

Extracting phylogenetic signal from phylogenomic data: Higher-level relationships of the nightbirds (Strisores)

Noor D. White*, Michael J. Braun

Department of Vertebrate Zoology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20560, USA

Behavior, Ecology, Evolution and Systematics Program, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Keywords:

Phylogenomics
Ultraconserved elements
Strisores
Nightbirds
GC content
Model selection

ABSTRACT

A well-resolved phylogeny would facilitate study of adaptation to nocturnality in the avian superorder Strisores, a group that includes both nocturnal and diurnal lineages. Based on previous estimates, it could be hypothesized that there were multiple independent origins of nocturnality in this group. In order to refine the Strisores phylogeny, we generated genome-scale datasets of 2289–4243 ultraconserved elements for 23 taxa representing all major living lineages in the group. Among the considerations for using genome-scale, molecular sequence data in phylogenomic analyses are issues related to GC content, GC variance and their effects on model selection. In this study, we employed a variety of analytical techniques to empirically investigate those issues in our data, as well as biases and errors resulting from alignment trimming, taxon selection, matrix completeness and evolutionary rate variation among sites and across lineages. Extensive analyses revealed conflict within the data, especially in regard to variation in GC content, that would not have been detected by more cursory study. Our results indicate that readily available models of molecular evolution are insufficient to encapsulate all phenomena present in genome-scale matrices, and that this problem may be at the root of many current issues in phylogenomic analysis. The analytical methods employed in this study are relevant to phylogenomic analysis of any large, heterogeneous matrix. In conclusion, we present a strongly supported estimate of the Strisores tree, and discuss visual adaptations for, and potential evolutionary pathways to, nocturnality in this clade.

1. Introduction

The analytical inquiries described herein were designed with the ultimate goal of resolving the evolutionary history of the Strisores, a superorder of birds that comprises both diurnal and nocturnal members. Strisores includes the diurnal Apodiformes (swifts and hummingbirds) as well as five lineages of nocturnal or crepuscular birds: the Caprimulgidae (nightjars and nighthawks), Nyctibiidae (potoos), Podargidae (frogmouths), Aegothelidae (owlet-nightjars), and Steatornithidae (oilbird). These five nocturnal 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 milk from goats. 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 Apodiformes within the clade of entirely nocturnal or crepuscular nightbirds ruffled some feathers in the ornithological community when it was first proposed based on

morphological data (Mayr, 2002). How could the swifts and hummingbirds—colorful, acrobatic fliers, clearly diurnally-adapted—have descended from the cryptically-colored, secretive nightbirds? Nevertheless, molecular evidence that corroborated the morphological hypothesis also existed (Braun and Huddleston, 2001). Since that early work, many independent phylogenetic studies, based on both molecular and morphological data, have confirmed this topology (e.g. Cracraft et al., 2004; Barrowclough et al., 2006; Ericson et al., 2006; Hackett et al., 2008; Braun and Huddleston, 2009; Mayr, 2010; Prum et al., 2015; Reddy et al., 2017). The group is now given the super-ordinal name Strisores (Mayr, 2010), a name 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. 1), encompassing 602 currently recognized species, 470 of which are diurnal (Chantler, 2017; Cleere, 2017; Holyoak, 2017a,b; Schuchmann and 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; Simpson et al., 2010; Cleere, 2010). Interestingly, Paleogene (Paleocene-Oligocene) fossils

* Corresponding author at: National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA.

E-mail address: noordwhite@gmail.com (N.D. White).

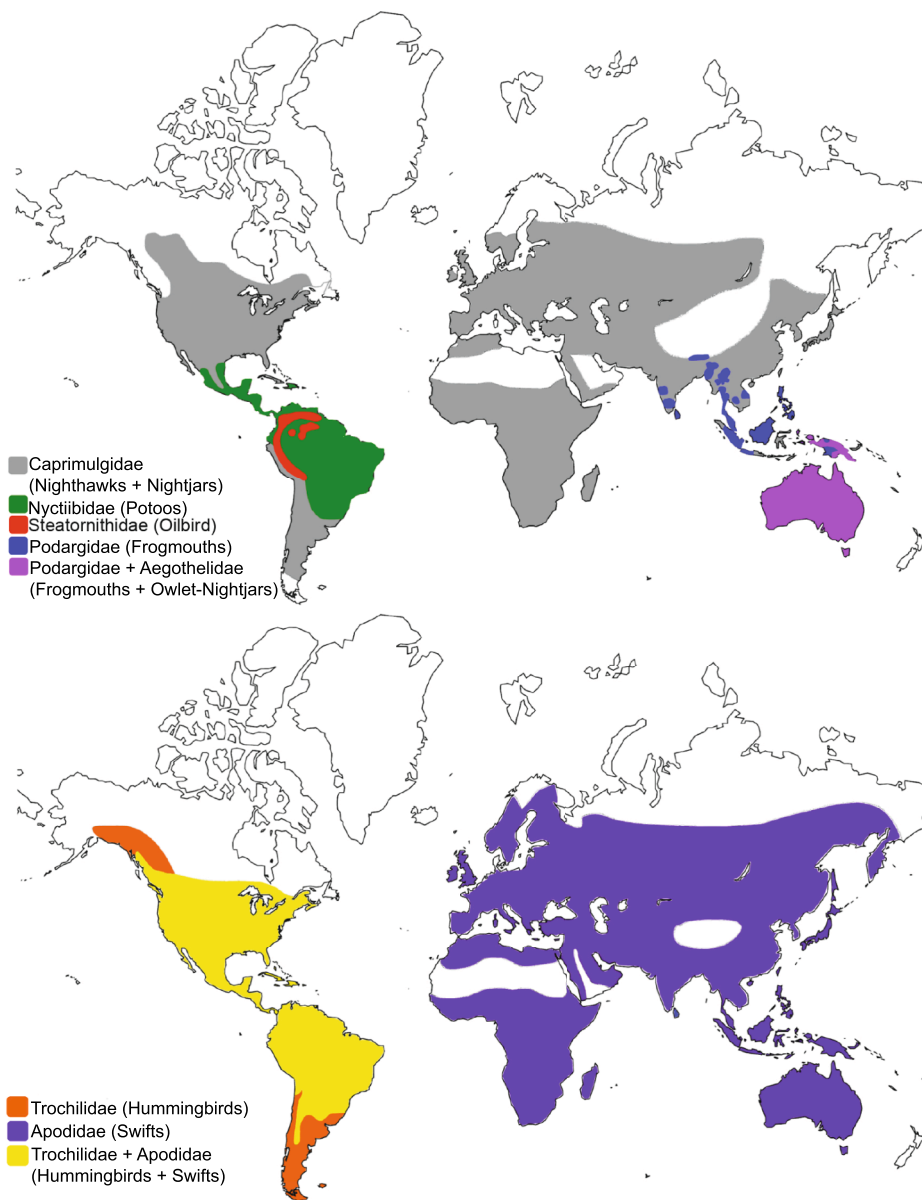


Fig. 1. Range map of the major Strisores lineages. Nighthawks + nightjars overlap the frogmouths + owlet-nightjars distribution, with the exception that nightjawks + nightjars do not occur on Tasmania. Map based on data from the Handbook of the Birds of the World (Chantler 2017; Cleere 2017; Holyoak 2017a,b; Schuchmann & Bonan, 2017; Thomas 2017; Wells 2017).

have been found in Europe for all Strisores lineages except the frogmouths and owlet-nightjars, suggesting a much more widespread distribution than what is observed today (Mayr, 2004; 2005a,b; 2009, 2017). Caprimulgids (nightjars and nighthawks) are found in tropical and temperate areas throughout the world, as are swifts. Hummingbirds currently have a broad distribution restricted to the New World, but appear in the fossil record of Europe (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. 2). The datasets used in these studies have progressed from a few genes (Ericson et al., 2006; Braun and Huddleston 2009) to tens of genes (Hackett et al., 2008; Reddy et al., 2017) to hundreds of genes (Prum

et al., 2015), but strong support for the earliest branches in the phylogeny has remained elusive.

In all of these studies, the oilbird (*Steatornis caripensis*) and potoos were either sister taxa or their positions were unresolved. 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 (their combined analysis was unresolved). 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. The most recent molecular estimates of Strisores topology, Prum et al. (2015) and Reddy et al. (2017), both recover oilbird + potoos, with relatively low support, but differ in the

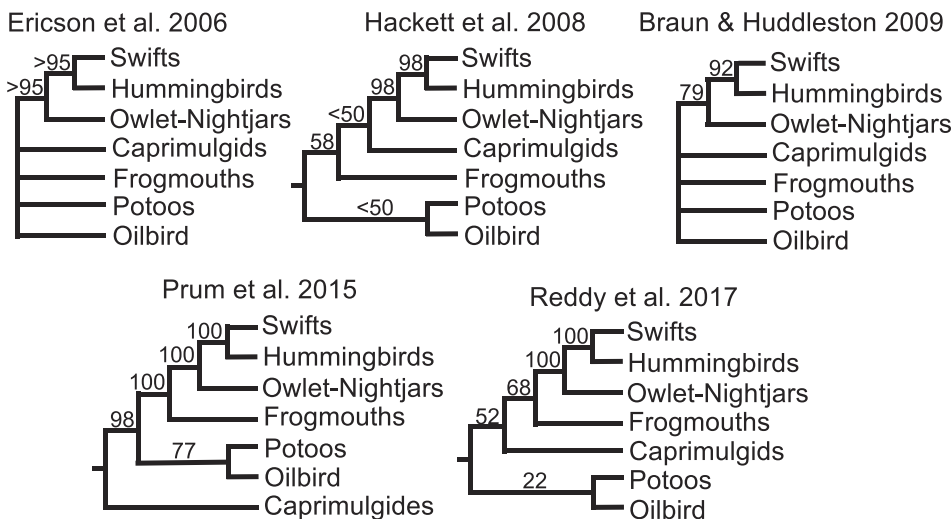


Fig. 2. Previous molecular estimates of the Strisores phylogeny. Support values are bootstraps, with the exception of [Ericson et al. \(2006\)](#), which are Bayesian posterior probabilities. [Ericson et al. \(2006\)](#) collapsed all nodes with PP < 0.95. The [Braun and Huddleston \(2009\)](#) topology shown is from their combined analysis. [Prum et al. \(2015\)](#) and [Reddy et al. \(2017\)](#) support values are from the maximum likelihood trees available in their respective supplementary material.

placement of the caprimulgids. Incongruence between these two analyses could be due to data type—the [Prum et al. \(2015\)](#) dataset is dominated by exons, while the [Reddy et al. \(2017\)](#) data is mostly intronic sequence—and warrants further investigation.

Morphological results are also unresolved, and independent studies disagree both in the result 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 oilbird + potoos in several independent studies ([Cracraft, 1981; 1988; Mayr, 2002, 2010; Mayr et al., 2003; Nesbitt et al., 2011](#)). However, most studies of molecular or combined morphological and molecular datasets have not recovered that group ([Mayr et al., 2003; Ericson et al., 2006; Hackett et al., 2008; Braun and Huddleston 2009; Prum et al., 2015; White et al., 2016; Reddy et al., 2017; White et al., 2017](#)), the one exception being the combined morphological and molecular analysis in [Nesbitt et al. \(2011\)](#). Additionally, extensive morphological work by [Livezey and Zusi \(2006, 2007\)](#) found two monophyletic groups, the Caprimulgiformes and the Apodiformes, sister to one another, in conflict with most other published work, including [Mayr \(2008\)](#), who directly contested Livezey and Zusi's scoring of two characters (beak morphology and the presence of a tapetum lucidum). 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.

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 et al., 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; Livezey and Zusi, 2007](#)).

A statistically sound phylogenetic tree will provide the groundwork necessary to elucidate the evolutionary origins of nocturnality in Strisores and allow mapping of molecular adaptations that might have facilitated the transition(s) between nocturnality and diurnality in this clade. In this study we wanted to take advantage of modern, genome-scale molecular markers to generate an unprecedentedly large matrix for resolving the Strisores. Thus, we employed targeted sequence capture.

Targeted sequence capture methods have enabled the efficient generation of genome-scale phylogenetic data matrices for dozens or hundreds of

taxa ([Glenn and Faircloth, 2016](#)). 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 and Braun, 2016; Branstetter et al., 2017; White et al., 2017](#)). UCE data are collected via an in-solution hybridization approach using oligonucleotide baits complementary to the ultraconserved sequence. These baits are used to capture DNA fragments of varying lengths from randomly-sheared DNA libraries, and through bioinformatic assembly of the resulting sequencing reads, loci longer than the ultraconserved cores are retrieved. These loci display increasing sequence variation among species in either direction from the core, and it is this flanking sequence that provides phylogenetic information.

UCE function is an active area of research in molecular genetics, as 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; Dickel et al., 2017](#)). UCEs, or other regions of similar conservation, have been identified in a wide variety of taxonomic groups, including vertebrates, insects, yeasts, and plants (e.g. [Lockton and Gaut, 2005; Siepel et al., 2005; Stephen et al., 2008; McCormack et al., 2012; Faircloth et al., 2015; Starrett et al., 2017](#)), and hold great potential for the resolution of the phylogenies of these groups. Due to the wide (and still growing) utility of UCEs, a substantial investigation of the effect of potential biases, such as GC content, or methodological choices, such as trimming algorithm, or investigation of the source of phylogenetic support, is warranted. These investigations allow us to explore the putative sources of phylogenetic support and potential conflict.

It has been well documented that use of currently available phylogenetic methods with genome-scale matrices (of any data type) can lead to systematic error—increased statistical confidence in the wrong answer with increased data points (reviewed in [Phillips et al., 2004; Kumar et al., 2012; Hahn and Nakhleh, 2015; Hosner et al., 2016](#)). 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 and Degnan, 2007; Philippe et al., 2011; Salichos and Rokas, 2014; Zwickl et al., 2014; Borowiec et al., 2015; Kocot et al., 2016; Suh, 2016](#)). In some cases, data type (e.g., coding vs. non-coding DNA sequence) may have a greater effect on phylogenetic inference than taxon sampling due to issues related to

model selection (Reddy et al., 2017). This is a potential issue for UCEs, as they likely overlap several functional categories of the genome and represent both coding and non-coding sequence (e.g., Harmston et al., 2013; Pirnie et al., 2016; Warnefors et al., 2016). However, best practices for analyzing a heterogeneous UCE matrix (comprised of both coding and non-coding sequence) have not been established. In this study, we made efforts to classify our UCEs by using functional annotations available in published genomes, and analyze loci from different functional categories separately to uncover potentially different signals.

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 and Lemey, 2009). One of the greatest changes the advent of phylogenomics has brought is that most researchers no longer visually inspect and manually edit alignments. With hundreds or thousands of loci in an individual analysis, manual inspection is onerous, and any effects of the researcher's subjective bias inherent in manual inspection will only be exacerbated when summed over the ever-increasing number of locus alignments. Thus, we rely on automated methods of alignment and trimming. Trimming is a particularly important issue with sequence capture datasets. Untrimmed alignments of sequence capture data have ragged ends due to random laboratory and sequencing effects resulting 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 and 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 as 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 and 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).

The effect of taxon sampling on phylogenetic inference has long been a topic of study, as the fewer taxa are included, the more difficult it is to identify homoplasy. One concern is the phenomenon of long-branch attraction, wherein two evolutionarily distant lineages that have undergone a lot of evolutionary change relative to the other taxa in the dataset are incorrectly inferred to be closely related in the tree (Felsenstein, 1987). Adding taxa intentionally chosen to break up putatively long branches alleviates potential risk of this artifact (Hendy and Penny, 1989). Following numerous independent studies, it is generally accepted that increased taxon sampling usually has beneficial effects on model-based phylogenetics, including increasing the accuracy of model estimation (e.g.; Graybeal, 1998; Omland et al., 1999; Johnson, 2001; Braun and Kimball, 2002; Pollock et al., 2002; Zwickl and Hillis, 2002; Reddy et al., 2017; Tamashiro et al., 2019), although the effects are context-dependent and can be detrimental in some cases (Poe and Swofford, 1999). In this study, we tested different taxon samplings centered around resolving the relationship between oilbird and potoos (as much as was possible due to the fact that oilbirds are a monotypic genus) in order to look for potential effects on phylogeny estimation, and increase our chances of recovering the correct evolutionary history.

When using a high-throughput hybrid-enrichment approach to collect data, some level of missing data is expected. One can either use as much data as possible (accepting that there will likely be some loci missing substantial numbers of taxa), define a threshold for the proportion of taxa that must have data at each locus in order for it to be included in the matrix, or require that all taxa have data for a locus to be included. The first strategy can produce matrices for which very few taxa are present at many loci, potentially introducing systematic error and increasing

computing times with little tradeoff benefit (Roure et al., 2012; Hosner et al., 2016). Alternatively, the last strategy can unnecessarily reduce the size of the usable dataset, as some locus/taxon combinations will be missing due to random processes in the associated labwork, rather than any biological reality. Studies (including some using UCE data) have demonstrated that it is beneficial to test different levels of matrix completeness (as in the second strategy), as they can affect topology inference, and further showed that including loci with some missing taxa was preferable to excluding all missing data (e.g. Philippe et al., 2004; McCormack et al., 2013; Streicher et al., 2015). In this study we chose to test both 100% complete matrices and matrices for which 75% of taxa must be present at a given locus in order for it to be included.

The generation of genome-scale datasets for phylogenetics exacerbates issues of 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 or expectation of conservation. Partitioning the data and applying different models of sequence evolution to subsets of a data matrix that have evolved under different functional constraints is a practical methodological choice (Bull et al., 1993; de Queiroz, 1993), and has been implemented in UCE analyses (e.g. Sun et al., 2014; Meiklejohn et al., 2016; Hosner et al., 2016). In this study, we sought to identify particular characteristics of our dataset to implement the most appropriate partitioning method possible. Initial inspections of the data indicated GC content varied greatly both across individual UCE loci, as well as across taxa within a given alignment. 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 and 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).

In this study, we sought to construct the best possible estimate of the Strisores phylogeny by designing analyses both to investigate potential issues related specifically to the Strisores tree, as well as to using UCEs as a phylogenetic marker. The analytical approaches that we selected addressed known and potential issues in phylogenomic analysis, and paid particular attention to issues that are especially pertinent to UCE data, including trimming, nucleotide composition and matrix completeness. We also tested the effect of taxon sampling on a key node in this tree, the potoo-oilbird relationship. The results of this study are relevant to any genome-scale dataset, whether those data result from target-capture methods or simply a very large, heterogeneous matrix.

2. Materials and methods

2.1. UCE functional characterization

The UCE probe sequences used in this study are those targeting tetrapod/amniote species, described by McCormack et al. (2012) and Faircloth et al. (2012), and available from < ultraconserved.org > . They include 5472 probes targeting 5060 UCE loci. To investigate the potentially disparate evolutionary constraints on these loci, we sought to identify where UCEs overlapped with existing functional annotations in published genomes. Assemblies of the chicken (galGal5; Hillier et al., 2004) and human (hg38; Lander et al., 2001) genomes were downloaded from the University of California Santa Cruz (UCSC) genome browser (Kent et al., 2002). The UCE probe sequences were mapped to each genome using Bowtie2 (v. 2.3.0; Langmead and Salzberg 2012), aligning "end-to-end" and without an upper limit on number of matches ('-a'). Only instances where an UCE probe mapped uniquely were kept going forward. Annotation files for CpG islands, 3' untranslated regions (UTRs), 5' UTRs, protein-coding sequences (CDS), exons and introns were downloaded from the UCSC genome table browser for each

genome (Kent et al., 2002; 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 and Hall, 2010), we identified where coordinates of our UCE probes mapped to each genome intersected coordinates of an annotation in their respective genome.

To investigate whether UCE probes intersected functional annotations more frequently than expected by chance, we simulated UCEs by randomly selecting short sequences from each genome that are the same length as our probes (120 bp). We 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" (White, 2017a), and was repeated 100 times. These 100 "simulated" region sets (each comprised of 5472 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.

2.2. Taxon Sampling, data collection and alignment

For each of the major Strisores lineages, two representatives were included, except for oilbird and potoos (all specimen information in Supplementary Table S1). The oilbird is monotypic, so no additional taxon sampling is possible for that lineage. Potoos are an unusual and ancient major lineage of the Strisores (Mariaux and Braun, 1996; Brumfield et al., 1997; White et al., 2017), and preliminary analyses of datasets with two potoos revealed some instability in placement of the oilbird relative to the potoos, so four taxon samplings were chosen to assess the effect of potoo inclusion/exclusion on phylogeny (Table 1). In particular, we wanted to test the effect of including or excluding the rufous potoo, *Phyllaemulor bracteatus*. *Phyllaemulor* is the oldest and most divergent living potoo lineage, whose systematic distinctiveness was recently recognized by assignment to a monotypic genus (Costa et al., 2017). It also had one of the lowest number and shortest average lengths of UCE contigs of taxa in this study (Supplementary Table S2), resulting in it being represented by the least data in many analyses, and warranting further investigation of the effect of its inclusion or exclusion. Lastly, three unambiguous outgroups were chosen for a total of 23 species.

Cleaned, aligned UCE data were available for potoos from a previous publication (White et al., 2017). For all other species, frozen tissue samples were assembled through our own fieldwork and loans from major museum collections, and DNA was extracted using a phenol–chloroform protocol (Rosel and Block, 1996).

UCE and flanking sequences 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 a KAPA Biosystems library preparation kit (KAPA Biosystems, Inc.). Libraries were enriched for UCE loci contained in the Tetrapod/Amniote probe set (commercially available from MYcroarray, now Arbor Biosciences) using chicken Cot-1 DNA as Block #1 (Applied Genetics Laboratories, Inc.), and a 24-hour hybridization at 65 °C.

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). We used the PHYLUCE pipeline (Faircloth, 2015) to process this data, beginning with quality control and trimming via Trimmomatic (Faircloth, 2013; Bolger et al., 2014). Reads were assembled *de novo* using Velvet (ver. 1.2.09; Zerbino and Birney, 2008), and contigs matched to UCE probe sequences using LASTZ (Harris, 2007). Contigs without a UCE match, that matched more than one UCE locus, or for which less than three taxa were represented were discarded. Single locus alignments were conducted using SATé-II based on MAFFT (ver. 2.2.7; Liu et al., 2011).

2.3. Trimming and matrix generation

In this study, we analyzed untrimmed UCE alignments along with those trimmed with algorithms available in the PHYLUCE pipeline and Aliscore software (v2.0). We ran the PHYLUCE trimming algorithm with default settings. This meant that 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.

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 and Meusemann, 2010; Misof and Misof, 2009). We ran Aliscore with the default parameters (sliding window of 6 bp, 4 × N random pairs checked).

We generated matrices based on two levels of matrix completeness, by which we mean the proportion of taxa for which a given locus is represented. Those levels were 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. 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, three trimming treatments and the four taxon samplings detailed in Table 1, we had 24 major concatenated matrices to analyze. Lastly, matrices containing only those loci that intersected different functional annotation categories were created for the matrices of taxon sets 1 and 4.

Dataset names follow a four-character convention: a "D" for dataset, a number from 1 to 4 indicating which taxon set they represent, a "C" or "I" indicating that they are complete or incomplete (75%) matrices, and a letter indicating what trimming method was used—"U" for untrimmed, "P" for PHYLUCE, and "A" for Aliscore.

2.4. Characterization of nucleotide composition

We wanted to build specific analyses into our study to account for two potential sources of compositional bias—GC variation among loci and GC variation among taxa—based on our preliminary observations that this dataset contained high GC variation. For each locus, we therefore calculated both the average GC content over the locus alignment, and the variance in GC content among the taxa in the locus alignment (GC variance). GC variance was calculated using a custom Perl script "GC_variance.pl" (White, 2017b), and is defined as the mean squared difference of an individual taxon's GC content from the average for that locus. We addressed compositional bias by recoding the nucleotide data into purines and pyrimidines (RY-coding) and rebuilding the 24 major matrices for additional analyses. We also ran analyses with the data partitioned by locus, allowing for GC content to be more accurately reflected by the model applied to each locus, rather than assuming an average GC content for all loci or subsets of loci. In order to remove some extreme cases of bias, we created matrices with reduced GC variance across taxa by excluding the 10% of loci showing the highest GC variance across taxa for all taxon set + completeness + trimmer combinations. We refer to these as the low GC variance matrices.

2.5. Model selection and topology searches

Model selection was conducted in PAUP* (ver. 4.0a151; Swofford, 2003) for every concatenated matrix used herein, and the most appropriate model was selected according to the corrected Akaike information criterion (Akaike, 1973; Hurvich, and Tsai, 1989). An exception to this are the RY-coded matrices, for which few models are available in a readily implemented form and the BINGAMMAI model in RAXML was used (ver. 8.2.9; Stamatakis, 2014).

Table 1

The four taxon sets used in this study vary in representation of the potoos (Nyctibiidae), particularly the presence or absence of the rufous potoo (*Phyllaemulor bracteatus*). Gray box = taxon present. Naming convention for the rufous potoo follows Costa et al. (2017).

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

GARLI (ver. 2.1; Zwickl, 2006) was used to conduct ML searches on the concatenated nucleotide matrices, the low GC variance matrices, and the matrices subset by UCE functional category. 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 further in White et al. (2016). One hundred non-parametric bootstrap replicates were conducted, with one search replicate each. Bootstrap values were plotted on the optimal topology using the SumTrees program in the python library DendroPy (ver. 4.2.0; Sukumaran and Holder, 2010).

RAxML was used to conduct ML searches on the concatenated RY matrices, as well as to construct all single locus trees (both nucleotide and RY-coded). ML searches of the RY matrices consisted of 20 search replicates, with bootstraps run using the bootstopping criterion ('-autoMRE'). Single locus tree searches 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.

IQTREE (ver. 1.6.5; Nguyen et al., 2015; Chernomor et al., 2016) was used to conduct model selection for individual UCE loci, to conduct the locus-partitioned ML searches, and to calculate concordance factors (ver. 1.7-betaX; Minh et al., 2018). IQTREE (ver. 1.6.9; Nguyen et al., 2015) was also used to conduct ML searches under the FreeRates ('-mrate R'; Yang 1995; Soubrier et al., 2012) and Heterotachy ('-mrate H'; Lopez et al., 2002) models. For these searches, 1000 ultrafast bootstrap replicates were performed (Minh et al., 2013).

ASTRAL (ver. 5.6.3; Zhang et al., 2018) was used to conduct gene tree summarization searches on both nucleotide and RY-coded single locus

trees. Gene tree summarization methods were designed to address gene-tree versus species-tree issues sometimes inherent in concatenated analysis. Coalescent-based species tree analyses were run using SVDquartets (Chifman and Kubatko, 2014, 2015) as implemented in PAUP* (ver. 4.0a151). SVDquartets analyses were run evaluating all possible quartets and implementing 100 non-parametric bootstrap replicates.

Indel number and length were calculated using the simple coding algorithm as implemented in 2matrix (Salinas and Little, 2014) and the output was processed using a custom Perl script "Indel_stats.pl" (White, 2017c). We used the "DescribeTree" function in PAUP* to score indel characters on the matrix generated by 2matrix for tree topologies of particular interest. The custom Perl script "PAUP_Indel_parser.pl" was used to process the output (White, 2017d). 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 our data, we downloaded the published data of Prum et al. (2015; doi:https://doi.org/10.5281/zenodo.28343; trimmed, concatenated matrix "Concatenated.phy") and Reddy et al. (2017; doi:https://doi.org/10.5061/dryad.6536v; "Reddy_sup_fileS5_ALLtaxset.nex"). Any taxa not within Strisores were pruned from the aligned matrices, 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). The Reddy et al. (2017) taxon name *Nyctibius bracteatus* was changed to *Phyllaemulor bracteatus* to reflect the taxonomic update of Costa et al. (2017). After the removal of any columns that were solely gaps, this

resulted in a 394,684 bp matrix with 16 taxa for the Prum et al. (2015) data, and a 71,075 bp matrix with 24 taxa for the Reddy et al. (2017) data. Data were RY-coded, and ML searches conducted in RAXML with 200 non-parametric bootstraps run.

3. Results

3.1. Characterization of UCE probes

Of the 5472 UCE probe sequences, all but one mapped to the chicken genome (overall alignment rate 99.98%) and 5222 mapped uniquely (Supplementary Table S3a). In the human genome, 1571 sequences did not map (overall alignment rate 71.29%), and 3810 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. All three results were statistically significant ($P > 0.05$), potentially indicating an important functional role for some UCEs, either as encoding functional elements themselves (exons, CDS), or in the regulation of expression (UTRs). However, only a relatively small fraction of mapped probes intersected functional annotations (e.g., 348 of 5222 for chicken exons, 590 of 3810 for human exons), suggesting that the role of many UCEs lies outside these domains. 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 AT-rich content of most UCE loci (Fig. 3).

3.2. Nucleotide composition of UCE data

While there was little variation in average GC content among concatenated data matrices (38.8–39%), initial characterization of our data revealed a high degree of variation in average GC content among UCE loci (24.9–71.2%; Fig. 3). Similarly, while the variance in GC content among taxa was low for most UCE loci, the range was large. GC content was affected little by trimming, but GC variance across taxa was slightly slower in datasets trimmed with Aliscore than those trimmed by PHYLUCE or left untrimmed. Looking at GC content across taxa, we observed little variation among taxa for the full matrices, but, for parsimony informative sites, we observed an anomalously high GC content shared by two lineages we were particularly interested in resolving—the oilbird and potoos (Fig. 4; the same pattern was observed for GC variance as well). These observations prompted us to incorporate analyses that would reduce the effect of base compositional bias.

3.3. Phylogenetic analysis

Sequence capture and data processing produced 4243 usable UCE loci. When loci with up to 25% missing taxa (75% completeness) were included, the number of loci in the 24 major datasets ranged from 4194 to 4243. Requiring 100% completeness reduced the number of loci substantially, to a range of 2289–2588 (Fig. 5). Trimming also had a dramatic effect on the size of the datasets, reducing their length by more than 50% in some cases. Nevertheless, concatenated matrices ranged from 1.2 to 4.1 Mbp in length, representing a sizeable increase in data from previous studies of the group (e.g., the full Reddy et al. [2017] matrix is 137,463 bp in length and the full Prum et al. [2015] matrix is 394,684 bp). The UCE matrices are heterogeneous in that they contain both coding and non-coding sequence, nearly invariant UCE core elements and more variable flanking regions, and the wide variation in GC content detailed above.

Across all analyses conducted in this study, we recovered three principal topologies for the Strisores tree, which differ only in the placement of oilbird in relation to potoos (Fig. 5). The first, Topology A, has potoos and oilbirds in a monophyletic clade. The second, topology

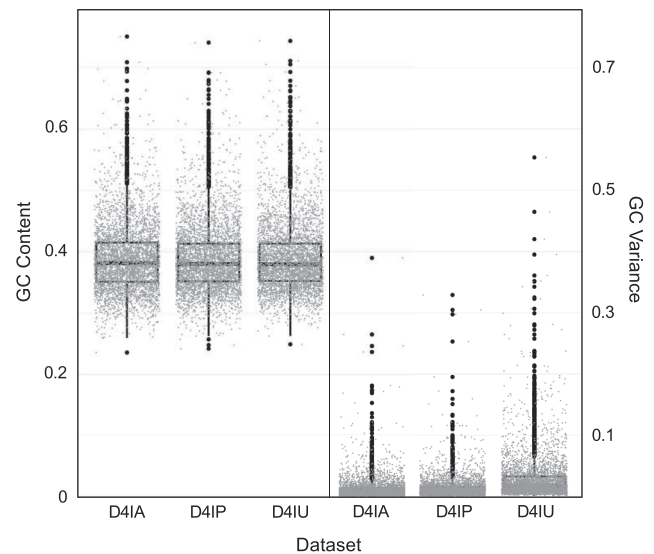


Fig. 3. Variation in GC content (left panel) and GC variance (right panel) for loci of the incomplete datasets of taxon set 4. Small gray dots are individual data points summarized by the underlying box and whisker plot. Median is represented by the line bisecting the box and separates the first and third quartiles. Whiskers extend to values no further than 1.5X the interquartile range from the median. Large black dots represent outliers. All other matrices showed similar patterns of variation as the examples shown here.

B, places oilbird as branching off the Strisores tree before potoos, and the third, topology C, has potoos branching off first. All other relationships in the tree were very highly supported (99–100% bootstrap support) in ML analyses of all matrices for all taxon sets. With minor exceptions, one of these three topologies were recovered by every analysis, but the frequency of recovery and support for each topology varied among analyses. Analyses of taxon sets 2 and 3, which did not include *Phyllaemulor*, produced topology A more often than taxon sets 1 and 4, which did include *Phyllaemulor*. Analyses of 75% complete matrices often produced higher bootstrap support values than the same analyses of 100% complete matrices, presumably because 75% complete matrices include more loci and more data. Similarly, untrimmed matrices often produced higher bootstrap support values than trimmed matrices, again presumably due to the inclusion of more data. However, none of these effects was consistent across all analyses.

On the other hand, accounting for variation in GC composition did appear to have a significant impact on our results. The wide ranges of GC content among loci and among taxa (Figs. 3 and 4) mean that loci with extreme values for these parameters will not be well fit by ML models that assume a single ratio of nucleotide composition over an entire matrix. An initial attempt to reduce GC bias was to implement RY coding of the data. For the concatenated analyses, this method resulted in a dramatic shift from the majority of nucleotide matrices (19 of 24) preferring topology A to all but one RY-coded dataset producing topology B or C (Fig. 5). This trend was not seen in gene tree summary analyses, however, where 14 of 24 nucleotide datasets and 13 of 24 RY-coded datasets produced topology A. While we don't have an explanation for the dramatic shift in concatenated results at this time, it could be due to a reduction in information content of the matrix. UCE loci individually have a low information content to begin with, with our overall matrices containing 7–8.8% parsimony-informative sites. RY-coding reduces the number of character states of the matrix from 4 to 2 and only considers transversions, inherently removing a substantial amount of potentially informative variation from the datasets.

We next attempted to reduce the effect of GC bias by dropping the 10% of loci with the highest GC variance among taxa, based on the patterns of GC variance seen in Fig. 3 (right panel). This data filter resulted in the topology preferred by dataset D1CP shifting from B to A

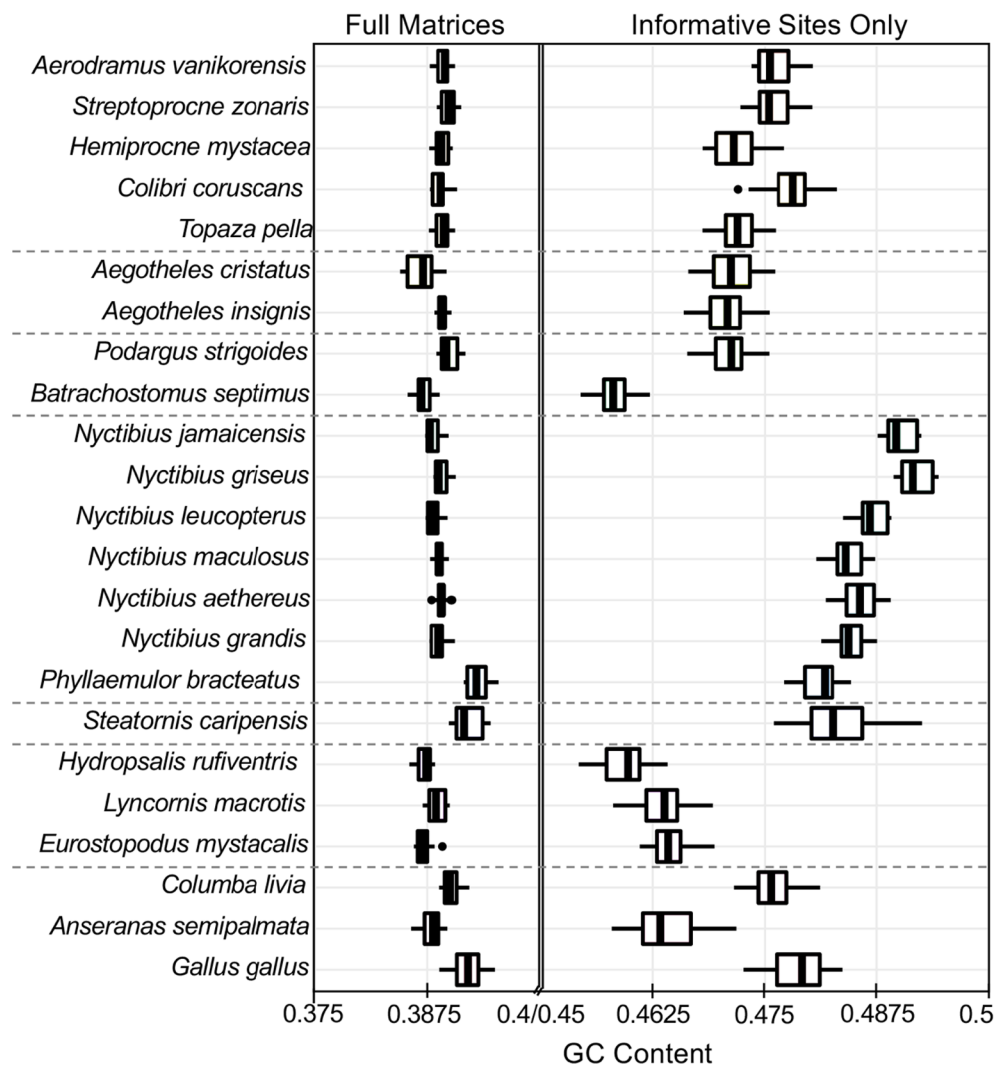


Fig. 4. Variation in GC content averaged by taxon for all matrices in which they appear. The left portion of the graph shows GC content over full alignments, the right over parsimony informative sites only. Dashed lines demarcate major Strisores lineages. Note the break in the x-axis at 0.40–0.45. Median 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 black dots represent outliers.

in concatenated nucleotide analyses, but had few other effects, indicating that this change did little to improve model fit to the data (Fig. 5).

We also wished to test whether evolutionary rate variation among sites or among lineages could be affecting our results. To do this, we ran those datasets for which majority topology A was not recovered in our concatenated nucleotide analyses (matrices D1CP, D1IP, D2IP, D4CP, D4CA) under the FreeRate and Heterotachy models implemented in IQTREE, which allow evolutionary rate variation among sites and among lineages, respectively. The analyses under the FreeRate model did not change topology in any instance, and increased support modestly at the relevant node in all cases but one (D4CA; [Supplementary Fig. S1](#)). This result suggests that among site rate variation was adequately accounted for by the conventional models with gamma-distributed rates applied in the concatenated analyses. In contrast, the analyses under the Heterotachy model shifted to topology A in all instances but one, with increased support at the relevant node except for the one dataset which did not shift (D4CA; [Supplementary Fig. S1](#)). This result suggests that rate variation among lineages, which is not addressed by conventional models, may have been sufficient to affect the concatenated analyses of these five datasets.

The analytical treatment that produced the most consistent results, in terms of topology across datasets, was partitioning the data by locus and

fitting an independent model of sequence evolution to each partition. This treatment produced topology A for all 24 major datasets, with bootstrap support ranging from 53 to 100% for the monophyly of potoos and oilbird (Fig. 5). This makes sense, as only by partitioning each locus individually do you allow the most appropriate GC content (and other model parameters) to be applied to the data. This treatment won't necessarily resolve high GC variance among taxa, but in our data, variation in GC content was a greater issue than GC variance among taxa.

Coalescent-based species tree analysis of the 24 major datasets resulted in topology A and four additional topologies, none of which were identical to our ML topologies B or C ([Supplementary Fig. S2](#)). However, three of the four novel topologies (and half of the analyses overall) did place oilbird and potoo in a monophyletic clade with reasonably high bootstrap support. Of note, these 3 "novel" topologies only differ from topology A in the placement of the outgroup root.

Considering the locus partitioned analyses as potentially being the most appropriate treatment of our UCE data, given its wide GC content variation, we now turned to other sources of data to ask if they also supported the hypothesis that topology A is the best estimate of the Strisores phylogeny.

Rare genomic changes in the form of insertion-deletion events (indels) found within our UCE loci also provided support for topology A. We found that there were 4 indels supporting topology A, and 2 supporting

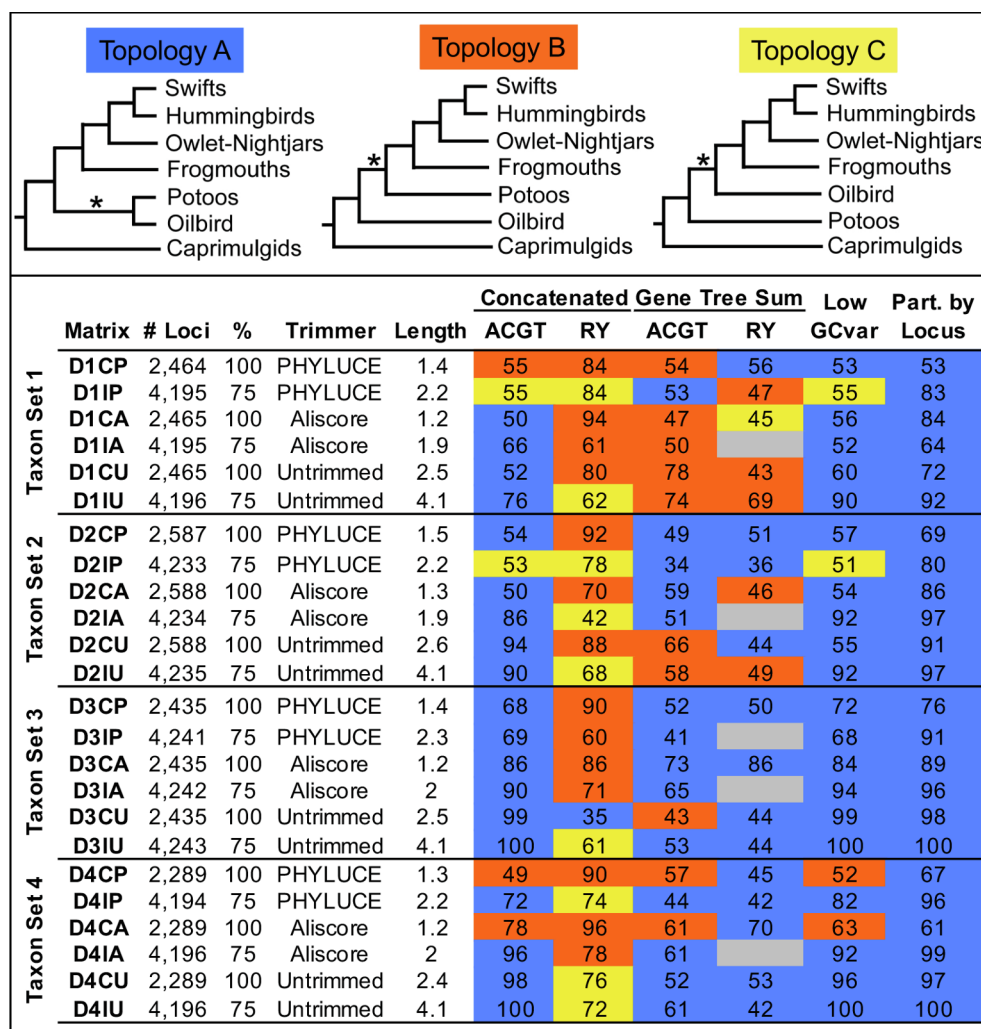


Fig. 5. Phylogenetic results of this study. *Upper panel:* The three alternative topologies for the major lineages of Strisores found across ML analyses. Each topology is color-coded to correspond to the lower panel. Asterisks mark nodes for which relevant bootstrap support is given in the lower panel and other figures and tables in this paper. *Lower panel:* Characteristics of the 24 data matrices used in this study, and topological results and support values from phylogenetic analyses of each. Naming conventions for the matrices are described in Methods. % = percent completeness of matrix. Length is in Mbp. Major categories of data analysis were concatenated matrices, gene tree summarization, low GC variance matrix (“Low GCvar”) analysis and locus partitioned analysis (“Part. by locus”) as described in Methods. ACGT and RY refer to matrices coded as nucleotides or purines/pyrimidines, respectively. Numbers indicate bootstrap support values for nodes labeled with an asterisk in the upper panel. Gray boxes indicate a topology other than those shown in the top panel was found.

topology C, and none supporting topology B. These indels were present in all datasets of taxon sets 1 and 4 (Supplementary Fig. S3).

Analyses of matrices created from the UCE functional categories found in this study resulted in some datasets preferring topology A (Supplementary Table S3b). Of particular note, introns, a popular marker choice for avian phylogenetics, preferred topology A in most instances. Introns have been demonstrated to be a very useful marker for avian phylogenetics (e.g., Prychitko and Moore, 1997; Chojnowski et al., 2008; Braun and Huddleston, 2009; Hackett et al., 2008; Reddy et al., 2017), and their preference for topology A provides some support for our hypothesis.

An alternative assessment of support is to use concordance factors to look within the data and consider the percentage of sites and/or genes (in this case, UCE loci) that support a given topology. We calculated concordance factor scores of both sites and genes for the three major ML topologies, A, B and C. Across all matrices considered, topology A was preferred by slim margins (Table 2). Gene concordance factors were two to five times higher for topology A than for B or C, but very low in number overall. This was unsurprising as UCE loci vary greatly in information content, and individually have relatively few informative sites relative to non-ultraconserved sequences. Site concordance factors favored topology A by a few percentage points in every case. Stronger evidence was given by the fact that RY-coded matrices showed the same pattern in concordance factor support for topology A as the nucleotide matrices, and did so without regard to matrix completeness, trimming or taxon sampling.

Lastly, we demonstrated sensitivity to GC content in independent, published matrices with similar taxon sampling of the Strisores. The GC content of Strisores matrices from 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*, *Anurostomus carolinensis* (Caprimulgidae) and *Chaetura pelagica* (Apodidae) 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. RY-coded Strisores matrices of Prum et al. (2015) and Reddy et al. (2017) data resulted in substantial topological differences from their nucleotide trees, which had originally placed oilbird and potoos as sister lineages (Fig. 2; Supplementary Fig. S4). RY-coded analysis of the Prum et al. (2015) Strisores dataset resulted in topology C, though with extremely low bootstrap support at the relevant node (3%; bootstrap majority-rule consensus resulted in the sister topology). This result could potentially be due to taxon sampling issues as that dataset included only one potoo, frogmouth and owlet-nightjar (Supplementary Fig. S4a). RY-coded analysis of the Reddy et al. (2017) Strisores matrix resulted in a clade including frogmouths, oilbird and potoos, which was not seen in any of our analyses (Supplementary Fig. S4b). These results could indicate that RY coding simply removes too much phylogenetic information from these datasets to resolve the deep Strisores phylogeny. We included these results here to demonstrate that treatment of GC content does have an impact on the Strisores phylogeny, and to highlight that we employed multiple methods of treating both GC content and GC variance among taxa in this study. Of the methods we utilized, partitioning by locus is the only one that did not require removing any of our data, and improved the fit of our model on as fine a scale as possible. Considering this in light of all of other lines of evidence presented above, we have significant support for that topology A being the best estimate of the Strisores phylogeny.

Table 2

Site (sCF) and gene (gCF) concordance factor support for tree topologies A, B and C (see Fig. 5), in both the nucleotide (ACGT) and RY-coded (RY) matrices. Numbers are shaded by sCF support from lower values (white) to higher values (dark gray), separately for nucleotide and RY-coded results.

Matrix	ACGT			RY		
	A	B	C	A	B	C
D1CP	35 / 8	33 / 4	32 / 3	36 / 5	33 / 1	30 / 1
D1IP	35 / 8	33 / 3	32 / 3	36 / 4	33 / 2	31 / 2
D1CA	35 / 9	33 / 3	32 / 3	36 / 5	32 / 1	32 / 1
D1IA	36 / 8	33 / 3	32 / 3	37 / 4	33 / 1	31 / 1
D1CU	35 / 8	33 / 3	32 / 2	36 / 4	33 / 1	30 / 1
D1IU	35 / 7	33 / 3	32 / 3	35 / 4	33 / 1	32 / 1
D4CP	35 / 8	33 / 3	32 / 3	37 / 5	33 / 1	32 / 1
D4IP	36 / 8	33 / 3	32 / 3	36 / 4	33 / 1	31 / 1
D4CA	36 / 8	32 / 3	32 / 3	36 / 5	32 / 1	32 / 1
D4IA	35 / 8	33 / 3	32 / 3	36 / 4	32 / 2	31 / 1
D4CU	35 / 8	33 / 3	32 / 3	36 / 5	33 / 1	31 / 1
D4IU	35 / 8	33 / 2	32 / 3	37 / 4	32 / 1	31 / 1

4. Discussion

4.1. 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 (Dickel et al., 2017, Pennisi 2017). UCEs appear to be enriched in UTRs, which are useful phylogenetic markers (Harshman et al., 2003; Bonilla et al., 2010). Our 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 modeling 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, as well as methods that can efficiently analyze genome-scale datasets partitioned by different evolutionary models.

4.2. Phylogenetic analysis of Strisores

Custom-designed analyses of the data in this study resolve the Strisores phylogeny as having potoos and oilbird in a monophyletic clade (Fig. 6). We depict this topology with the branch lengths and support values from analysis of the D4IU matrix, as it is the largest dataset we generated. The phylogeny presented in Fig. 6 illustrates why resolving the phylogeny of the major lineages of Strisores has been difficult—internodes in this region of the tree are very short. This

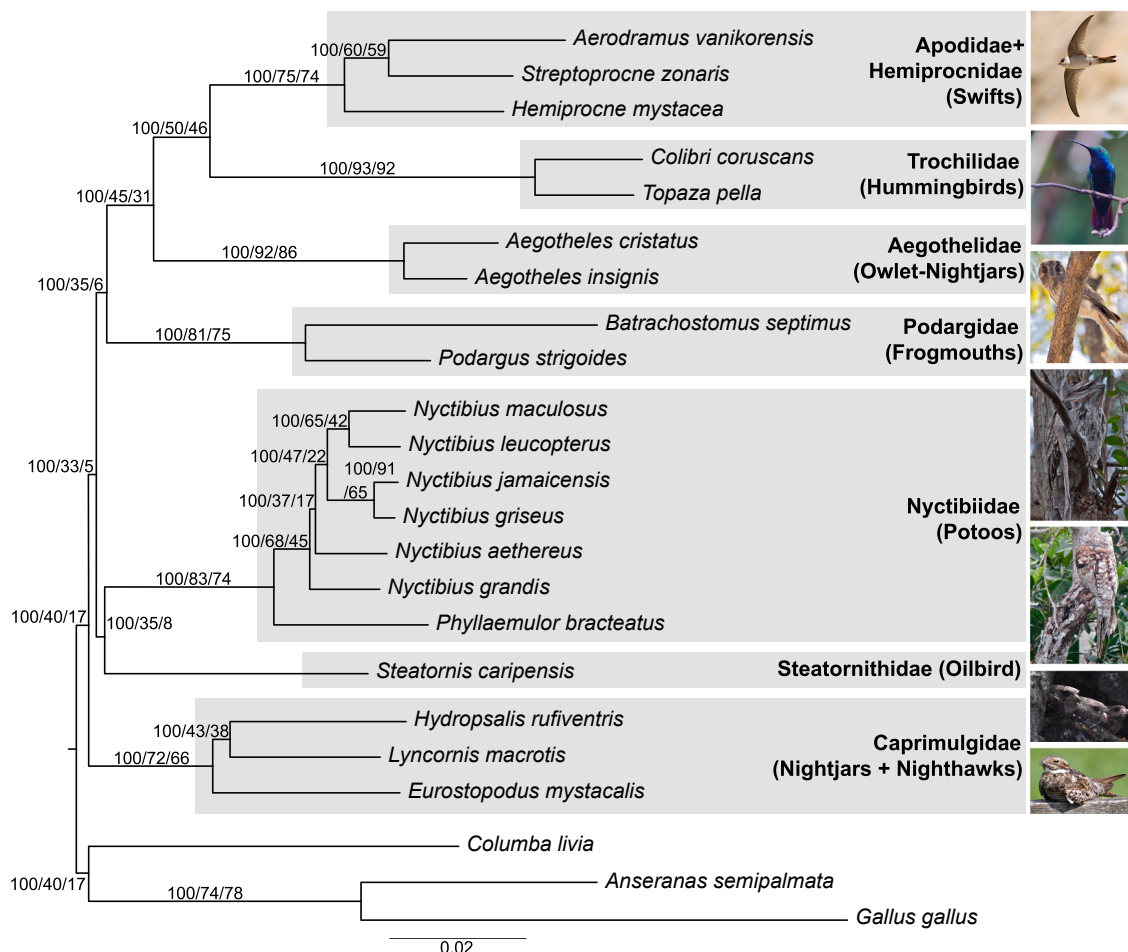


Fig. 6. Best estimate of the Strisores phylogeny. Tree based on ML analysis of the nucleotide D4IU matrix. Support values are bootstraps/site concordance factor/gene concordance factor. Scale units are substitutions per site. Photos taken by Daniel J. Field, University of Cambridge.

feature of the tree may suggest that there is little information in the UCE data that would allow one to prefer one topology over another. However, because the UCE dataset samples the genome broadly 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. While either (or both) could be true, our dataset represents an unprecedented amount of data collected to resolve the deepest Strisores divergences. We designed the analyses in this paper to bring to light hidden biases and sources of conflict in that data, and present the dominant signal recovered across analyses, topology A, as the best estimate of the Strisores phylogeny. Of note, topology A was that found by all prior molecular studies for which the major lineages of Strisores are resolved (Fig. 2).

While we have uncovered substantial evidence for topology A, we recognize conflicting signal present in our analyses in the form of results favoring topology B, and to a far lesser extent, topology C. While we were not able to pinpoint the signal resulting in these topologies in these few instances, we consider their random presence as evidence that our results are not consistently biased by the inclusion or exclusion of any particular taxon, nor by any methodology used in this study. An appropriate future test of the effect of GC content on phylogenetic estimation would be to conduct computational simulations of data at varying levels of GC content and deeply investigate this potential source of bias in genome-scale datasets.

Despite the focus on GC content in this study, it is of note to remember that UCEs are technically more AT-rich than GC-rich. 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. 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 and 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 thereby obscure phylogenetic analysis through incomplete lineage sorting (Kubatko and Degnan, 2007; Degnan and 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 seen with these data, and is likely present in our dataset to some extent. Our ratios of concordance factor support for each topology appear to fit expectations of ILS—that is we have one predominant topology, A, that appears at higher frequency than two minority topologies, B and C, which appear at roughly equal frequency (Table 2). Implementation of the coalescent model to account for ILS yielded a majority of topology A trees, as well as topologies not found in any of our ML analyses (Supplementary Fig. S2). We have no evidence that ILS is an overwhelming signal or major biasing factor in these analyses, and the results presented demonstrate that our analyses do respond to changes in accounting for GC content.

Similarly, we 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 our datasets; in fact they are among the shorter terminals in the Strisores tree (Fig. 6). 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 between oilbird and potoos (Fig. 4) in biasing our 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. We used the best methods available to choose models of sequence evolution that fit our data, whether concatenated or

partitioned analyses, and these methods 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. Additional analyses under the FreeRates model yielded modest increases in support and no change in topology, indicating that among site rate variation was modelled adequately by our previous analyses of the concatenated matrices. Of note, a recent comparison of the FreeRate model with gamma-distributed models indicates that the FreeRate model may behave poorly in some parts of parameter space (see Tamashiro et al., 2019). Additional analyses under the Heterotachy model resulted in a shift to topology A in 4 out of 5 tests, indicating that the lineages of this tree may be experiencing different rates of evolution. We note that the five datasets tested were all among the smaller datasets in length (Fig. 5) and four of them included *Phyllaemulor bracteatus*, a taxon for which we recovered fewer and shorter UCE loci during sequence capture. Thus, while *Phyllaemulor bracteatus* is certainly a potoo (Costa et al., 2017), and there is no doubt about its position in the tree (White et al., 2017), this may be a case in which additional taxon sampling was detrimental to phylogenetic reconstruction (Poe and Swofford, 1999).

Our results indicate that there are complex evolutionary forces acting on the large-scale DNA sequence matrices that are becoming commonplace in phylogenetics. 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). Our results indicate that conventional models of molecular evolution can be insufficient to encapsulate all phenomena present in genome-scale matrices, and this problem may be at the root of some current issues in phylogenomic analysis.

Through the numerous analyses presented here, we were able to uncover different sources of signal in our data. Had we simply made one matrix with this data, and conducted one phylogenetic analysis, we would have produced a tree with apparently high statistical support and accepted that result as a rigorous basis for our 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 and Nakhleh, 2015). Only by using multiple analytical approaches were we able to detect the conflict present in our dataset.

4.3. Evolutionary history of nocturnality in Strisores

With a best estimate of phylogeny in hand, we can hypothesize scenarios by which nocturnality arose in the Strisores (Fig. 7). In debating these potential scenarios, one can take into consideration adaptations to nightvision 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. 7b). Caprimulgids (nightjars and nighthawks) have a tapetum lucidum—a mirror-like structure behind the retina that reflects light back through photoreceptor cells, providing a second chance for the absorption of incoming photons (Nicol and Arnott, 1974). Tapeta lucida are present in many nocturnal vertebrates, and their presence can be observed as a brilliant, or mirror-like, 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 potentially independent evolutionary origins. 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 our current best estimate of the tree suggests an independent origin could be possible (Fig. 6). Frogmouths and owl-nightjars do not have a brilliant eye shine, and so are not believed to have tapeta, but their visual adaptations have not been studied further. A different adaptation to nocturnality is

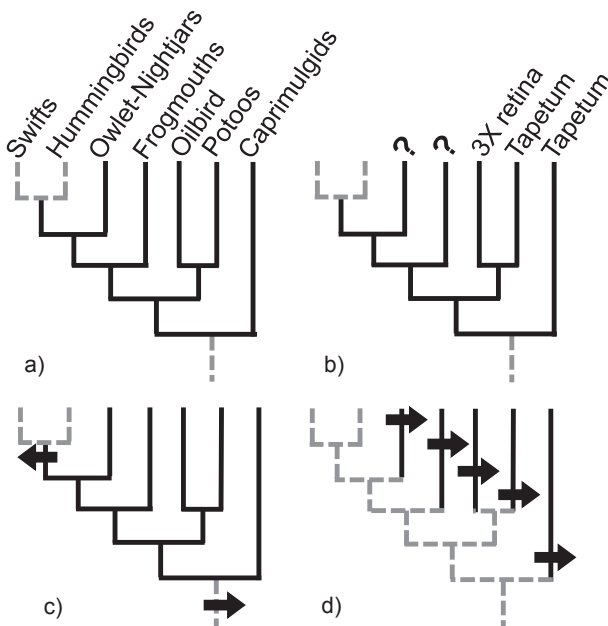


Fig. 7. Evolutionary scenarios for transition/s to nocturnality in Strisores. (a) Best estimate of the Strisores backbone phylogeny based on UCE data. (b) Visual adaptations to nocturnality are known in the oilbird, potoos and caprimulgids, but have not been studied in the frogmouths or owlet-nightjars. (c, d) The topology of the Strisores phylogeny can be used to hypothesize two alternatives of the evolutionary origin of nocturnality representing the extreme scenarios. 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 gray, dashed lines. A similar figure has appeared in Mayr (2010).

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 is compatible with multiple scenarios for the origination of nocturnality in this group—too many to depict here. Instead, for the sake of discussion we illustrate two extreme scenarios; (1) that nocturnality evolved only once in the ancestor of the Strisores, with a single reversal to diurnality in the swifts and hummingbirds (Fig. 7c), or (2) that the Strisores originated from a diurnal ancestor, and there were multiple independent adaptations to nocturnality within this clade (Fig. 7d). Simple parsimony principles would prefer the single origin scenario, as was suggested by Mayr, 2010. However, this only holds if the transitions from diurnality to nocturnality and nocturnality to diurnality have similar probabilities of occurring (Wagner parsimony). If the transition probabilities are substantially different, other scenarios are plausible. In the extreme case where the transition from diurnality to nocturnality is much more common than the reverse (Dollo parsimony), the scenario in Fig. 7d would be preferred. The known variety of adaptations to nightvision among Strisores described above suggests that multiple transitions to nocturnality within the Strisores are plausible, with distinct adaptations to low-light vision in each case (also discussed in Braun and Huddlestone 2009; Mayr, 2010). Further research into the morphological and/or molecular adaptations to nightvision in all Strisorean lineages will help clarify the evolutionary history of nocturnality in the clade. This information may facilitate the application of Bayesian models that allow characters to exhibit both Wagner and Dollo parsimony characteristics through time (e.g., Alekseyenko et al. 2008). Such models have the potential to distinguish between phylogenetic hypotheses like those considered here.

Perhaps of greater evolutionary importance, having both a topology and the knowledge of adaptations to nightvision in hand, as we begin to infer the deeper evolutionary origin of nocturnality in Strisores, we can shed light on the evolutionary history of nocturnality in all 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. Studying the vision of these birds in more detail, such as by looking at changes in the phototransduction cascade, will help to unravel the evolutionary history of nocturnality in these birds, all birds, and to shed light on its origin/s across life.

Data Availability

The raw Illumina fastq sequence reads generated by this study are available in the NCBI Short Read Archive under BioProject ID PRJNA563602. In the Mendeley Data archive associated with this publication, concatenated data matrices are provided with character sets which will allow the isolation of individual UCE loci for any taxon. The treefiles for all analyses presented in Fig. 5 are also available in the Mendeley Data archive.

Acknowledgments

We thank Edward L. Braun for thorough discussion of theory, methodological issues and drafts of this paper. We thank Derrick J. Zwickl, Marek L. Borowiec and David L. Swofford for discussion of methodological issues. We thank an anonymous reviewer for their suggestions and additional insight. We thank Daniel J. Field for use of his photographs. We thank Karen L. Carleton for her continued support and guidance. We acknowledge the genetic repositories listed in Supplementary Table S1, and especially thank the field collectors and 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 et al., 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.

Declaration of Competing Interest

The authors have no competing interests to declare.

Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ympev.2019.106611>.

References

- Akaike, H., 1973. Information theory and extension of the maximum likelihood principle. *Akademiai Kiado, Budapest*.
- Alekseyenko, A.V., Lee, C.J., Suchard, M.A., 2008. Wagner and Dollo: a stochastic duet by composing two parsimonious solos. *Syst. Biol.* 57, 772–784.
- 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. *Mol. Phylogenet. Evol.* 41, 238–248.
- Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004a. Ultraconserved elements in the human genome. *Science* 304, 1321–1325.
- Bejerano, G., Haussler, D., Blanchette, M., 2004b. Into the heart of darkness: large-scale clustering of human non-coding DNA. *Bioinformatics* 20, 140–148.
- Bergsten, J., 2005. A review of long-branch attraction. *Cladistics* 21, 163–193.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for illumina sequence data. *Bioinformatics* 30, 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. *Mol. Phylogenet. Evol.* 56, 536–542.
- Borowiec, M.L., Lee, E.K., Chiu, J.C., Plachetzki, D.C., 2015. Extracting phylogenetic

- signal and accounting for bias in whole-genome data sets supports the Ctenophora as sister to remaining Metazoa. *BMC Genomics* 16, 1–15.
- 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. *Mol. Phylogenet. Evol.* 111, 149–157.
- 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. *Curr. Biol.* 27, 1019–1025.
- Braun, E.L., Kimball, R.T., 2002. Examining basal avian divergences with mitochondrial sequences: model complexity, taxon sampling, and sequence length. *Syst. Biol.* 51, 614–625.
- Braun, M.J., Huddleston, C.J., 2001. Molecular phylogenetics of caprimulgiform night-birds. Paper presented at: Meeting of the American Ornithologist's Union. Abstracts of the 119th Stated Meeting of the American Ornithologists' Union; Seattle, WA.
- Braun, M.J., Huddleston, C., 2009. A molecular phylogenetic survey of caprimulgiform nightbirds illustrates the utility of non-coding sequences. *Mol. Phylogenet. Evol.* 53, 948–960.
- Brumfield, R.T., Swofford, D.L., Braun, M.J., 1997. Evolutionary relationships among the potoos (Nyctibiidae) based on isozymes. *Ornithol. Monogr.* 48, 129–145.
- Bull, J.J., Huelsenbeck, J.P., Cunningham, C.W., Swofford, D.L., Waddell, P.J., 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42, 384–397.
- 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 Res.* 15, 1601–11072.
- Chantler, P., 2017. Swifts (Apodidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://www.hbw.com/node/52266/>.
- Chernomor, O., von Haeseler, A., Minh, B.Q., 2016. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst. Biol.* 65, 997–1008.
- Chifman, J., Kubatko, L., 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30, 3317–3324.
- Chifman, J., Kubatko, L., 2015. Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *J. Theor. Biol.* 374, 35–47.
- Chojnowski, J.L., Kimball, R.T., Braun, E.L., 2008. Introns outperform exons in analyses of basal avian phylogeny using clathrin heavy chain genes. *Gene* 410, 89–96.
- Cleere, N., 1998. Nightjars: a guide to the nightjars, nighthawks, and their relatives, 1st ed. Yale University Press, New Haven, CT.
- Cleere, N., 2010. Nightjars, potoos, frogmouths, oilbird and owl-nightjars of the world, 1st ed. Princeton University Press, Princeton, NJ.
- Cleere, N., 2017. Nightjars (Caprimulgidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://www.hbw.com/node/52265/>.
- Costa, T.V.V., Whitney, B.M., Braun, M.J., White, N.D., Silveria, L.F., Cleere, N., 2017. A Systematic reappraisal of the rufous potoo *Nyctibius bracteatus* (Nyctibiidae) and description of a new genus. *J. Ornithol.* 159, 367–377.
- Cracraft, J., 1981. Toward a phylogenetic classification of the recent birds of the world (Class Aves). *Auk* 98, 681–714.
- Cracraft, J., 1988. The major clades of birds. In: Benton, M.J. (Ed.), *The phylogeny and classification of the tetrapods, volume 1: amphibians, reptiles, birds*. Clarendon Press, Oxford, pp. 339–361.
- Cracraft, J., Barker, F.K., Braun, M., Harshman, J., Dyke, G.J., Feinstein, J., Stanley, S., Cibois, A., Schikler, P., Beresford, P., García-Moreno, J., Sorenson, M.D., Yuri, T., Mindell, M.P., 2004. Phylogenetic relationships among modern birds (Neornithes): toward an avian tree of life. In: Cracraft, J., Donoghue, M. (Eds.), *Assembling the tree of life*. Oxford University Press, New York, pp. 468–489.
- 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 archosaurs. *Biol. Lett.* 8, 783–786.
- de Queiroz, A., 1993. For consensus (sometimes). *Syst. Biol.* 42, 368–372.
- Degnan, J.H., Rosenberg, N.A., 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol. Evol.* 24, 332–340.
- Delsuc, F., Brinkmann, H., Philippe, H., 2005. Phylogenomics and the reconstruction of the tree of life. *Nat. Rev. Genet.* 6, 361–375.
- Dickel, D.E., Ypsilanti, A.R., Pla, R., Zhu, Y., Barozzi, I., Mannion, B.J., Khin, Y.S., Fukuda-Yuzawa, Y., Plajzer-Frick, I., Pickle, C.S., Lee, E.A., Harrington, A.N., Pham, Q.T., Garvin, T.H., Kato, M., Osterwalder, M., Akiyama, J.A., Afzal, V., Rubenstein, J.L.R., Pennacchio, L.A., Vise, A., 2017. Ultraconserved Enhancers Are Required for Normal Development. *Cell* 172, 491–499.
- Ericson, P.G.P., Anderson, C.L., Britton, T., Elzanowski, A., Johansson, U.S., Kallersjö, M., Ohlson, J.I., Parsons, T.J., Zuccon, D., Mayr, G., 2006. Diversification of neaves: integration of molecular sequence data and fossils. *Biol. Lett.* 2, 543–547.
- Eyre-Walker, A., 1993. Recombination and mammalian genome evolution. *P. Roy. Soc. B-Biol. Sci.* 252, 237–243.
- Faircloth, B.C., McCormack, J.E., Crawford, N.G., Harvey, M.G., Brumfield, R.T., Glenn, T.C., 2012. Ultraconserved elements anchor thousands of genetic markers spanning multiple evolutionary timescales. *Syst. Biol.* 61, 717–726.
- Faircloth, B.C., 2013. Illumiprocessor: a trimmomatic wrapper for parallel adapter and quality trimming. Available from: <https://doi.org/10.6079/J9ILL>.
- Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G., 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. *Mol. Ecol. Resour.* 15, 489–501.
- Faircloth, B.C., 2015. PHYLUCE is a software package for the analysis of conserved genomic loci. *Bioinformatics* 32, 786–788.
- Felsenstein, J., 1987. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27, 401–410.
- Figuat, E., Ballenghien, M., Romiguier, J., Galtier, N., 2015. Biased gene conversion and GC-content evolution in the coding sequences of reptiles and vertebrates. *Genome Biol. Evol.* 7, 240–250.
- Foster, P.G., Hickey, D.A., 1999. Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *J. Mol. Evol.* 48, 284–290.
- Galtier, N., Piganeau, G., Mouchiroud, D., Duret, L., 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* 159, 907–911.
- Galtier, N., Duret, L., 2007. Adaptation or biased gene conversion? extending the null hypothesis of molecular evolution. *Trends Genet.* 23, 273–277.
- Glenn, T.C., Faircloth, B.C., 2016. Capturing Darwin's dream. *Mol. Ecol. Resour.* 16, 1051–1058.
- Graybeal, A., 1998. Is it better to add taxa or characters to a difficult phylogenetic problem? *Syst. Biol.* 47, 9–17.
- Griffiths, C.S., Barrowclough, G.F., Groth, J.G., Mertz, L., 2004. Phylogeny of the Falconidae (Aves): a comparison of the efficacy of morphological, mitochondrial, and nuclear data. *Mol. Phylogenet. Evol.* 32, 101–109.
- Hackett, S.J., Kimball, R.T., Reddy, S., Bowie, R.C.K., Braun, E.L., Braun, M.J., Chojnowski, J.L., Cox, W.A., Han, K.L., 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, 1763–1768.
- Hahn, M.W., Nakhleh, L., 2015. Irrational exuberance for resolved species trees. *Evolution* 70, 7–17.
- Han, K.-L., Robbins, M.B., Braun, M.J., 2010. A multi-gene estimate of phylogeny in the nightjars and nighthawks (Caprimulgidae). *Molecul. Phylogenet. Evolut.* 55, 443–453.
- Harmston, N., Barešić, A., Lenhard, B., 2013. The mystery of extreme non-coding conservation. *Phil. Trans. R. Soc. B.* 368, 20130021.
- Harris, R.S., 2007. Improved pairwise alignment of genomic DNA [dissertation]. [University Park (PA)]: The Pennsylvania State University.
- 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. *Syst. Biol.* 52, 386–402.
- 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. *Syst. Biol.* 65, 910–924.
- Hendy, M.D., Penny, D., 1989. A framework for the quantitative study of evolutionary trees. *Syst. Zool.* 38, 297–309.
- Higgins, D., Lemey, P., 2009. Multiple sequence alignment. In: Lemey, P., Salemi, M., Vandamme, A.M. (Eds.), *The phylogenetic handbook: a practical approach to DNA and protein phylogeny*, 1st ed. Cambridge University Press, New York, pp. 68–96.
- Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A.M., Delany, M.E., Dodgson, J.B., Chinwalla, A.T., Clifton, P.F., Clifton, S.W., Delehaunty, K.D., Fronick, C., Fulton, R.S., Graves, T.A., Kremitzki, C., Layman, D., Magrini, V., McPherson, J.D., Miner, T.L., Minx, P., Nash, W.E., Nhan, M.N., Nelson, J.O., Oddy, L.G., Pohl, C.S., Randall-Maher, J., Smith, S.M., Wallis, J.W., Yang, S.-P., Romanov, M.N., Rondelli, C.M., Paton, B., Smith, J., Morrice, D., Daniels, L., Tempest, H.G., Robertson, L., Masabanda, J.S., Griffin, D.K., Vignal, A., Fillon, V., Jacobsson, L., Kerje, S., Andersson, L., Croijmans, R.P.M., Aerts, J., van der Poel, J.J., Ellegren, H., Caldwell, R.B., Hubbard, S.J., Grafham, D.V., Kierzek, A.M., McLaren, S.R., Overton, I.M., Arakawa, H., Beattie, K.J., Bezzubov, Y., Boardman, P.E., Bonfield, J.K., Croning, M.D.R., Davies, R.M., Francis, M.D., Humphray, S.J., Scott, C.E., Taylor, R.G., Tickle, C., Brown, W.R.A., Rogers, J., Buerstedde, J.-M., Wilson, S.A., Stubbs, L., Ovcharenko, I., Gordon, L., Lucas, S., Miller, M.M., Inoko, H., Shiina, T., Kaufman, J., Salomonsen, J., Skjoeft, K., Ka-Shu, Wong, G., Wang, J., Liu, B., Yu, J., Yang, H., Nefedov, M., Koriabine, M., deJong, P.J., Goodstadt, L., Webber, C., Dickens, N.J., Letunic, I., Suyama, M., Torrents, D., von Mering, C., Zdobnov, E.M., Makova, K., Nekrutenko, A., Elitski, L., Eswara, P., King, D.C., Yang, S., Tyekucheva, S., Radakrishnan, A., Harris, R.S., Chiaromonte, F., Taylor, J., He, J., Rijnkels, M., Griffiths-Jones, S., Ureta-Vidal, A., Hoffman, M.M., Severin, J., Searle, S.M.J., Law, A.S., Speed, D., Waddington, D., Cheng, Z., Tuzun, E., Eichler, E., Bao, Z., Flicke, P., Shteynberg, D.D., Brent, M.R., Bye, J.M., Huckle, E.J., Chatterji, S., Dewey, C., Pachter, L., Kouranov, A., Mourelatos, Z., Hatzigeorgiou, A.G., Paterson, A.H., Ivarie, R., Brandstrom, M., Axelsson, E., Backstrom, N., Berlin, S., Webster, M.T., Pourquie, O., Reymond, A., Ucla, C., Antonarakis, S.E., Long, M., Emerson, J.J., Betrán, E., Dupanloup, I., Kaessmann, H., Hinrichs, A.S., Bejerano, G., Furey, T.S., Harte, R.A., Raney, B., Siepel, A., James, Kent W., Haussler, D., Eyraes, E., Castelo, R., Abril, J.F., Castellano, S., Camara, F., Parra, G., Guigo, R., Bourque, G., Tesler, G., Pevzner, P.A., Smit, A., Fulton, L.A., Mardis, E.R., Wilson, R.K., 2004. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716.
- Hoff, K.M., 1966. A comparative study of the appendicular muscles of Strigiformes and Caprimulgiformes [dissertation]. Washington State University, [Seattle (WA)].
- Holyoak, D.T., 2017a. Frogmouths (Podargidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://www.hbw.com/node/52263/>.
- Holyoak, D.T., 2017b. Owllet-Nightjars (Aegothelidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://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). *Mol. Biol. Evol.* 33, 1110–1125.

- Hurvich, C.M., Tsai, C.L., 1989. Regression and time series model selection in small samples. *Biometrika* 76, 297–307.
- Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? *Trends Genet.* 22, 225–231.
- Johnson, K.P., 2001. Taxon sampling and the phylogenetic position of Passeriformes: evidence from 916 avian cytochrome b sequences. *Syst. Biol.* 50, 128–136.
- 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 Res.* 12, 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., Halanach, K.M., 2016. Phylogenomics of Lophotrochozoa with consideration of systematic error. *Syst. Biol.* 66, 256–282.
- Kubatko, L.S., Degnan, J.H., 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Syst. Biol.* 56, 17–24.
- Kumar, S., Filipski, A.J., Battistuzzi, F.U., Kosakovsky Pond, S.L., Tamura, K., 2012. Statistics and truth in phylogenomics. *Mol. Biol. Evol.* 29, 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. *Front. Zool.* 7, 10.
- Lander, E.S., Linton, L.M., Birren, B., Nusbaum, C., Zody, M.C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., LeHoczy, J., Levine, R., McEwan, P., McKernan, K., Meldrum, J., Mesirov, J.P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd, C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J.C., Mungall, A., Plumb, R., Ross, M., Showlken, R., Sims, S., Waterston, R.H., Wilson, R.K., Hillier, L.W., McPherson, J.D., Marra, M.A., Mardis, E.R., Fulton, L.A., Chinwalla, A.T., Pepin, K.H., Gish, W.R., Chissoe, S.L., Wendt, M.C., Delehaunty, K.D., Miner, T.L., Delehaunty, A., Kramer, J.B., Cook, L.L., Fulton, R.S., Johnson, D.L., Minx, P.J., Clifton, S.W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J.-F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R.A., Muzny, D.M., Scherer, S.E., Bouck, J.B., Sodergren, E.J., Worley, K.C., Rives, C.M., Gorrell, J.H., Metzker, M.L., Naylor, S.L., Kucherlapati, R.S., Nelson, D.L., Weinstock, G.M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brodtier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Smith, D.R., Doucette-Stamm, L., Rubinfeld, M., Weinstock, K., Mei, L., Dubois, J., Yang, H., Yu, J., Wang, J., Huang, G., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S., Davis, R.W., Federspiel, N.A., Pia, Abola, A., Proctor, M.J., Roe, B.A., Chen, F., Pan, H., Ramser, J., Lehrach, H., Reinhardt, R., Richard, McCombie W., de la Bastide, M., Dedhia, N., Böcker, H., Hornisch, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J.A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D.G., Burge, C.B., Cerutti, L., Chen, H.-C., Church, D., Clamp, M., Copley, R.R., Doerks, T., Eddy, S.R., Eichler, E.E., Furey, T.S., Galagan, J., Gilbert, J.G.R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W., Steven, Johnson, L., Jones, T.A., Kasif, S., Kasprzyk, A., Kennedy, S., James, Kent W., Kitts, P., Koonin, E.V., Korf, I., Kulp, D., Lancet, D., Lowe, T.M., McLysaght, A., Mikkelsen, T., Moran, J.V., Mulder, N., Pollara, V.J., Ponting, C.P., Schuler, G., Schultz, J., Slater, G., Smit, A.F.A., Stupka, E., Zsuzskowski, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y.I., Wolfe, K.H., Yang, S.-P., Yeh, R.-F., Collins, F., Guyer, M.S., Peterson, J., Felsenfeld, A., Wetterstrand, K.A., Myers, R.M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D.R., Olson, M.V., Kaul, R., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G.A., Athanasiou, M., Schultz, R., Patrinos, A., Morgan, M.J., 2001. Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with bowtie 2. *Nat. Methods* 9, 357–359.
- Lockton, S., Gaut, B.S., 2005. Plant conserved non-coding sequences and paralogous evolution. *Trends Genet.* 21, 60–65.
- Lopez, P., Casane, D., Philippe, H., 2002. Heterotachy, an important process of protein evolution. *Mol. Biol. Evol.* 19, 1–7.
- 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. *Re* 16, 1204–1223.
- Livezey, B.C., Zusi, R.L., 2006. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. I. Methods and characters. *Bull. Carnegie Mus. Nat. Hist.* 37, 1–544.
- Livezey, B.C., Zusi, R.L., 2007. Higher-order phylogeny of modern birds (Theropoda, Aves: Neornithes) based on comparative anatomy. II. Analysis and discussion. *Zool. J. Linn. Soc.-Lond.* 149, 1–95.
- 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. *Syst. Biol.* 61, 90–106.
- Liu, L., Yu, L., Kubatko, L., Pearl, D.K., Edwards, S.V., 2009. Coalescent methods for estimating phylogenetic trees. *Mol. Phylogenet. Evol.* 53, 320–328.
- Mariaux, J., Braun, M.J., 1996. A molecular phylogenetic survey of the nightjars and allies (Caprimulgiformes) with special emphasis on the potoos (Nyctibiidae). *Mol. Phylogenet. Evol.* 6, 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). *J. Für Ornithol.* 143, 82–97.
- Mayr, G., Manegold, A., Johansson, U.S., 2003. Monophyletic groups within “higher land birds”—comparison of morphological and molecular data. *J. Zool. Syst. Evol. Res.* 41, 233–248.
- Mayr, G., 2004. Old world fossil record of modern-type hummingbirds. *Science* 304, 861–864.
- Mayr, G., 2005a. The paleogene fossil record of birds in Europe. *Biol. Rev.* 80, 515–542.
- Mayr, G., 2005b. The paleogene old world potoo *Paraprefica* Mayr, 1999 (Aves, Nyctibiidae): its osteology and affinities to the new world pterocninae Olson, 1987. *J. Syst. Paleo.* 3, 359–370.
- Mayr, G., 2008. Avian higher-level phylogeny: well-supported clades and what we can learn from a phylogenetic analysis of 2954 morphological characters. *J. Zool. Syst. Evol. Res.* 46, 63–72.
- Mayr, G., 2009. Paleogene fossil birds, 1st ed. Springer-Verlag, Berlin, Germany.
- Mayr, G., 2010. Phylogenetic relationships of the paraphyletic “caprimulgiform” birds (nightjars and allies). *J. Zool. Syst. Evol. Res.* 48, 126–137.
- Mary, G., 2017. Avian evolution: the fossil record of birds and its paleobiological significance, 1st ed. John Wiley & Sons, West Sussex, United Kingdom.
- 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 Res.* 22, 746–754.
- McCormack, J.E., Harvey, M.G., Faircloth, B.C., Crawford, N.G., Glenn, T.C., Brumfield, R.T., 2013. A phylogeny of birds based on over 1,500 loci collected by target enrichment and high-throughput sequencing. *PLoS ONE* 8, e54848.
- Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Kimball, R.T., Braun, E.L., 2016. Analysis of a rapid evolutionary radiation using ultraconserved elements: evidence for a bias in some multispecies coalescent methods. *Syst. Biol.* 65, 612–627.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. Creating the CIPRES science gateway for inference of large phylogenetic trees. Paper presented at: Gateway Computing Environments Workshops. In Proceedings of the Gateway Computing Environments Workshop GCE; New Orleans, LA.
- Minh, B.Q., Nguyen, M.A.T., von Haeseler, A., 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* 30, 1188–1195.
- Minh, B.Q., Hahn, M., Lanfear, R., 2018. New methods to calculate concordance factors for phylogenomic datasets. <https://doi.org/10.1101/487801>.
- Misof, B., Misof, K., 2009. A Monte Carlo approach successfully identifies randomness in multiple sequence alignments: a more objective means of data exclusion. *Syst. Biol.* 58, 21–34.
- Nabholz, B., Künstner, A., Wang, R., Jarvis, E.D., Ellegren, H., 2011. Dynamic evolution of base composition: causes and consequences in avian phylogenomics. *Mol. Biol. Evol.* 28, 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. *Dev. Biol.* 327, 526–540.
- Nesbitt, S.J., Ksepka, D.T., Clarke, J.A., 2011. Podargiform affinities of the enigmatic *Fluviviridavis platyrhamphus* and the early diversification of Strisores (Caprimulgiformes) + Apodiformes. *PLoS ONE* 6, e26350.
- Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., Minh, B.Q., 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274.
- Nicol, J., Arnott, H., 1974. Tapeta lucida in the eyes of goatsuckers (Caprimulgidae). *P. Roy. Soc. B-Biol. Sci.* 187, 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.
- Omland, K.E., Lanyon, S.M., Fritz, S.J., 1999. A molecular phylogeny of the new world orioles (Icteridae): the importance of dense taxon sampling. *Mol. Phylogenet. Evol.* 12, 224–239.
- Pennacchio, L.A., Ahituv, N., Moses, A.N., Prabhakar, S., Nobrega, M.A., Shoukry, M., Minovitsky, 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, 499–502.
- Pennisi, E., 2017. Mysterious unchanging DNA finds a purpose in life. *Science* 356, 892.
- 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 Biol.* 9, e1000602.
- Philippe, H., Snell, E.A., Baptiste, E., Lopez, P., Holland, P.W.H., Casane, D., 2004. Phylogenomics of eukaryotes: impact of missing data on large alignments. *Mol. Biol. Evol.* 21, 1740–1752.
- Phillips, M.J., Delsuc, F., Penny, D., 2004. Genome-scale phylogeny and the detection of systematic biases. *Mol. Biol. Evol.* 21, 1455–1458.
- Pirnie, S.P., Osman, A., Zhu, Y., Carmichael, G.G., 2016. An ultraconserved element (UCE) controls homeostatic splicing of *ARGL1* mRNA. *Nucleic Acids Res.* 45, 3473–3486.
- Poe, S., Swofford, D.L., 1999. Taxon sampling revisited. *Nature* 398, 299–300.
- Pollock, D.D., Zwickl, D.J., McGuire, J.A., Hillis, D.M., 2002. Increased taxon sampling is advantageous for phylogenetic inference. *Syst. Biol.* 51, 664–671.
- Prychitko, T.M., Moore, W.S., 1997. The utility of DNA sequences of an intron from the β -fibrinogen gene in phylogenetic analysis of woodpeckers (Aves: Picidae). *Mol. Phylogenet. Evol.* 8, 193–204.
- Prum, R.O., Berv, J.S., Dornburg, A., Field, D.J., Townsend, J.P., Lemmon, E.M., Lemmon, A.R., 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* 526, 569–573.
- Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing

- genomic features. *Bioinformatics* 26, 841–842.
- Reddy, S., Kimball, R.T., Pandey, A., Hosner, P.A., Braun, M.J., Hackett, S.J., Han, K.-L., Harsham, 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. *Syst. Biol.* 66, 857–879.
- 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 albigollis*). *Brain Behav. Evol.* 64, 19–33.
- 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 Res.* 20, 1001–1009.
- 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. *Mol. Biol. Evol.* 30, 2134–2144.
- 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. *Mol. Biol. Evol.* 33, 670–678.
- Rosel, P.E., Block, B.A., 1996. Mitochondrial control region variability and global population structure in the swordfish, *Xiphias gladius*. *Mar. Biol.* 125, 11–22.
- Rosenbloom, K.R., Armstrong, J., Barber, G.P., Casper, J., Clawson, H., Diekhans, M., Dreszer, T.R., Fujita, P.A., Guruvadoo, L., Haussler, 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 Res.* 43, D670–D681.
- Roure, B., Baurain, D., Philippe, H., 2012. Impact of missing data on phylogenies inferred from empirical phylogenomic data sets. *Mol. Biol. Evol.* 30, 197–214.
- Salichos, L., Rokas, A., 2014. Inferring ancient divergences requires genes with strong phylogenetic signals. *Nature* 497, 27–33.
- Salinas, N.R., Little, D.P., 2014. 2matrix: a utility for indel coding and phylogenetic matrix concatenation. *Appl. Plant Sci.* 2, 1300083.
- Schuchmann, K.L., Bonan, A., 2017. Hummingbirds (Trochilidae). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://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. *Syst. Biol.* 58, 381–394.
- Shubin, N., Tabin, C., Carroll, S., 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.
- 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 Res.* 15, 1034–1050.
- Simpson, K., Day, N., Trusler, P., 2010. *Birds of Australia*, 8th ed. Penguin Group, Camberwell.
- Soubrier, J., Steel, M., Lee, M.S.Y., Der Sarkissian, C., Guindon, S., Ho, S.Y.W., Cooper, A., 2012. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol. Biol. Evol.* 29, 3345–3358.
- 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. *Syst. Biol.* 48, 65–75.
- Stamatakis, A., 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313.
- Starrett, J., Derkarabetian, S., Hedin, M., Bryson Jr., R.W., McCormack, J.E., Faircloth, B.C., 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. *Mol. Ecol. Res.* 17, 812–823.
- 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. *Mol. Biol. Evol.* 25, 402–408.
- Streicher, J.W., Schulte, J.A., Weins, J.J., 2015. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Syst. Biol.* 65, 128–145.
- Suh, A., 2016. The phylogenomic forest of bird trees contains a hard polytomy at the root of Neoaves. *Zool. Scripta* 45, 50–62.
- Sukumaran, J., Holder, M.T., 2010. DendroPy: a python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571.
- Sun, K., Meiklejohn, K.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., Kimball, R.T., 2014. The evolution of peafowl and other taxa with ocelli (eyespot): a phylogenomic approach. *Proc. Roy. Soc. B* 281, 20140823.
- Swofford, D.L., 2003. PAUP*. Phylogenetic analysis using parsimony (*and other methods), 4 ed. Sinauer Associates, Sunderland, MA.
- Tamashiro, R.A., White, N.D., Braun, M.J., Faircloth, B.C., Braun, E.L., Kimball, R.T., 2019. What are the roles of taxon sampling and model fit in tests of cyto-nuclear discordance using avian mitogenomic data? *Mol. Phylogenet. Evol.* 130, 132–142.
- 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. *Syst. Biol.* 64, 778–791.
- Thomas, B.T., 2017. Oilbird (*Steatornithidae*). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://www.hbw.com/node/52261/>.
- Walls, G.L., 1963. *The vertebrate eye and its adaptive radiation*, secost ed. Hafner Publishing Company, New York (NY).
- Warnefors, M., Hartmann, B., Thomsen, S., Alonso, C.R., 2016. Combinatorial gene regulatory functions underlie ultraconserved elements in *Drosophila*. *Mol. Biol. Evol.* 33, 2294–2306.
- Wells, D., 2017. Treeswifts (*Hemiprocridae*). In: del Hoyo, J., Elliott, A., Sargatal, J., Christie, D.A., de Juana, E. (Eds.). *Handbook of the birds of the world alive*. Barcelona: Lynx Edicions; (accessed May 22, 2017). <http://www.hbw.com/node/52267/>.
- 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). *Ornitol. Neotrop.* 27, 223–236.
- White, N.D., Mitter, C., Braun, M.J., 2017. Ultraconserved elements resolve the phylogeny of potoos (Aves: Nyctibiidae). *J. Avian Biol.* 48, 872–880.
- White, N.D., 2017a. Locus Simulator: v1.0. Perl script version 1.0. Available from: URL <http://doi.org/10.5281/zenodo.832352>.
- White, N.D., 2017b. GCvariance: Another Release. Perl script version 1.1. Available from: URL <http://doi.org/10.5281/zenodo.582627>.
- White, N.D., 2017c. Indel-Stats: v1.0. Perl script version 1.0. Available from: URL <http://doi.org/10.5281/zenodo.438047>.
- White, N.D., 2017d. PAUP_parsers: v1.0. Perl script version 1.0. Available from: URL <http://doi.org/10.5281/zenodo.832357>.
- 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 Biol.* 3, e7.
- Yang, Z., 1995. A space-time process model for the evolution of DNA sequences. *Genetics* 139, 993–1005.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829.
- Zhang, Chao, Rabiee, Maryam, Sayyari, Erfan, Mirarab, Siavash, 2018. ASTRAL-III: Polynomial Time Species Tree Reconstruction from Partially Resolved Gene Trees. *BMC Bioinfo.* 19 (S6), 153.
- Zwickl, D.J., Hillis, D.M., 2002. Increased taxon sampling greatly reduces phylogenetic error. *Syst. Biol.* 51, 588–598.
- Zwickl, D.J., 2006. Genetic algorithm approaches for the phylogenetic analysis of large biological sequence datasets under the maximum likelihood criterion [dissertation]. University of Texas at Austin, [Austin (TX)].
- 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. *Syst. Biol.* 63, 645–659.