classification of the Recent Birds of the World (Class Aves). *The Auk, 98,* 681-714.

Cracraft, J. (1988). The Major Clades of Birds (M. J. Benton, Ed.) In *The Phylogeny and Classification of the Tetrapods, Volume 1: Amphibians, Reptiles, Birds* (339-361). Oxford.

Crawford, N. G., Faircloth, B. C., McCormack, J. E., Brumfield, R. T., Winker, K., & Glenn, T. C. (2012). More than 1000 Ultraconserved Elements Provide Evidence that Turtles are the Sister Group of Archosuars. *Biology letters, 8*(5), 783-786.

Darriba, D., Taboada, G. L., Doallo, R., & Posada, D. (2012). jModelTest 2: More Models, new Heuristics and Parallel Computing. *Nature Methods*, *9*(8), 772-772.

de Queiroz, A. (1993). For Consensus (sometimes). *Systematic Biology*, *42*(3), 368-

372.

Degnan, J. H., & Rosenberg, N. A. (2009). Gene Tree Discordance, Phylogenetic Inference and the Multispecies Coalescent. *Trends in Ecology & Evolution*, *24*(6), 332-340.

Delsuc, F., Brinkmann, H., & Philippe, H. (2005). Phylogenomics and the Reconstruction of the Tree of Life. *Nature Reviews Genetics, 6,* 361–375.

Dobzhansky, T. G. (1973). Nothing in Biology Makes Sense Except in the Light of Evolution. *American Biology Teacher, 35*, 125-129.

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 Phylogenetics. *Molecular Phylogenetics and Evolution*, *94*(Part A), 447-462.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE, 6*(5): 319379.

Endler, J. A. (1993). The Color of Light in Forests and its Implications. *Ecological Monographs, 63*(1), 1-27.

Ericson, P. G. P., Anderson, C. L., Britton, T., Elzanowski, A., Johansson, U. S., Kallersjo, M., Ohlson, J. I., Parsons, T. J., Zuccon, D., & Mayr, G. (2006). Diversification of Neoaves: Integration of Molecular Sequence Data and Fossils. *Biology Letters*, *2*(4), 543-547.

Eyre-Walker, A. (1993). Recombination and Mammalian Genome Evolution.

*Proceedings of the Royal Society of London B: Biological Sciences*, *252*(1335), 237-243.

Faircloth, B. C., Sorenson, L., Santini, F., & Alfaro, M. E. (2013a). A Phylogenomic Perspective on the Radiation of the Ray-Finned Fishes Based upon Targeted Sequencing of Ultraconserved Elements (UCEs). *PLoS ONE, 8*(6), e65923.

Faircloth, B. C. (2013). Illumiprocessor: A Trimmomatic Wrapper for Parallel Adapter and Quality Trimming. www.dx.doi.org/10.6079/J9ILL.

Faircloth, B. C., Branstetter, M. G., White, N. D., & Brady, S. G. (2015a). Target Enrichment of Ultraconserved Elements from Arthropods Provides a Genomic Perspective on Relationships among Hymenoptera. *Molecular Ecology Resources, 15*(3), 489-501.

Faircloth, B. C. (2015). PHYLUCE is a Software Package for the Analysis of Conserved Genomic Loci. *Bioinformatics*, *32*(5), 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. *Systematic Biology*, *61*(5), 717-726.

Figuet, E., Ballenghien, M., Romiguier, J., & Galtier, N. (2015). Biased Gene Conversion and GC-Content Evolution in the Coding Sequences of Reptiles and Vertebrates. *Genome Biology and Evolution*, *7*(1), 240-250.

Foster, P. G. & Hickey, D. A. (1999). Compositional Bias May Affect both DNA-Based and Protein-Based Phylogenetic Reconstructions. *Journal of Molecular Evolution, 48,* 284–290.

Galtier, N., Piganeau, G., Mouchiroud, D., & Duret, L. (2001). GC-Content Evolution in Mammalian Genomes: The Biased Gene Converstion Hypothesis. *Genetics, 159,* 907-911.

Galtier, N., & Duret, L. (2007). Adaptation or Biased Gene Conversion? Extending the Null Hypothesis of Molecular Evolution. *Trends in Genetics*, *23*(6), 273-277.

Galtier, N., & Gouy, M. (1998). Inferring Pattern and Process: Maximum-likelihood Implementation of a Nonhomogeneous Model of DNA Sequence Evolution for Phylogenetic Analysis. *Molecular Biology and Evolution, 15*(7), 871-879.

Gatesy, J., & Springer, M. S. (2014). Phylogenetic Analysis at Deep Timescales: Unreliable Gene Trees, bypassed Hidden Support, and the Coalescence/Concatalescence Conundrum. *Molecular Phylogenetics and Evolution*, *80*, 231-266.

Glenn, T. C., & Faircloth, B. C. (2016). Capturing Darwin's Dream. *Molecular Ecology Resources*, *16*(5), 1051-1058.

Govardovskii, V. I., Fyhrquist, N., Reuter, T., Kuzmin, D. G., & Donner, K. (2000). In Search of the Visual Pigment Template. *Visual Neuroscience*, *17*(4), 509-528.

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*(7), 644-652.

Griffiths, C. S., Barrowclough, G. F., Groth, J. G., & Mertiz, L. (2004). Phylogeny of

the Falconidae (Aves): A Comparison of the Efficacy of Morphological, Mitochondrial, and Nuclear Data. *Molecular Phylogenetics and Evolution, 32,* 101-109.

Groth, J. G., & Barrowclough, G. F. (1999). Basal Divergences in Birds and the Phylogenetic Utility of the Nuclear RAG-1 Gene. *Molecular Phylogenetics and Evolution*, *12*(2), 115-123.

Hackett, S. J., Kimball, R. T., Reddy, S., Bowie, R. C. K., Braun, E. L., Braun, M. J., Chojnowski, J. L., Cox, W. A., Kin-Lan, H., Harshman, J., Huddleston, C. J., Marks, B. D., Miglia, K. J., Moore, W. S., Sheldon, F. H., Steadman, D. W., Witt, C. C., * Yuri, T. (2008). A Phylogenomic Study of Birds Reveals Their Evolutionary History. *Science*, *320*(5884), 1763-1768.

Hahn, M. W., & Nakhleh, L. (2015). Irrational Exuberance for Resolved Species Trees. *Evolution*, *70*(1), 7-17.

Han, K.-L., Robbins, M. B., & Braun, M. J. (2010). A Multi-gene Estimate of Phylogeny in the Nightjars and Nighthawks (Caprimulgidae). *Molecular Phylogenetics and Evolution*, *55*(2), 443-453.

Harmston, N., Barešić, A., & Lenhard, B. (2013). The Mystery of Extreme Non-Coding Conservation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *368*(1632), 20130021.

Harris, R. S. (2007). Improved Pairwise Alignment of Genomic DNA. (Doctoral dissertation, The Pennsylvania State University, pp. 1-84). University Park.

Harshman, J., Huddleston, C. J., Bollback, J. P., Parsons, T. J., & Braun, M. J. (2003). True and False Gharials: A Nuclear Gene Phylogeny of Crocodylia. *Systematic*

*Biology*, *52*(3), 386-402.

Hart, N. S., Partridge, J. C., & Cuthill, I. C. (1999). Visual Pigments, Cone Oil Droplets, Ocular Media and Predicted Spectral Sensitivity in the Domestic Turkey (*Meleagris gallopavo*). *Vision Research, 39*(20), 3321-3328.

Hart, N. S. (2001). The Visual Ecology of Avian Photoreceptors. *Progress in Retinal and Eye Research*, *20*(5), 675-703.

Harvey, M. G., Smith, B. T., Glenn, T. C., Faircloth, B. C., & Brumfield, R. T. (2016). Sequence Capture versus Restriction Site Associated DNA Sequencing for Shallow Systematics. *Systematic Biology, 65*(5), 910-924.

Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A. J., Searle, S. M. J., Amode, R., Brent, S., Spooner, W., Kulesha, E., Yates, A., & Flicek, P. (2016). Ensembl Comparative Genomics Resources. *Database*, *2016*, 1-17.

Higgins, D., & Lemey, P. (2003). Multiple Sequence Alignment. In P. Lemey, M. Salemi, & A. M. Vandamme (Eds.), *The Phylogenetic Handbook: A Practical approach to DNA and Protein Phylogeny* (1st ed., pp. 68-96). New York: Cambridge University Press.

Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., et al. (2004). Sequence and Comparative Analysis of the Chicken Genome provide Unique Perspectives on Vertebrate Evolution. *Nature*, *432*(7018), 695-716.

Hoff, K. M. (1966). A Comparative Study of the Appendicular Muscles of Strigiformes and Caprimulgiformes. (Doctoral dissertation, Washington State University, pp. 1-183). Seattle.

Holyoak, D. T. (2001). *Nightjars and their Allies: The Caprimulgiformes.* (1st ed., Vol. 7, Bird Families of the World)(C. M. Perrins, W. J. Bock, & J. Kikkawa, Eds.). Oxford: Oxford University Press.

Holyoak, D.T. (2017a). Frogmouths (Podargidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52263

Holyoak, D.T. (2017b). Owlet-Nightjars (Aegothelidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52262

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). *Molecular Biology and Evolution*, *33*(4), 1110-1125.

Hunt, D. M., Carvalho, L. S., Cowing, J. A., & Davies, W. L. (2009). Evolution and spectral tuning of visual pigments in birds and mammals. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1531), 2941-2955.

Hurvich, C. M., & Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika*, *76*(2), 297-307.

Imai, H., Kojima, D., Oura, T., Tachibanaki, S., Terakita, A., & Shichida, Y. (1997). Single Amino Acid Residue as a Functional Determinant of Rod and Cone Visual Pigments. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(6), 2322-2326.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., & Li, C. (2014). Whole-

Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science*, *346*(6215), 1320-1331.

Jarvis, E. D., Mirarab, S., Aberer, A. J., Li, B., Houde, P., Li, C., et al. (2014). Whole-Genome Analyses Resolve Early Branches in the Tree of Life of Modern Birds. *Science*, *346*(6215), 1320-1331.

Jeffroy, O., Brinkmann, H., Delsuc, F., & Philippe, H. (2006). Phylogenomics: The Beginning of Incongruence? *Trends in Genetics, 22,* 225–231.

Kelley, J. L., & Davies, W. I. L. (2016). The Biological Mechanisms and Behavioral Functions of Opsin-Based Light Detection by the Skin. *Frontiers in Ecology and Evolution, 4*(106).

Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, *12*(4), 656-664.

Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, A. D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996-1006.

Kocot, K. M., Struck, T. H., Merkel, J., Waits, D. S., Todt, C., Brannock, P. M., Weese, D. A., Cannon, J. T., Moroz, L. L., Lieb, B., & Halanych, K. M. (2016). Phylogenomics of Lophotrochozoa with Consideration of Systematic Error. *Systematic Biology*, *66*(2), 256-282.

Kubatko, L. S., & Degnan, J. H. (2007). Inconsistency of Phylogenetic Estimates from Concatenated Data under Coalescence. *Systematic Biology*, *56*(1), 17-24.

Kumar, S., Filipski, A. J., Battistuzzi, F. U., Kosakovsky Pond, S. L., & Tamura, K. (2012). Statistics and Truth in Phylogenomics. *Molecular Biology and Evolution*,

*29*(2), 457-472.

Kück, P., & Meusemann, K. (2010). Parametric and Non-Parametric Masking of Randomness in Sequence Alignments can be Improved and Leads to Better Resolved Trees. *Frontiers in Zoology, 7*(1), 10.

Lamb, T. D., Patel, H., Chuah, A., Natoli, R. C., Davies, W. I. L., Hart, N. S., Colling, S. P., & Hunt, D. M. (2016). Evolution of Vertebrate Phototransduction: Cascade Activation. *Molecular Biology and Evolution, 33*(8), 2064-2087.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial Sequencing and Analysis of the Human Genome. *Nature, 409*, 860-921.

Lanfear, R., Calcott, B., Ho, S. Y. W., & Guindon, S. (2012). PartitionFinder: Combined Selection of Partitioning Schemes and Substitution Models for Phylogenetic Analyses. *Molecular Biology and Evolution*, *29*(6), 1695-1701.

Langmead, B., & Salzberg, S. L. (2012). Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods*, *9*(4), 357-359.

Larsen, C., Speed, M., Harvey, N., & Noyes, H. A. (2007). A Molecular Phylogeny of the Nightjars (Aves: Caprimulgidae) Suggests Extensive Conservation of Primitive Morphological Traits across Multiple Lineages. *Molecular Phylogenetics and Evolution*, *42*(3), 789-796.

Lemmon, A. R., Emme, S. A., & Lemmon, E. M. (2012). Anchored Hybrid Enrichment for Massively High-Throughput Phylogenomics. *Systematic Biology*, *61*(5), 727-744.

Lockton, S., & Gaut, B. S. (2005). Plant Conserved Non-Coding Sequences and

Paralogue Evolution. *TRENDS in Genetics, 21*(1), 60-65.

Lim, H. C., & Braun, M. J. (2016). High-Throughput SNP Genotyping of Historical and Modern Samples of five Bird Species via Sequence Capture of Ultraconserved Elements. *Molecular Ecology Resources*, *16*(5), 1204-1223.

Liu, K., Warnow, T. J., Holder, M. T., Nelesen, S. M., Yu, J., Stamatakis, A. P., & Linder, C. R. (2011). SATe-II: Very Fast and Accurate Simultaneous Estimation of Multiple Sequence Alignments and Phylogenetic Trees. *Systematic Biology*, *61*(1), 90-106.

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*(1), 36-53.

Liu, L., Yu, L., Kubatko, L., Pearl, D. K., & Edwards, S. V. (2009). Coalescent Methods for Estimating Phylogenetic Trees. *Molecular Phylogenetics and Evolution*, *53*(1), 320-328.

Lythgoe, J. N. (1979). *The Ecology of Vision.* Clarendon Press, Oxford.

Mariaux, J., & Braun, M. J. (1996). A Molecular Phylogenetic Survey of the Nightjars and Allies (Caprimulgiformes) with Special Emphasis on the Potoos (Nyctibiidae). *Molecular Phylogenetics and Evolution*, *6*(2), 228-244.

Martin, G., Rojas, L., & Ramírez, Y. (2004). The Eyes of Oilbirds (Steatornis caripensis): Pushing at the Limits of Sensitivity. *Naturwissenschaften, 91*, 26-29.

Mayr, G. (2002). Osteological Evidence for Paraphyly of the Avian Order Caprimulgiformes (Nightjars and Allies). *Journal Für Ornithologie*, *143*(1), 82-97.

Mayr, G. (2004). Old World Fossil Record of Modern-Type Hummingbirds. *Science*, *304*(5672), 861-864.

Mayr, G. (2005a). The Palaeogene Old World potoo Paraprefica Mayr, 1999 (Aves, Nyctibiidae): Its Osteology and Affinities to the New World Preficinae Olson, 1987. *Journal of Systematic Palaeontology*, *3*(4), 359-370.

Mayr, G. (2005b). The Paleogene Fossil Record of Birds in Europe. *Biological Reviews*, *80*(4), 515-542.

Mayr, G. (2010). Phylogenetic Relationships of the Paraphyletic "Caprimulgiform" Birds (Nightjars and Allies). *Journal of Zoological Systematics and Evolutionary Research*, *48*(2), 126-137.

Mayr, G., Manegold, A., & Johansson, U. S. (2003). Monophyletic Groups within "Higher Land Birds"—Comparison of Morphological and Molecular Data. *Journal of Zoological Systematics and Evolutionary Research*, *41*(4), 233-248.

McCormack, J. E., Faircloth, B. C., Crawford, N. G., Gowaty, P. A., Brumfield, R. T., & Glenn, T. C. (2012). Ultraconserved Elements are Novel Phylogenomic Markers that Resolve Placental Mammal Phylogeny when Combined with Species-Tree Analysis. *Genome Research*, *22*(4), 746-754.

McGill, R., Tukey, J. W., Larsen, W. A. (1978). Variations of Box Plots. *The American Statistician, 32,* 12-16.

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. *Systematic Biology*, *65*(4), 612-627.

Meredith, R. W., Gatesy, J., Emerling, C. A., York, V. M., & Springer, M. S. (2013). Rod Monochromacy and the Coevolution of Cetacean Retinal Opsins. *PLoS Genetics*, *9*(4), e1003432.

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees. In *Proceedings of the Gateway Computing Environments Workshop GCE* (pp. 1-8). New Orleans, LA.

Miller, W., Rosenbloom, K., Hardison, R. C., Hou, M., Taylor, J., Raney, B., Burhans, R., King, D. C., Baertsch, R., Blankenberg, D., Kosakovsky Pond, S. L., Nekrutenko, A., Giardine, B., Harris, R. S., Tyekucheva, S., Diekhans, M., Pringle, T. H., Murphy, W. J., Lesk, A., Weinstock, G. M., Linblad-Toh, K., Gibbs, R. A., Lander, E. S., Siepel, A., Haussler, D., & Kent, W. J. (2007). 28-Way Vertebrate Alignment and Conservation Track in the UCSC Genome Browser. *Genome Research*, *17*(12), 1797-1808.

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.

Misof, B., & Misof, K. (2009). A Monte Carlo Approach Successfully Identifies Randomness in Multiple Sequence Alignments: A More Objective Means of Data Exclusion. *Systematic Biology*, *58*(1), 21-34.

Moustafa, A., & Bhattacharya, D. (2008). PhyloSort: a User-Friendly Phylogenetic Sorting Tool and its Application to Estimating the Cyanobacterial Contribution to the Nuclear Genome of Chlamydomonas. *BMC Evolutionary Biology*, *8*(1), 6-7.

Nabholz, B., Künstner, A., Wang, R., Jarvis, E. D., & Ellegren, H. (2011). Dynamic

Evolution of Base Composition: Causes and Consequences in Avian Phylogenomics. *Molecular Biology and Evolution, 28*(8), 2197-2210.

Navratilova, P., Fredman, D., Hawkins, T. A., Turner, K., Lenhard, B., & Becker, T. S. (2009) Systematic Human/Zebrafish Comparative Identification of *Cis*-Regulatory Activity around Vertebrate Developmental Transcription Factory Genes. *Developmental Biology, 327,* 526-540.

Nesbitt, S. J., Ksepka, D. T., & Clarke, J. A. (2011). Podargiform Affinities of the Enigmatic Fluvioviridavis platyrhamphus and the Early Diversification of Strisores ("Caprimulgiformes" + Apodiformes). *PLoS ONE, 6*(11), e26350.

Nicol, J., & Arnott, H. (1974). Tapeta Lucida in the Eyes of Goatsuckers (Caprimulgidae). *Proceedings of the Royal Society of London Series B: Biological Sciences*, *187*(1088), 349-352.

Nóbrega, M. A., Zhu, Y., Plajzer-Frick, I., Afzal, V. & Rubin, E. M. (2004). Megabase Deletions of Gene Deserts Result in Viable Mice. *Nature, 431,* 988-993.

Oberholser, H. C. (1914). A Monograph of the Genus Chordeiles Swanson, Type of a New Family of Goatsuckers. *Bulletin of the United States National Museum*, *86*, 1-123.

Ödeen, A. (2003). Complex Distribution of Avian Color Vision Systems Revealed by Sequencing the SWS1 Opsin from Total DNA. *Molecular Biology and Evolution*, *20*(6), 855-861.

Ödeen, A., Hart, N. S., & Håstad, O. (2008). Assessing the Use of Genomic DNA as a Predictor of the Maximum Absorbance Wavelength of Avian SWS1 Opsin

Visual Pigments. *Journal of Comparative Physiology A: Sensory, Neural, and Behavioral Physiology*, *195*(2), 167-173.

Ödeen, A., Pruett-Jones, S., Driskell, A. C., Armenta, J. K., & Håstad, O. (2012). Multiple Shifts between Violet and Ultraviolet Vision in a Family of Passerine Birds with Associated Changes in Plumage Coloration. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1732), 1269-1276.

Ödeen, A., & Håstad, O. (2013). The Phylogenetic Distribution of Ultraviolet Sensitivity in Birds. *BMC Evolutionary Biology, 13*(36).

Pennacchio, L. A., Ahituv, N., Moses, A. N., Prabhakar, S., Nobrega, M. A., Shoukry, M., Minovitksy, S., Dubchak, I., Holt, A., Lewis, K. D., Plajzer-Frick, I., Akiyama, J., De Val, S., Afzal, V., Black, B. L., Couronne, O., Eisen, M. B., Visel, A., & Rubin, E. M. (2006). *In vivo* Enhancer Analysis of Human Conserved Non-Coding Sequences. *Nature, 444*(7118), 499-502.

Pennisi, E. (2017). Mysterious unchanging DNA finds a purpose in life. *Science*, *356*(6341), 892.

Peters, J. L. (1940). *Checklist of the Birds of the World.* Harvard University Press, Cambridge.

Peterson, R. T. (2002). A Field Guide to the Birds of Eastern and Central North America (5th ed.). Houghton Mifflin Harcourt, Boston.

Philippe, H., Brinkmann, H., Lavrov, D. V., Littlewood, D. T. J., Manuel, M., Wörheide, G., & Baurain, D. (2011). Resolving Difficult Phylogenetic Questions: Why More Sequences are not Enough. *PLoS Biology*, *9*(3), e1000602.

Pirnie, S. P., Osman, A., Zhu, Y., & Carmichael, G. G. (2016). An Ultraconserved

Element (UCE) controls homeostatic splicing of ARGLU1 mRNA. *Nucleic Acids Research, 45*(6), 3473-3486.

Posada, D. (2008). jModelTest: Phylogenetic Model Averaging. *Molecular Biology and Evolution*, *25*(7), 1253-1256.

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*(7574), 569-573.

Pugh, E. N., Jr, Nikonov, S., & Lamb, T. D. (1999). Molecular Mechanisms of Vertebrate Photoreceptor Light Adaptation. *Current Opinion in Neurobiology*, *9*(4), 410-418.

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics*, *26*(6), 841-842.

Reddy, S., Kimball, R. T., Pandey, A., Hosner, P. A., Braun, M. J., Hackett, S. J., Han, K-L., Harshamn, 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 Phylogenomics Data Sets Yield Conflicting Trees? Data Type Influences the Avian Tree of Life more than Taxon Sampling. *Systematic Biology, 0*(0), 1-23.

Remsen, J. V., Jr., Areta, J. I., Cadena, C. D., Claramunt, S., Jaramillo, A., Pacheco, J. F., Pérez-Emán, J., Robbins, M. B., Stiles, F. G., Stotz, D. F., & Zimmer, K. J. (2016). A Classification of the Bird Species of South America. Retrieved May 6, 2017, from www.museum.lsu.edu/~Remsen/SACCBaseline.htm

Ricklefs, R. E. (1970). Stage of taxon cycle and distribution of birds on Jamaica,

Greater Antilles. *Evolution, 24*(2), 475-477.

Ridgway, R. (1914). The Birds of North and Middle America. *Bulletin of the United States National Museum, 50*.

Robinson, D. F., & Foulds, L. R. (1981). Comparison of Phylogenetic Trees. *Mathematical Biosciences*, *53*(1-2), 131-147.

Rodieck, R. W. (1973). *The Vertebrate Retina: Principles of Structure and Function*. W. H. Freeman & Co. Ltd.

Rojas, L. M., Ramírez, Y., McNeil, R., Mitchell, M., & Marín, G. (2004). Retinal Morphology and Electrophysiology of Two Caprimulgiformes Birds: The Cave-Living and Nocturnal Oilbird *(Steatornis caripensis)*, and the Crepuscularly and Nocturnally Foraging Common Pauraque *(Nyctidromus albicollis)*. *Brain, Behavior and Evolution*, *64*(1), 19-33.

Romiguier, J., Cameron, S. A., Woodard, S. H., Fischman, B. J., Keller, L., & Praz, C. J. (2016). Phylogenomics Controlling for Base Compositional Bias Reveals a Single Origin of Eusociality in Corbiculate Bees. *Molecular Biology and Evolution*, *33*(3), 670-678.

Romiguier, J., Ranwez, V., Delsuc, F., Galtier, N., & Douzery, E. J. P. (2013). Less Is More in Mammalian Phylogenomics: AT-Rich Genes Minimize Tree Conflicts and Unravel the Root of Placental Mammals. *Molecular Biology and Evolution*, *30*(9), 2134-2144.

Romiguier, J., Ranwez, V., Douzery, E. J. P., & Galtier, N. (2010). Contrasting GC-Content Dynamics across 33 Mammalian Genomes: Relationship with Life-History Traits and Chromosome Sizes. *Genome Research*, *20*(8), 1001-1009.

Rosel, P. E., & Block, B. A. (1996). Mitochondrial Control Region Variability and Global Population Structure in the Swordfish, Xiphias gladius. *Marine Biology*, *125*(1), 11-22.

Rosenbloom, K. R., Armstrong, J., Barber, G. P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T. R., Fujita, P. A., Guruvadoo, L., Haeussler, M., Harte, R. A., Heitner, S., Hickey, G., Hinrichs, A. S., Hubley, R., Karolchik, D., Learned, K., Lee, B. T., Li, C. H., Miga, K. H., Nguyen, N., Paten, B., Raney, B. J., Smit, A. F. A., Speir, M. L., Zweig, A. S., Haussler, D., Kuhn, R. N., & Kent, W. J. (2015). The UCSC Genome Browser database: 2015 update. *Nucleic Acids Research*, *43*(D1), D670-D681.

Salichos, L., & Rokas, A. (2014). Inferring Ancient Divergences Requires Genes with Strong Phylogenetic Signals. *Nature*, *497*(7449), 327-331.

Salinas, N. R., & Little, D. P. (2014). 2matrix: A Utility for Indel Coding and Phylogenetic Matrix Concatenation. *Applications in Plant Sciences*, *2*(1), 1300083.

Schodde, R., & Mason, I. J. (1980). *Nocturnal Birds of Australia.* Lansdowne Editions, East Melbourne.

Schott, R., Panesar, B., Card, D. C., Preston, M., & Castoe, T. A. (2017). Targeted Capture of Complete Coding Regions across Divergent Species. *Genome Biology and Evolution*, *9*(2), 398-414.

Schuchmann, K.L. & Bonan, A. (2017). Hummingbirds (Trochilidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of*

*the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52268

Sheffield, N. C., Song, H., Cameron, S. L., & Whiting, M. F. (2009). Nonstationary Evolution and Compositional Heterogeneity in Beetle Mitochondrial Phylogenomics. *Systematic Biology, 58,* 381-394.

Shimodaira, H. (2002). An Approximately Unbiased Test of Phylogenetic Tree Selection. *Systematic Biology*, *51*(3), 492-508.

Shimodaira, H., & Hasegawa, M. (1999). Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution, 16*(8), 1114.

Shubin, N., Tabin, C., & Carroll, S., (2009). Deep Homology and the Origins of Evolutionary Novelty. *Nature*, *457,* 818-823.

Sibley, C. G., & Ahlquist, J. E. (1983). Phylogeny and Classification of Birds Based on the Data of DNA-DNA Hybridization. In R. F. Johnston (Ed.), *Current Ornithology* (1st ed., Vol. 1, pp. 245-292). Springer US.

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D., (2005). Evolutionarily Conserved Elements in Vertebrate, Insect, Worm, and Yeast Genomes. *Genome Research, 15,* 1034-1050.

Sigurdsson, S., & Cracraft, J. (2014). Deciphering the Diversity and History of New World Nightjars (Aves: Caprimulgidae) using Molecular Phylogenetics. *Zoological Journal of the Linnean Society*, *170*(3), 506-545.

Simmons, M. P., & Gatesy, J. (2015). Coalescence vs. Concatenation: Sophisticated Analyses vs. First Principles Applied to Rooting the Angiosperms. *Molecular Phylogenetics and Evolution*, *91*, 98-122.

Simpson, K., Day, N., & Trusler, P. (2010). Birds of Australia (8th ed.). Penguin Group, Camberwell.

Smith, B. T., Harvey, M. G., Faircloth, B. C., Glenn, T. C., & Brumfield, R. T. (2013). Target Capture and Massively Parallel Sequencing of Ultraconserved Elements for Comparative Studies at Shallow Evolutionary Time Scales. *Systematic Biology, 63*(1), 83-95.

Snow, D. W. (1961). The Natural History of the Oilbird, *Steatornis caripensis*, in Trinidad, W. I. Part 1. General Behavior and Breeding Habits. *Zoologica: New York Zoological Society, 46,* 27-47.

Snow, D. W. (1962). Notes on the Biology of some Trinidad Swifts. *Zoologica: New York Zoological Society, 47*(12)*,* 129-139.

Springer, M. S., Amrine, H. M., Burk, A., & Stanhope, M. J. (1999). Additional Support for Afrotheria and Paenungulata, the Performance of Mitochondrial versus Nuclear Genes, and the Impact of Data Partitions with Heterogeneous Base Composition. *Systematic Biology, 48*(1), 65-75.

Stamatakis, A. (2014). RAxML version 8: a Tool for Phylogenetic Analysis and Post-Analysis of Large Phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Stephen, S., Pheasant, M., Makunin, I. V., & Mattick, J. S. (2008). Large-Scale Appearance of Ultraconserved Elements in Tetrapod Genomes and Slowdown of the Molecular Clock. *Molecular Biology and Evolution, 25*(2), 402-408.

Streicher, J. W., Schulte, J. A., II, & Wiens, J. J. (2015). How Should Genes and Taxa be Sampled for Phylogenomic Analyses with Missing Data? An Empirical Study in Iguanian Lizards. *Systematic Biology*, *65*(1), 128-145.

Suh, A. (2016). The Phylogenomic Forest of Bird Trees contains a Hard Polytomy at the Root of Neoaves. *Zoologica Scripta*, *45*(S1), 50-62.

Sukumaran, J., & Holder, M. T. (2010). DendroPy: a Python Library for Phylogenetic Computing. *Bioinformatics*, *26*(12), 1569-1571.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods) (4 ed.). Sunderland, MA: Sinauer Associates.

Sytsma, K. J., & Pires, J. C. (2001). Plant Systematics in the Next 50 Years – Re-Mapping the New Frontier. *Taxon, 50*, 713-732.

Takahashi, Y., & Ebrey, T. G. (2003). Molecular Basis of Spectral Tuning in the Newt Short Wavelength Sensitive Visual Pigment. *Biochemistry, 42*(20), 6025-6034.

Tan, G., Muffato, M., Ledergerber, C., Herrero, J., Goldman, N., Gil, M., & Dessimoz, C. (2015). Current Methods for Automated Filtering of Multiple Sequence Alignments Frequently Worsen Single-Gene Phylogenetic Inference. *Systematic Biology*, *64*(5), 778-791.

Thewissen, J. G. M., & Nummela, S. (Eds.). (2008). *Sensory Evolution on the Threshold: Adaptations in Secondarily Aquatic Vertebrates* (1st ed.). University of California Press, Berkeley.

Thomas, B.T. (2017). Oilbird (Steatornithidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52261

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F., & Higgins, D. G. (1997). The CLUSTAL_X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Research*, *25*(24), 4876-4882.

Tonini, J., Moore, A., Stern, D., Shcheglovitova, M., & Orti, G. (2015). Concatenation and Species Tree Methods Exhibit Statistically Indistinguishable Accuracy under a Range of Simulated Conditions. *PLoS Currents, 7,* ecurrents.tol.34260cc27551a527b124ec5f6334b6be.

L. (1963). *The Vertebrate Eye and its Adaptive Radiation.* New York, NY: Hafner Publishing Company.

Warnefors, M., Hartmann, B., Thomsen, S., & Alonso, C. R. (2016). Combinatorial Gene Regulatory Functions Underlie Ultraconserved Elements in Drosophila. *Molecular Biology and Evolution*, *33*(9), 2294-2306.

Wells, D. (2017). Treeswifts (Hemiprocnidae). In: J. del Hoyo, A. Elliott, J. Sargatal, D. A. Christie & E. de Juana (Eds.), *Handbook of the Birds of the World Alive*. Retrieved May 22, 2017, from www.hbw.com/node/52267

Wetmore, A. (1919). On the Anatomy of Nyctibius with Notes on Allied Birds. *Proceedings of the US National Museum, 54,* 577-586.

White, N. D., Barrowclough, G. F., Groth, J. G., & Braun, M. J. (2016). A Multi-Gene Estimate of Higher-Level Phylogenetic Relationships among Nightjars

(Aves: Caprimulgidae). *Ornitología Neotropical*, *27*, 223-236.

White, N. D., Mitter, C., & Braun, M. J. (2017). Ultraconserved Elements Resolve the Phylogeny of Potoos (Aves: Nyctibiidae). *Journal of Avian Biology*, *48*, 872-880.

White, N. D. (2017, March 24). *Indel-Stats: v1.0*. Zenodo. Retrieved from http://doi.org/10.5281/zenodo.438047

White, N. D. (2017, May 23). *GC_variance: Another Release*. Zenodo. Retrieved from http://doi.org/10.5281/zenodo.582627

White, N. D. (2017, July 19). *PAUP_parsers: v1.0*. Zenodo. Retrieved from http://doi.org/10.5281/zenodo.832357

White, N. D. (2017, July 19). *Locus_Simulator: v1.0*. Zenodo. Retrieved from http://doi.org/10.5281/zenodo.832352

Whitney, B. M., & Pacheco, J. F. (2003). Reassignment of Chordeiles vielliardi Lencioni-Neto, 1994, to Nyctiprogne Bonaparte, 1857, with Comments on the Latter Genus and Some Presumably Related Chordeilines (Caprimulgidae). *Bulletin of the British Ornithologists' Club, 123*(2), 103-112.

Wilkie, S. E., Robinson, P. R., Cronin, T. W., Poopalasundaram, S., Bowmaker, J. K., & Hunt, D. M. (2000). Spectral Tuning of Avian Violet- and Ultraviolet-Sensitive Visual Pigments. *Biochemistry*, *39*(27), 7895-7901.

Woolfe, A., Goodson, M., Goode, D. K., Snell, P., McEwen, G. K., Vavouri, T., Smith, S. F., North, P., Callaway, H., Kelly, K., Walter, K., Abnizova, I., Gilks, W., Edwards, Y. J. K., Cooke, J. E., Elgar, G. (2005). Highly Conserved Non-Coding Sequences are Associated with Vertebrate Development. *PLoS Biology, 3*(1), e7.

Wu, Y., Hadly, E. A., Teng, W., Hao, Y., Liang, W., Liu, Y., & Wang, H. (2016). Retinal Transcriptome Sequencing Sheds Light on the Adaptation to Nocturnal and Diurnal Lifestyles in Raptors. *Nature Scientific Reports, 6*(33578).

Yang, Z. (2007). PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and Evolution*, *24*(8), 1586-1591.

Yates, A., Akanni, W., Amode, M. R., Barrell, D., Billis, K., Carvalho-Silva, D., Cummins, C., Clapham, P., Fitzgerald, S., Gil, L., Girón, C. G., Gordon, L., Hourlier, T., Hunt, S. E., Janacek, S. H., Johnson, N., Juettemann, T., Keenan, S., Lavidas, I., Martin, F. J., Maurel, T., McLaren, W., Murphy, D. N., Nag, R., Nuhn, M., Parker, A., Patricio, M., Pignatelli, M., Rahtz, M., Riat, H. S., Sheppard, D., Taylor, K., Thormann, A., Vullo, A., Wilder, S. P., Zadissa, A., Birney, E., Harrow, J., Muffato, M., Perry, E., Ruffier, M., Spudich, G., Trevanion, S. J., Cunningham, F., Aken, B. L., Zerbino, D. R., & Flicek, P. (2016). Ensembl 2016. *Nucleic Acids Research*, *44*(Database issue), D710-D716.

Yokoyama, S., Radlwimmer, F. B., & Blow, N. S. (2000). Ultraviolet Pigments in Birds Evolved from Violet Pigments by a Single Amino Acid Change. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(13), 7366-7371.

Yokoyama, S., Yang, H., & Starmer, W. T. (2008). Molecular Basis of Spectral Tuning in the Red- and Green-Sensitive (M/LWS) Pigments in Vertebrates. *Genetics*, *179*(4), 2037-2043.

Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de Novo Short Read Sssembly using de Bruijn Graphs. *Genome Research*, *18*(5), 821-829.

Zhang, X., Wensel, T. G., & Yuan, C. (2006). Tokay Gecko Photoreceptors Achieve Rod-Like Physiology with Cone-Like Proteins. *Photochemistry and Photobiology*, *82*(6), 1452-1460.

Zhao, H., Rossiter, S. J., Teeling, E. C., Li, C., Cotton, J. A., & Zhang, S. (2009). The Evolution of Color Vision in Nocturnal Mammals. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(22), 8980-8985.

Zwickl, D. J. (2006). Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Datasets under the Maximum Likelihood Criterion. (Doctoral dissertation, University of Texas at Austin, pp. 1-125). Austin.

Zwickl, D. J., Stein, J. C., Wing, R. A., Ware, D., & Sanderson, M. J. (2014). Disentangling Methodological and Biological Sources of Gene Tree Discordance on Oryza (Poaceae) Chromosome 3. *Systematic Biology*, *63*(5), 645-659.