Genomic signatures of the evolution of a diurnal lifestyle in Strigiformes

Pamela Espíndola-Hernández (ID) ,* Jakob C. Mueller (ID) , Bart Kempenaers (ID)

Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany

*Corresponding author: Department of Behavioural Ecology and Evolutionary Genetics, Max Planck Institute for Ornithology, 82319 Seewiesen, Germany. Email: pamela.dola@gmail.com

Abstract

Understanding the targets of selection associated with changes in behavioral traits represents an important challenge of current evolutionary research. Owls (Strigiformes) are a diverse group of birds, most of which are considered nocturnal raptors. However, a few owl species independently adopted a diurnal lifestyle in their recent evolutionary history. We searched for signals of accelerated rates of evolution associated with a diurnal lifestyle using a genome-wide comparative approach. We estimated substitution rates in coding and noncoding conserved regions of the genome of seven owl species, including three diurnal species. Substitution rates of the noncoding elements were more accelerated than those of protein-coding genes. We identified new, owl-specific conserved noncoding elements as candidates of parallel evolution during the emergence of diurnality in owls. Our results shed light on the molecular basis of adaptation to a new niche and highlight the importance of regulatory elements for evolutionary changes in behavior. These elements were often involved in the neuronal development of the brain.

Keywords: parallel evolution; comparative genomics; diel-activity pattern; diurnality; adaptation; CNEEs; protein-coding genes

Introduction

Even though owls are considered one of the most iconic nocturnal birds, species vary considerably in their diel activity patterns. The spectrum of phenotypes ranges from exclusively nocturnal owls (family Tytonidae) to diurnal ones (the snowy owl *Bubo scandiacus*, the northern hawk owl *Surnia ulula* and the burrowing owl *Athene cunicularia*), with many intermediate activity patterns (e.g. crepuscular or cathemeral) (del Hoyo *et al.* 1999; König and Weick 2008; Duncan 2018). Diurnality in owls is absent in the family Tytonidae, but has emerged independently at least twice among the family Strigidae (König and Weick 2008; Wink *et al.* 2009; Salter *et al.* 2020). This provides an opportunity to study genomic signatures of a recent case of parallel evolution in birds.

The owls belong to the clade of the Afroaves and presumably evolved from an ancestral diurnal landbird with raptorial features (Ericson *et al.* 2006; Hackett *et al.* 2008; Jarvis *et al.* 2014; Prum *et al.* 2015; McClure *et al.* 2019). Currently, 250 species of owls live in a variety of ecosystems around the world (del Hoyo *et al.* 1999; König and Weick 2008). Their diversification from the rest of the Afroaves was probably fostered by increasing opportunities to hunt small nocturnal mammals, which experienced a rapid radiation during the Eocene (56–33 Ma) (Feduccia 1995, 1999, 2003). Many of the owls' early adaptations to nocturnality have been shaped by positive selection on genes functionally associated with visual perception, including phototransduction and chromatin packaging (Espíndola-Hernández *et al.* 2020). However, little is known about the mechanisms and targets of selection

that shaped the more recent shift into a diurnal activity pattern observed in some owls.

The diurnal owls, as well as their cathemeral relatives, have been described as "time-shifter" species. Despite phylogenetic constraints on the evolution of diel activity patterns (Roll *et al.* 2006; Anderson and Wiens 2017), the "time-shifter" species might have changed their activity pattern in response to competition for food (Schoener 1974; Jaksić 1982; Carothers and Jaksić 1984). Shorter nights during summer and interference competition might have been the main drivers of diurnality in the owls included in this study (Pei *et al.* 2018).

Modern evolutionary biology tries to understand whether genetic correlates of parallel evolution of novel phenotypic traits exist, and of which type these are. The emergence of diurnality in different owl clades is a case of parallel evolution (Gould 2002; Pearce 2012; Rosenblum *et al.* 2014), and has likely occurred from similar genomic elements of the common Strigidae ancestor. A general distinction is often made between regulatory and structural changes, and evidence for both exists. In birds, for instance, loss of flight evolved independently in different clades and has been linked to protein-coding genes (Burga *et al.* 2017; Pan *et al.* 2019), as well as to noncoding elements (Sackton *et al.* 2019).

Nonsynonymous changes in protein-coding regions affect the structure of the gene product and, therefore, the function of the protein itself. Because of the supposed strong phenotypic effect, these structural modifications have been considered as major evolutionary factors. However, nonsynonymous changes are relatively rare, and closely related species are often almost identical

in protein-coding regions of the genome. Thus, King and Wilson (1975) suggested that the phenotypic differences observed between closely related species, such as human and chimpanzee, are likely due to mutations in regulatory regions of the genome. Many studies have now shown that changes in the regulation of gene expression contribute to differences in a multitude of phenotypic traits (Wray 2007; Rubinstein and de Souza 2013; Stern 2013; Hill *et al.* 2021). Conserved nonexonic elements (CNEEs), which have been used as markers for avian phylogenomic inferences (Edwards *et al.* 2017; Tiley *et al.* 2020), are usually located in the *cis*-regulatory domain of genes, and mutations in these regions have been linked to a wide variety of phenotypic changes that often constitute evolutionary innovations (Wray 2007; Rubinstein and de Souza 2013). In birds, CNEEs have been used to study the evolution of the development of avian limbs and flight feathers (Seki *et al.* 2017), convergent evolution associated with the loss of flight in ratites (Sackton *et al.* 2019), and the diversification of bill shape (Yusuf *et al.* 2020).

Here, we report on a search for signals of accelerated evolution linked to the emergence of diurnality. We compared the substitution rates in the genomes of seven owl species, of which four are strictly nocturnal and 3 are consistently diurnal. To obtain the maximum contrast in diel activity patterns, we did not include species with intermediate or cathemeral phenotypes. We used a genome-wide comparative approach to estimate substitution rates in conserved coding regions (CDS: coding sequences) and noncoding regions (CNEE) of the respective genomes. Our study aims to answer the following questions. (1) Are there CDS and CNEEs that evolved under accelerated substitution rates among diurnal owls? (2) Is there an enrichment of functions linked with these CDS and CNEEs, and therefore with a diurnal lifestyle? (3) Are these genomic signatures predominantly structural (CDS) or regulatory (CNEEs)?

Materials and methods
Study species, reference genome, and multispecies alignment

We used the genome assembly and annotation of *A. cunicularia* (Burrowing owl) as reference for the studied species (Mueller *et al.* 2018). The reference genome was annotated by the NCBI Eukaryotic Genome Annotation Pipeline (NCBI *A. cunicularia* Annotation Release 100; NCBI Assembly Accession GCA_003259725.1 of athCun1).

The genomes of *Asio otus* (long-eared owl), *Bubo bubo* (Eurasian eagle owl), *B. scandiacus* (snowy owl), and *S. ulula* (Northern hawk owl) have been sequenced and mapped to the reference for a previous study (Espíndola-Hernández *et al.* 2020). The genome assemblies of *Strix occidentalis* (spotted owl, Hanna *et al.* 2017), *Tyto alba* (barn owl, Ducrest *et al.* 2020), and *Leptosomus discolor* (cuckoo roller, Zhang *et al.* 2014, used as outgroup) were downloaded from NCBI and mapped to the reference genome using LAST v. 921 (Kiełbasa *et al.* 2011). Despite some ambiguity in the phylogeny of owls, the topological relationships among the included owl species is well established and remained the same in studies using different markers (mitochondrial and ultra-conserved genome-wide markers) (Wink *et al.* 2009; Salter *et al.* 2020). We used the consensus topology of these phylogenetic trees with the Cuckoo roller as the outgroup for all analyses (Fig. 1). We used an unrooted tree that is a modified version of the same topology for the acceleration rate tests in coding genes (see Extended Methods section of the Supplementary File 1). The Supplementary Material provides the accession numbers of the downloaded
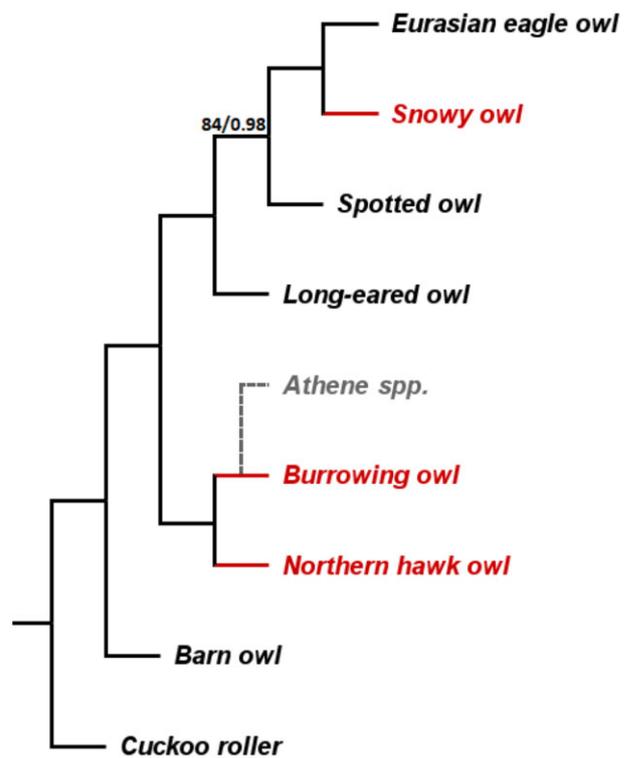
Fig. 1. Phylogenetic topology of the included owl species. A maximum likelihood and Bayesian inference analyses (Salter *et al.* 2020) showed that all nodes received 100% bootstrap support except for the ancestral node of *Strix* and *Bubo* spp., which is labeled with the exact value (ML bootstrap support/Bayesian posterior probability). The tip branches of the three diurnal owl species, where the transition to diurnality occurred, were tested for accelerated evolution (in red). The non-included sister species of the burrowing owl (the rest of the *Athene* spp.) are mostly nocturnal (in gray).

genomes (Supplementary Table 1 in Supplementary File 2), a general workflow diagram of the analyses (Supplementary Fig. 1 in Supplementary File 1), and a more detailed description of the pipelines and parameters.

We produced a single genome-wide, reference-mapped sequence for each species in four steps. (1) Compilation ("piling up") of all the reads or sequences of the whole genome using samtools (Li *et al.* 2009). (2) Variant calling with bcftools (Danecek and McCarthy 2017). (3) Producing the reference-mapped, species-specific sequence with bcftools, choosing the allele with more reads or better mapping quality in case of heterozygous sites. (4) Soft-masking (change to lowercase) of the repetitive regions (based on the reference genome), and hard-masking (change to "N") of sites with zero-read coverage (per species).

To produce multispecies alignments, we first extracted the sequence of each element (gene or CNEE; see below) from the reference-mapped sequence of each species using bedtools (Quinlan and Hall 2010; Dale *et al.* 2011). We then concatenated the extracted sequences of all species in a single, multispecies FASTA file and ran a multispecies aligner for each element, either using MACSE (Ranwez *et al.* 2011) for the genes, or PRANK (Löytynoja 2014) for the CNEEs. We used MACSE for protein-coding gene sequences because it corrects for potentially erroneous frameshifts (e.g. indels smaller than triplets) without disrupting the underlying codon structure. Finally, we removed high-entropy regions and gaps with BMGE v. 1.12 (Criscuolo and Gribaldo 2010).

Avian-specific CNEEs and identification of owl-specific CNEEs

We used the 284,001 avian-specific CNEEs identified and described by Sackton *et al.* (2019), which are conserved among 35 species across the avian clade, are at least 50 bp long, and include a large fraction of known regulatory elements. The positions of these avian-specific CNEEs are publicly available in the coordinates of the Chicken 4.0 assembly (Sackton *et al.* 2019). We used MafFilter (Dutheil *et al.* 2014) to transfer ("liftover") the avian-specific CNEEs from the Chicken 4.0 coordinates to the Burrowing owl (athCun1) coordinates.

Additionally, we identified new owl-specific CNEEs that are shared among nocturnal owls. First, we used PhyloFit (from the software package PHAST: PHylogenetic Analysis with Space-Time models, Hubisz *et al.* 2011) to estimate a neutral model based on 4-fold degenerate sites (4d sites) of all coding regions of the four nocturnal-owl genomes (Eurasian eagle owl, long-eared owl, spotted owl, and barn owl). We used msa_view (from PHAST) to extract these 4d sites (Hubisz *et al.* 2011). Then, we used this neutral model (also referred to as the nonconserved model) as null model for the identification of the "most-conserved" regions in the noncoding regions of the four nocturnal-owl genomes with PhastCons (Siepel *et al.* 2005). We excluded all CNEE sequences <50 bp.

Test for accelerated substitution rates in coding sequences

We followed the method used in Espíndola-Hernández *et al.* (2020) to test for accelerated rates of evolution in CDS. In brief, we estimated the nonsynonymous to synonymous substitution rate ratio (ω = dN/dS; for a review, see Nielsen 2005) to measure the direction and magnitude of selection on protein-coding genes. A value of $\omega < 1$ indicates purifying selection, $\omega = 1$ neutral evolution, and $\omega > 1$ positive selection. We used the maximum-likelihood method implemented in the CodeML program of PAML 4.9h (Yang 2007), based on the branch model (Yang 1998) and the branch-site model (Yang and Nielsen 2002; Zhang *et al.* 2005; Yang and dos Reis 2011). For both models, we used a preset unrooted tree topology with the branches of the diurnal owls labeled as the foreground (see Supplementary File 1), and the rest of the tree branches as background.

We consider as an "accelerated substitution rate" each case where the alternative model had a significant better fit to the data and had a $\omega_{foreground} > \omega_{background}$, which mostly indicates positive selection at specific sites, but might also include cases of relaxed purifying selection (Espíndola-Hernández *et al.* 2020). We applied the false discovery rate (FDR) correction to control for multiple testing for the CodeML models. To complement the selection test results based on CodeML, we used the aBSREL model (Smith *et al.* 2015), implemented in the HyPhy package, to test for selection signals that are specific for the diurnal owls. This test implements a modified version of the branch-site model to test for selection exclusively in the foreground branches. In this test, we included all the genome-wide significant protein-coding genes according to the CodeML tests. To control for multiple testing, we used the Holm–Bonferroni sequential rejection procedure from the HyPhy package (Smith *et al.* 2015).

Test for accelerated substitution rates in CNEEs

We tested different evolutionary models to identify an accelerated substitution rate in the foreground branches leading to the three diurnal owls using the Bayesian approach implemented in PhyloAcc (Hu *et al.* 2019). PhyloAcc uses a hierarchical Bayesian phylogenetic model to identify branches on a phylogeny on which particular genomic elements change their substitution rate, from a conserved or neutral to an accelerated substitution rate (Hu *et al.* 2019). The conservation or acceleration is estimated in relation to a neutral model. The neutral model was first built using PhyloFit (Hubisz *et al.* 2011) based on the 4d sites of all coding regions from the complete set of eight bird genomes used in this study (similar to the model used for the detection of owl-specific CNEEs except for the set of species). PhyloAcc considers that the elements have initially evolved at a neutral rate ($r_0 = 1$, having the same substitution rate as the initial neutral model), and then become conserved at the root or some other branch on the phylogeny ($r_1 < 1$, having a lower substitution rate than the neutral model). The elements might then evolve with an accelerated rate ($r_2 > r_1$, having a higher substitution rate than the conserved state) (Hu *et al.* 2019). The program PhyloAcc restricts the possible shift patterns in 3 nested models: the null model (M_0), where the substitution rate in any branch is not allowed to shift to an accelerated rate; the lineage-specific model (M_1), where the substitution rates of the diurnal owls are allowed to shift to an accelerated rate; and the full model (M_2), where the substitution rate of any branch is allowed to shift. The marginal likelihood of the data under each model is compared by two Bayes factors, BF1 and BF2 (Hu *et al.* 2019). Briefly, BF1 is the ratio of the marginal likelihoods of the data under M_1 and M_0, indicating how much the data support M_1 in relation to M_0. BF2 is the ratio of the marginal likelihoods of the data under M_1 and M_2, indicating how much the data support M_1 in relation to M_2. To identify DNA elements accelerated exclusively in target lineages, Hu *et al.* (2019) recommend considering only cases with high values in both Bayes factors. Thus, we considered a CNEE as a candidate for parallel accelerated evolution during the emergence of diurnality in owls when the following conditions were met: $\log_{BF1} \geq 10$, $\log_{BF2} \geq 1$, and the posterior probability for accelerated evolution under $M_2 > 0.8$ for at least two of the three diurnal owl species. The distribution of both Bayes factors across all tested CNEEs is shown in Supplementary Fig. 2 in Supplementary File 1.

Functional overrepresentation analysis

Within each group of elements (genes, avian-specific and owl-specific CNEEs), we ranked the elements according to the strength of evidence for accelerated evolution in the diurnal owls, and applied a Wilcoxon rank-sum test with the R package GOfuncR (Grote 2018). We used a custom-made gene ontology (GO) annotation database made for all annotated athCun1 genes, combining human (org.Hs. e.g. db) and chicken (org.Gg. e.g. db) annotations to GOs. We ranked the genes by the test statistic (log-likelihood ratio value) of the branch or branch-site test, and included only the genes with an accelerated substitution rate in the diurnal owls ($\omega_{background} < \omega_{foreground}$) in the case of the branch model. The CNEEs were ranked by a custom-made parameter based on the posterior probability of acceleration (pp) along the phylogenetic tree. For each branch, we estimated the probability of acceleration relative to that of the respective ancestral branch ($pp_{branch} - pp_{ancestral\ branch}$). Then, we summed these relative probabilities for all diurnal owls and for the other branches from the nocturnal species. The custom-made parameter for the CNEEs is then the difference between the sum from the diurnal species and the sum from the nocturnal species (see the formula in Supplementary File 1).

For the ranked genes, the gene-GO annotation file was used directly, while for the ranked CNEEs, we produced a CNEE-GO

annotation file using the GOs of nearby genes. We linked CNEEs to genes by intersecting the CNEEs with the putative "Gen Regulatory Domain Region" of "One Closest" genes established by GREAT (McLean *et al.* 2010). The "One Closest" option of GREAT determines for each gene a potential regulatory domain that extends maximally 1 Mb from the Transcription Start Site (TSS) in both directions until the mid-point between this TSS and the TSS of the adjacent gene (McLean *et al.* 2010). To account for multiple testing and for potential clustering of CNEEs around genes, we used the family-wise error rate (FWER) estimation procedure of GOfuncR, which permutes the ranking parameter while the annotations of CNEEs or genes to GO categories stay fixed and re-estimates the statistics for every GO term (Grote 2018).

Comparison of evolutionary rates between protein-coding genes and CNEEs

The comparison between rates of evolution in coding and non-coding regions of the genome is not straightforward. The codon structure of genes adds another level of complexity in evolutionary models, due to the different constraints of substitutions for each of the nucleotide positions in a codon. The sites in the noncoding CNEEs apparently do not show systematic patterns of evolutionary constraints. However, depending on the definition of CNEEs, they likely also include neutrally evolving and more or less conserved sites. Thus, we compared the acceleration rates of evolution of CDS and CNEEs by a simple substitution model without considering the codon structure, using PhyloP (Pollard *et al.* 2010). The scale estimates indicate the rate of evolution relative to the neutral model (same model as for the PhyloAcc analysis described above). To this end, we ran the likelihood ratio test of PhyloP with the lineage-specific option to compare a null model having one single scale parameter with an alternative model having two estimated scale parameters: one scale for the branches leading to the diurnal owls (foreground scale) and a second scale for all remaining branches (background scale). We compared the distributions of the estimated subscale (ratio between foreground and background scale in the alternative model) between coding genes and CNEEs. We used ggplot2 to visualize these distributions (Wickham 2016). To account for variation in sequence length of the tested elements (a potential confounder for scale estimates), we plotted subscale for different length intervals. This allows the comparison of subscale values between protein-coding genes and CNEEs of similar length.

Results
Accelerated evolutionary rates in coding genes and CNEEs and their functional enrichment

Of the 12,298 tested protein-coding genes, 69 showed a significantly higher ω-value in the diurnal owls compared to the background of nocturnal species (branch model: FDR $\leq$ 0.05, Supplementary Table 2 in Supplementary File 2), and 15 showed accelerated substitution rates on specific sites of the diurnal owl sequences (branch-site model: FDR $\leq$ 0.05, Supplementary Table 2 in Supplementary File 2). Seven of these genes showed evidence for positive selection at specific sites in at least two diurnal owls and not in any other species (*IKZF2, SOX18, JPH2, WNT4, CAMK1D, GIT2,* and *CASP8*), according to the aBSREL model (Supplementary Table 7 in Supplementary File 2).

Based on the branch-model tests, we found no evidence for functional enrichment among the high-ranked genes. For the branch-site model tests, high-ranked genes were significantly enriched for the GO term "HAUS complex" (Wilcoxon rank-sum test,

FWER = 0.001, Supplementary Table 3 in Supplementary File 2). According to the complementary aBSREL model, the functions of the significantly accelerated 7 genes are predominantly related to regulatory functions, including transcription regulation (Supplementary Table 7 in Supplementary File 2).

Among the 265,599 tested avian-specific CNEEs (Sackton *et al.* 2019), 113 elements showed significantly accelerated rates of evolution in diurnal owls based on the Bayes factor thresholds, 13 of these were accelerated in 2 diurnal owl species, and only two showed evidence for accelerated evolution in all three diurnal owl species according to the threshold of the posterior probability of acceleration in the full model (Fig. 2 and Supplementary Table 4 in Supplementary File 2). The high-ranked avian-specific CNEEs were significantly enriched for elements linked to one GO term associated to the axolemma, the plasma membrane of the neurons' axon (Wilcoxon rank-sum test, FWER < 0.05, Supplementary Table 5 in Supplementary File 2). There are 629 avian-CNEEs in the putative regulatory domain of 12 genes (*ADORA1, ADORA2A, ANK1, CNTNAP2, EPB41L3, KCNC2, KCNJ11, MAPT, MYO1D, ROBO2, SPTBN1,* and *THY1*) related to this GO term (Supplementary Table 5 in Supplementary File 2). We manually annotated these genes using public gene databases and found that most of them are functionally linked to neuronal development and connectivity (Supplementary Table 6 in Supplementary File 2).

We identified 2,364 new owl-specific CNEEs present among all four nocturnal owl species. From these, 31 showed evidence for accelerated evolution in at least one of the diurnal owl species based on the Bayes factor thresholds. Only three of them had a posterior probability of accelerated evolution above the threshold in the full model in two diurnal owl species (Fig. 2), and none in all three diurnal species. Twenty-eight showed evidence for accelerated evolution in the snowy owl only. There was no genome-wide significant functional enrichment of GO terms among the ranked owl-specific CNEEs.

The genome-wide detected genes (CDS) and the genes associated with the CNEEs with evidence for accelerated evolution in at least two diurnal owl species do not have elements in common. This is true for both, owl-specific CNEEs and avian-specific CNEEs.

## Comparison of acceleration rates between genes and CNEEs

According to the LRT from PhyloP and after correction for multiple testing, 2.3% of the genes (278 out of 12,298 genes tested), 2.8% of the owl-specific CNEEs (67 out of 2,364 CNEEs tested), and 0.1% of the avian-specific CNEEs (329 out of 265,599 CNEEs tested) showed evidence for accelerated evolution in the diurnal owls (PhyloP results with subscale > 1 and FDR corrected P-value < 0.05, i.e. genome-wide significance).

Among the genome-wide significantly accelerated elements, the sub-scale values of CNEEs were generally higher than those of the protein-coding genes (Figs. 3 and 4). The CNEEs also showed outlier groups of extreme values. To account for the fact that CNEEs are on average shorter than protein-coding genes, we compared subscale values within intervals of sequence lengths (Fig. 4). Most of the elements with extreme subscale values had also extreme sequence lengths. Those cases were excluded from Fig. 4, which only shows the length intervals for which data from all three categories of elements were available (see legend).
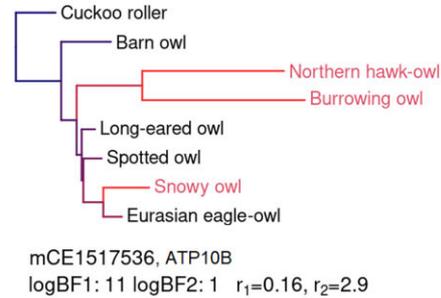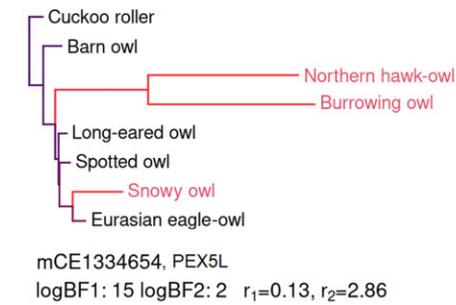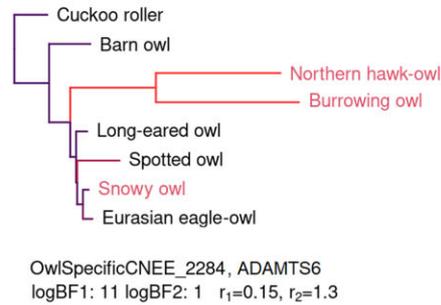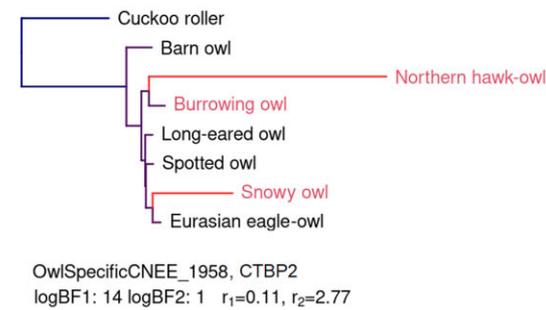
Fig. 2. Avian- and owl-specific CNEEs with evidence of accelerated rates of evolution in diurnal owls. The phylogenetic tree illustrates the shift in substitution rates under the full model [according to Hu *et al.* (2019) and Sackton *et al.* (2019)]. Diurnal species are indicated in red. The branch lengths are proportional to the posterior mean substitution rate. The line below each tree shows the name of the CNEE and its associated gene, the values of the two log-BFs, and the conserved (r_1) and accelerated substitution rate (r_2). The sequence data support a parallel shift from a conserved CNEE indicating purifying selection (blue) to an accelerated substitution rate (red) in two or three diurnal owl species.
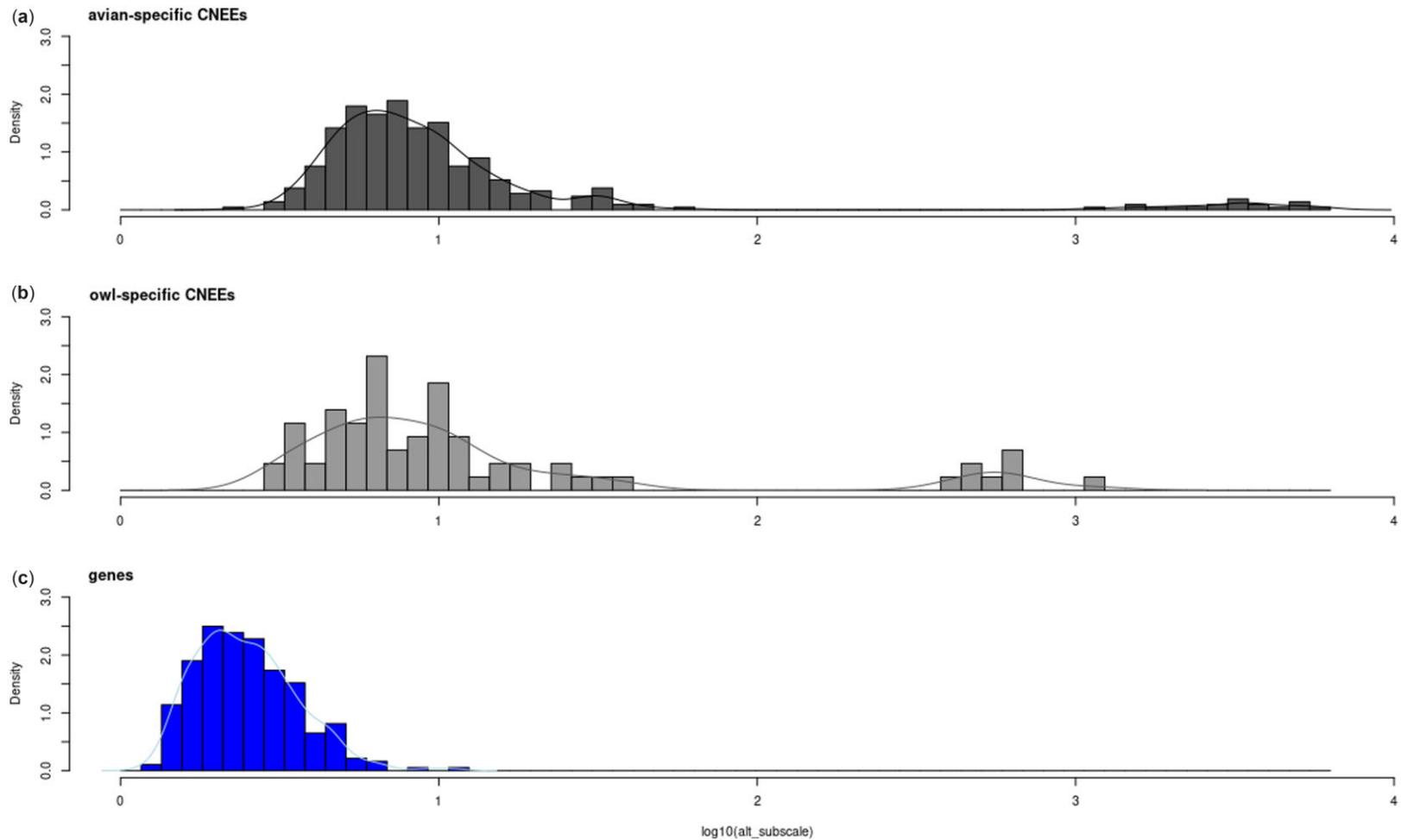
Fig. 3. Distributions of subscale values in the alternative model (alt_subscale) from significantly accelerated elements (PhyloP results with subscale > 1 and FDR-corrected P-value < 0.05). The plot compares the histograms and density curves of $\log_{10}$-transformed subscale values of avian-specific CNEEs (a, black, N = 329), owl-specific CNEEs (b, gray, N = 67), and genes (c, blue, N = 286). Most of the elements with values on the tail ends of these distributions have also extreme sequence lengths (genes longer than 1,000 bp, or CNEEs shorter than 200 bp; see Fig. 4).
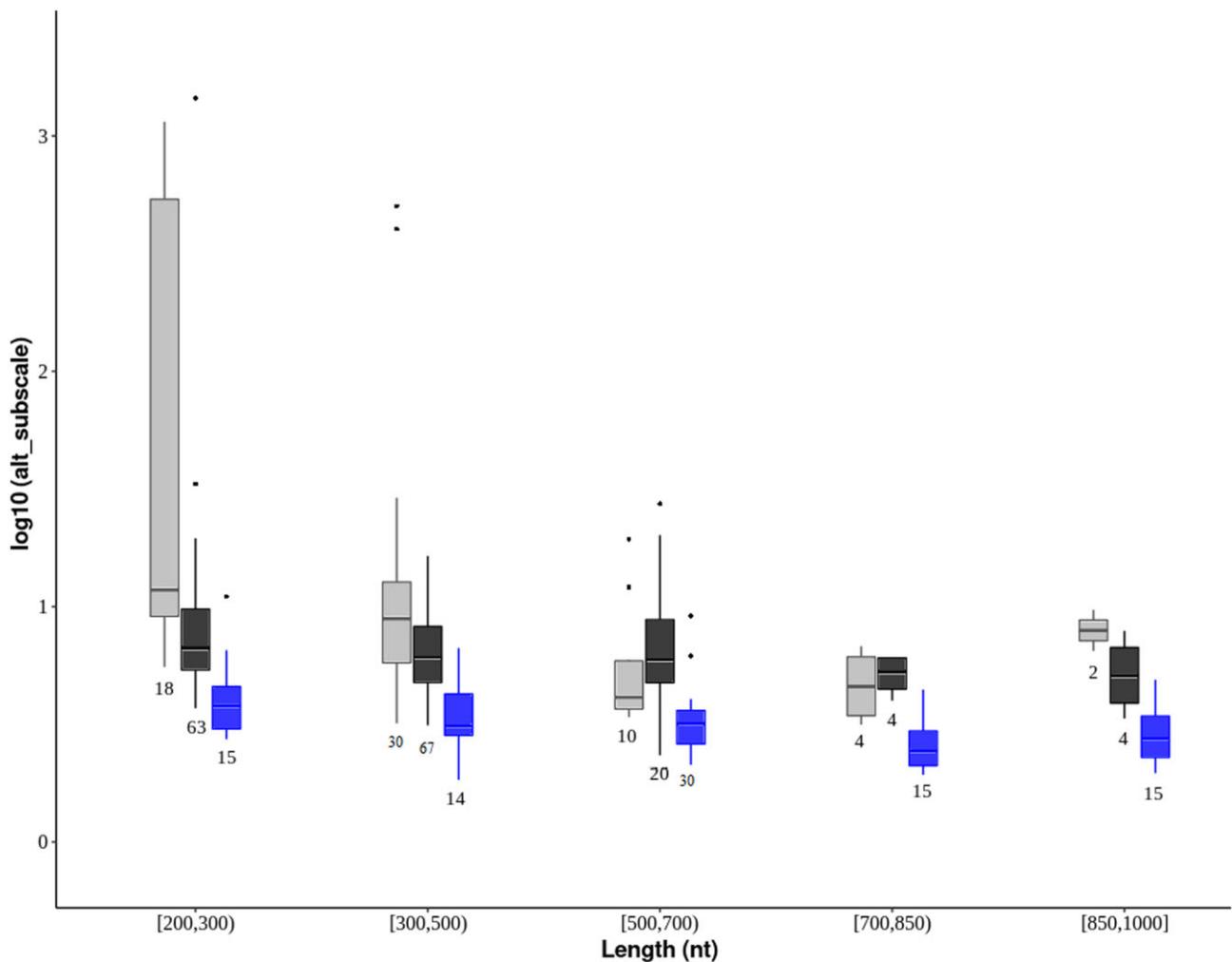
Fig. 4. Comparison of subscale values between owl-specific CNEEs (gray), avian-specific CNEEs (black), and genes (blue) in relation to sequence length (intervals). Only elements with genome-wide significance are included (PhyloP results with subscale >1 and FDR-corrected *P*-value < 0.05). Shown are box plots with sample sizes (number of elements). We excluded genes longer than 1,000 bp, and avian-specific CNEEs shorter than 200 bp, because there were no owl-specific CNEEs outside of the range from 200 to 1,000 bp.

Discussion

Our study aimed at detecting genomic signals of selection linked to the evolution of a diurnal lifestyle in owls, whereby we searched for accelerated substitution rates in protein-coding and noncoding elements of the genome. Our results showed that accelerated substitution rates during the evolution of diurnality in owls occurred in both coding and noncoding regions of the genome. The absolute number of significantly accelerated elements was comparable between protein-coding genes and CNEEs. However, among those elements with evidence for accelerated evolution, the magnitude of acceleration (subscale value) was larger in the CNEEs than in the protein-coding genes. Our comparison between these genomic regions is based on a general substitution model without considering the codon structure of the protein-coding genes or the expected variable evolutionary constraints among the sites in CNEEs. Hence, our approach can only serve as a rough average comparison across all elements. Further, most of the protein-coding genes with signals of positive selection at specific sites exclusively in the diurnal owls are associated with regulatory processes of gene expression. This functional association with regulatory processes and the higher

magnitude of acceleration in potentially *cis*-regulatory elements (CNEEs) suggest that regulatory evolution might have been more relevant than structural evolution during the shift to a diurnal lifestyle in owls.

Structural and regulatory changes as mechanisms for adaptation have long been discussed. Several papers have reviewed evidence about which part of the genome plays a more relevant role in adaptative evolution (Macintyre 1982; Carroll 2005; Wray 2007; Hoekstra and Coyne 2007; Romero *et al.* 2012; Hill *et al.* 2021). Different levels of pleiotropy are important for evolutionary hypotheses about why genetic substitutions might occur more frequently in regulatory noncoding regions than in structural protein-coding genes. Many of the CNEEs are in *cis*-regulatory genomic regions with modular organization, such that they regulate the expression of only one nearby gene and are affected by only a single transcription factor. This implies that a mutation in one of the many regulatory modules might selectively affect only one aspect of the gene-expression network, e.g. only in a specific tissue (Wray 2007; Molodtsova *et al.* 2014). Another important aspect regarding the evolution of noncoding regions is their functional redundancy. In some cases, regulatory elements share functions and this redundancy acts as a buffer against genetic

disturbances, allowing genetic changes without compromising essential biological functions. The buffering conferred by the cis-regulatory redundancy might mediate the recruitment of novel regulatory binding sites from existing ones and eventually the achievement of novel gene regulation pathways (Hong et al. 2008; Frankel et al. 2010; Perry et al. 2010; Wittkopp and Kalay 2012; Rubinstein and de Souza 2013). Many behavioral traits are inherently dynamic and this might need "fine tuning" by regulatory responses to a dynamic environment (Macintyre 1982; Wray 2007). Diel activity is such a behavioral trait that might require a dynamic and flexible control, and therefore is expected to predominantly evolve through regulatory mutations affecting specific gene regulatory network interactions.

The significance of regulatory evolution has been highlighted in other recent studies in birds. For instance, Seki et al. (2017) found that birds have a higher proportion of conserved elements in the non-coding part of the genome in comparison to mammals. The authors showed that these avian-specific, highly conserved elements in the noncoding region are associated with genes that participate in the development of avian limbs and flight feathers. Their results support the hypothesis that changes in noncoding regulatory sequences might have played an important role in the emergence of avian evolutionary innovations (Seki et al. 2017). Additional support for this hypothesis came from a comparative study among palaeognathous species (Sackton et al. 2019). This study showed that noncoding elements with accelerated rates of evolution were overrepresented near key limb developmental genes. They further proved the cis-regulatory activity of the CNEEs through their effect of open chromatin states during embryonic development. Thus, the study suggested that convergent morphological evolution and loss of flight in ratites were more strongly associated with changes in the regulatory noncoding part of the genome than in protein-coding genes (Sackton et al. 2019). In another study, Yusuf et al. (2020) identified candidate loci related to macro-evolutionary shifts in bird beak shape evolution across distantly related avian taxa, and studied whether those morphological shifts were explained by shifts in molecular rates of coding and noncoding genomic regions. The study found that signals in the noncoding regions were more often associated with avian bill shape diversification.

Each identified signal of selection or accelerated evolution provides a candidate element, either coding or noncoding, for further study of parallel evolution of diurnality in owls. We attempted to interpret and summarize these signals using functional enrichment analyses of GO terms. Among the protein-coding elements, the high-ranked genes showed a significant association with the GO term "HAUS complex" (HAUS1, HAUS2, HAUS3, HAUS6, and HAUS8; Supplementary Table 3 in Supplementary File 2). This GO term refers to a microtubule-binding complex involved in the generation of the mitotic spindle (Goshima et al. 2008), that also plays a key role in neuronal migration, polarization, and development through local regulation of the cytoskeleton in axons and dendrites (Cunha-Ferreira et al. 2018). Due to its effects on the development of neuronal connectivity in the brain, it might play a role in the evolution of behavior (Mueller et al. 2020), and consequently in adaptations to a diurnal lifestyle in owls.

Among the avian-CNEEs, only PEX5L and ATP10B showed accelerated substitution rates in all three diurnal species. These two genes are involved in the organization and maintenance of organelles in the cytoplasm, especially in the brain cells. In particular, they function in the cortical neurons (PEX5L in the peroxisomes

and ATP10B in the maintenance of lysosome membrane integrity). Considering the difference between the diurnal and nocturnal species in terms of their probability of shift to acceleration (ranking parameter), there was an overrepresentation of high-ranked avian-CNEEs placed around genes functionally linked to the axolemma GO term. We inspected the functions of the genes in this GO term, using information from GeneCards (www.genecards.org, last accessed: 11.03.2022), NCBI gene (www.ncbi.nlm.nih.gov/search/, last accessed: 11.03.2022), and amiGO2 (http://amigo.geneontology.org/amigo, last accessed: 11.03.2022). Most of these genes are involved in interactions between the intra- and extra-cellular environment through the plasma membrane, especially in the brain cells, and several of these genes were related to the development of neurons and the regulation of membrane potentials in the neurons (ADORA1, ADORA2A, CNTNAP2, EPB41L3, KCNC2, KCNJ11, MAPT, MYO1D, ROBO2, and THY1; Supplementary Table 6 in Supplementary File 2). Four of these genes are related to human phenotypes that involve a variety of abnormalities in the development of eyes and ears (ANK1, MAPT, MYO1D, and SPTBN1; Supplementary Table 6 in Supplementary File 2), and two of these genes are related to regulation of the circadian rhythm and sleep (ADORA1, and ADORA2A; Supplementary Table 6 in Supplementary File 2). These genes therefore seem to be good candidates in the context of adaptation to a diurnal lifestyle in the owls.

In addition to the avian-specific CNEEs (Sackton et al. 2019), we identified 2,364 new owl-specific CNEEs among the nocturnal owls. These owl-specific CNEEs are candidates for regulatory elements during the evolution of owls. Only three of these elements are strong candidates for regulatory changes during the evolution of diurnality in owls, showing accelerated substitution rates in at least two of the three diurnal species. These three owl-specific CNEEs are linked to the genes ADAMTS6, CTBP2, and DIPK2B. ADAMTS6 is generally involved in proteolysis, and kidney and heart development, but also encodes two isoforms that are upregulated by tumor necrosis factor alpha (TNFα) in retinal pigment epithelial cells (Bevitt et al. 2003; Lu et al. 2013). One of the isoforms encoded by CTBP2 (ribeye) is a major component of specialized synapses known as synaptic ribbons. These specialized synapses are involved in visual (Schmitz et al. 2000) and auditory perception (West and McDermott 2011), as well as circadian timing and the pupillary light reflex (Hannibal and Fahrenkrug 2006; Østergaard et al. 2007). Mutations in the human CTBP2 have been linked to retinitis pigmentosa, night blindness, and deafness (GeneCards, www.genecards.org, last accessed: 11.03.2022). DIPK2B (DIA1R) encodes signal peptides for protein targeting in the secretory pathway, and is expressed in embryonic and adult brain tissues.

We found no genes in common between those identified as showing evidence for genome-wide accelerated substitution rates and those associated with the CNEEs that showed such evidence in at least two diurnal owls. This result is in line with another comparative study (Yusuf et al. 2020), which showed 2 different sets of genes associated with signals of accelerated evolution in coding and noncoding regions, even though both were implicated in beak development.

In summary, our results showed that accelerated evolution occurs in coding and noncoding conserved genomic regions during the emergence of diurnality in owls. Acceleration rates were higher in the noncoding elements than in the protein-coding genes, and accelerated protein-coding genes in diurnal owls are functionally associated with regulation of gene expression. Our results suggest that regulatory evolution might have played a

predominant role in the shift to a diurnal lifestyle in owls. In addition, as expected for a shift to a diurnal lifestyle with sensory and behavioral adaptations, several accelerated noncoding and coding elements are functionally linked to nervous system development and brain connectivity.

Data availability

Sequence data are publicly available, and their references and accession numbers are listed in Supplementary Table 1 in Supplementary File 2. The multispecies alignments of all elements are available in repository https://doi.org/10.25387/g3.19369118.

Supplemental material is available at G3 online.

Acknowledgments

Conflicts of interest

The authors declare to have no competing interests.

Literature cited

Anderson SR, Wiens JJ. Out of the dark: 350 million years of conservatism and evolution in diel activity patterns in vertebrates. Evolution. 2017;71(8):1944–1959.

Bevitt DJ, Mohamed J, Catterall JB, Li Z, Arris CE, Hiscott P, Sheridan C, Langton KP, Barker MD, Clarke MP, *et al.* Expression of ADAMTS metalloproteinases in the retinal pigment epithelium derived cell line ARPE-19: transcriptional regulation by TNFα. Biochim Biophys Acta. 2003;1626(1–3):83–91.

Burga A, Wang W, Ben-David E, Wolf PC, Ramey AM, Verdugo C, Lyons K, Parker PG, Kruglyak L. A genetic signature of the evolution of loss of flight in the Galapagos cormorant. Science. 2017;356(6341):eaal3345.

Carothers JH, Jaksic FM. Time as a niche difference: the role of interference competition. Oikos. 1984;42(3):403–406.

Carroll SB. Evolution at two levels: on genes and form. PLoS Biol. 2005;3(7):e245.

Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. BMC Evol Biol. 2010;10:210.

Cunha-Ferreira I, Chazeau A, Buijs RR, Stucchi R, Will L, Pan X, Adolfs Y, van der Meer C, Wolthuis JC, Kahn OI, *et al.* The HAUS complex is a key regulator of non-centrosomal microtubule organization during neuronal development. Cell Rep. 2018;24(4):791–800.

Dale RK, Pedersen BS, Quinlan AR. Pybedtools: a flexible Python library for manipulating genomic datasets and annotations. Bioinformatics. 2011;27(24):3423–3424.

Danecek P, McCarthy SA. BCFtools/csq: haplotype-aware variant consequences. Bioinformatics. 2017;33(13):2037–2039.

del Hoyo J, Elliott A, Sargatal J, Cabot J. Handbook of the Birds of the World. Barcelona: Lynx Edicions; 1999.

Ducrest A-L, Neuenschwander S, Schmid-Siegert E, Pagni M, Train C, Dylus D, Nevers Y, Warwick Vesztrocy A, San-Jose LM, Dupasquier M, *et al.* New genome assembly of the barn owl (*Tyto alba alba*). Ecol Evol. 2020;10(5):2284–2298.

Duncan JR. Owls of the World: Their Lives, Behavior and Survival. Baltimore (MD): Johns Hopkins University Press; 2018.

Dutheil JY, Gaillard S, Stukenbrock EH. MafFilter: a highly flexible and extensible multiple genome alignment files processor. BMC Genomics. 2014;15:53.

Edwards SV, Cloutier A, Baker AJ. Conserved nonexonic elements: a novel class of marker for phylogenomics. Syst Biol. 2017;66(6):1028–1044.

Ericson PGP, Anderson CL, Britton T, Elzanowski A, Johansson US, Källersjö M, Ohlson JI, Parsons TJ, Zuccon D, Mayr G, *et al.* Diversification of Neoaves: integration of molecular sequence data and fossils. Biol Lett. 2006;2(4):543–547.

Espíndola-Hernández P, Mueller JC, Carrete M, Boerno S, Kempenaers B. Genomic evidence for sensorial adaptations to a nocturnal predatory lifestyle in owls. Genome Biol Evol. 2020;12(10):1895–1908.

Feduccia A. Explosive evolution in tertiary birds and mammals. Science. 1995;267(5198):637–638.

Feduccia A. The Origin and Evolution of Birds. New Haven (CT): Yale University Press; 1999.

Feduccia A. 'Big bang' for tertiary birds? Trends Ecol. Evol. 2003;18(4):172–176.

Frankel N, Davis GK, Vargas D, Wang S, Payre F, Stern DL. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. Nature. 2010;466(7305):490–493.

Goshima G, Mayer M, Zhang N, Stuurman N, Vale RD. Augmin: a protein complex required for centrosome-independent microtubule generation within the spindle. J Cell Biol. 2008;181(3):421–429.

Gould SJ. The Structure of Evolutionary Theory. Cambridge (MA): Harvard University Press; 2002.

Grote S. 2018. GOfuncR: gene ontology enrichment using FUNC. R package version 1.2.0.

Hackett SJ, Kimball RT, Reddy S, Bowie RCK, Braun EL, Braun MJ, Chojnowski JL, Cox WA, Han K-L, Harshman J, *et al.* A phylogenomic study of birds reveals their evolutionary history. Science. 2008;320(5884):1763–1768.

Hanna ZR, Henderson JB, Wall JD, Emerling CA, Fuchs J, Runckel C, Mindell DP, Bowie RCK, DeRisi JL, Dumbacher JP, *et al.* Northern spotted owl (*Strix occidentalis caurina*) genome: divergence with the barred owl (*Strix varia*) and characterization of light-associated genes. Genome Biol Evol. 2017;9(10):2522–2545.

Hannibal J, Fahrenkrug J. Neuronal input pathways to the brain's biological clock and their functional significance. Adv Anat Embryol Cell Biol. 2006;182:1–71.

Hill MS, Vande Zande P, Wittkopp PJ. Molecular and evolutionary processes generating variation in gene expression. Nat Rev Genet. 2021;22(4):203–215.

Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007;61(5):995–1016.

Hong JW, Hendrix DA, Levine MS. Shadow enhancers as a source of evolutionary novelty. Science. 2008;321(5894):1314.

Hu Z, Sackton TB, Edwards SV, Liu JS. Bayesian detection of convergent rate changes of conserved noncoding elements on phylogenetic trees. Mol Biol Evol. 2019;36(5):1086–1100.

Hubisz MJ, Pollard KS, Siepel A. Phast and Rphast: phylogenetic analysis with space/time models. Brief Bioinform. 2011;12(1):41–51.

Jaksić FM. Inadequacy of activity time as a niche difference: the case of diurnal and nocturnal raptors. Oecologia. 1982;52(2):171–175.

Jarvis ED, Mirarab S, Aberer AJ, Li B, Houde P, Li C, Ho SYW, Faircloth BC, Nabholz B, Howard JT, *et al.* Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 2014; 346(6215):1320–1331.

Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011;21(3): 487–493.

King M, Wilson a C. Evolution at two levels in humans and chimpanzees. Science. 1975;188(4184):107–116.

König C, Weick F. Owls of the World. London: Christopher Helm. A & C Black; 2008.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25(16):2078–2079.

Löytynoja A. Phylogeny-aware alignment with PRANK. In: D Russell, editor. Multiple Sequence Alignment Methods. Totowa (NJ): Humana Press; 2014. p. 155–170.

Lu Y, Vitart V, Burdon KP, Khor CC, Bykhovskaya Y, Mirshahi A, Hewitt AW, Koehn D, Hysi PG, Ramdas WD, *et al.*; NEIGHBOR Consortium. Genome-wide association analyses identify multiple loci associated with central corneal thickness and keratoconus. Nat Genet. 2013;45(2):155–163.

Macintyre RJ. Regulatory genes and adaptation. In: MK Hecht, B Wallace, and GT Prance, editors. Evolutionary Biology. Boston (MA): Springer; 1982. p. 247–285.

McClure CJW, Schulwitz SE, Anderson DL, Robinson BW, Mojica EK, Therrien J-F, Oleyar MD, Johnson J. Commentary: defining raptors and birds of prey. J. Raptor Res 2019;53(4):419.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol. 2010;28(5):495–501.

Molodtsova D, Harpur BA, Kent CF, Seevananthan K, Zayed A. Pleiotropy constrains the evolution of protein but not regulatory sequences in a transcription regulatory network influencing complex social behaviours. Front Genet 2014;5:431–437.

Mueller JC, Carrete M, Boerno S, Kuhl H, Tella JL, Kempenaers B. Genes acting in synapses and neuron projections are early targets of selection during urban colonization. Mol Ecol. 2020;29(18): 3403–3412.

Mueller JC, Kempenaers B, Kuhl H, Boerno S, Tella JL, Carrete M. Evolution of genomic variation in the burrowing owl in response to recent colonization of urban areas. Proc R Soc B Biol Sci. 2018; 285:1–9.

Nielsen R. Molecular signatures of natural selection. Annu. Rev. Genet. 2005;39:197–218.

Østergaard J, Hannibal J, Fahrenkrug J. Synaptic contact between melanopsin-containing retinal ganglion cells and rod bipolar cells. Invest Ophthalmol Vis Sci. 2007;48(8):3812–3820.

Pan S, Lin Y, Liu Q, Duan J, Lin Z, Wang Y, Wang X, Lam SM, Zou Z, Shui G, *et al.* Convergent genomic signatures of flight loss in birds suggest a switch of main fuel. Nat Commun. 2019;10:1–11.

Pearce T. Convergence and parallelism in evolution: a neo-Gouldian account. Br J Philos Sci. 2012;63(2):429–448.

Pei Y, Valcu M, Kempenaers B. Interference competition pressure predicts the number of avian predators that shifted their timing of activity. Proc R Soc B. 2018;285(1880):20180744.

Perry MW, Boettiger AN, Bothma JP, Levine M. Shadow enhancers foster robustness of Drosophila gastrulation. Curr Biol. 2010; 20(17):1562–1567.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of non-neutral substitution rates on mammalian phylogenies. Genome Res. 2010;20(1):110–121.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. Nature. 2015; 526(7574):569–573.

Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26(6):841–842.

Ranwez V, Harispe S, Delsuc F, Douzery EJP. MACSE: multiple alignment of coding SEquences accounting for frameshifts and stop codons. PLoS One. 2011;6(9):e22594.

Roll U, Dayan T, Kronfeld-Schor N. On the role of phylogeny in determining activity patterns of rodents. Evol Ecol. 2006;20(5): 479–490.

Romero IG, Ruvinsky I, Gilad Y. Comparative studies of gene expression and the evolution of gene regulation. Nat Rev Genet. 2012; 13(7):505–516.

Rosenblum EB, Parent CE, Brandt EE. The molecular basis of phenotypic convergence. Annu Rev Ecol Evol Syst. 2014;45(1): 203–226.

Rubinstein M, de Souza FSJ. Evolution of transcriptional enhancers and animal diversity. Philos Trans R Soc B Biol Sci. 2013;368: 20130017.

Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, *et al.* Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019; 364(6435):74–78.

Salter JF, Oliveros CH, Hosner PA, Manthey JD, Robbins MB, Moyle RG, Brumfield RT, Faircloth BC. Extensive paraphyly in the typical owl family (Strigidae). Auk. 2020;137:1–15.

Schmitz F, Königstorfer A, Südhof TC. RIBEYE, a component of synaptic ribbons: a protein's journey through evolution provides insight into synaptic ribbon function. Neuron. 2000; 28(3):857–872.

Schoener TW. Resource partitioning in ecological communities. Science. 1974;185(4145):27–39.

Seki R, Li C, Fang Q, Hayashi S, Egawa S, Hu J, Xu L, Pan H, Kondo M, Sato T, *et al.* Functional roles of Aves class-specific cis-regulatory elements on macroevolution of bird-specific features. Nat Commun. 2017;8:1–14.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S, *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. Genome Res. 2005;15(8):1034–1050.

Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: An adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol Biol Evol. 2015;32(5):1342–1353.

Stern DL. The genetic causes of convergent evolution. Nat Rev Genet. 2013;14(11):751–764.

Tiley GP, Pandey A, Kimball RT, Braun EL, Burleigh JG. Whole genome phylogeny of Gallus: introgression and data-type effects. Avian Res. 2020;11(1):15.

West MC, McDermott BM. Ribeye a-mCherry fusion protein: A novel tool for labeling synaptic ribbons of the hair cell. J Neurosci Methods. 2011;197(2):274–278.

Wickham H. 2016. ggplot2: Elegant Graphics for Data Analysis. New York: Springer-Verlag. ISBN 978-3-319-24277-4.

Wink M, El-Sayed A-A, Sauer-Gürth H, Gonzalez J. Molecular phylogeny of owls (Strigiformes) inferred from DNA sequences of the

mitochondrial cytochrome *b* and the nuclear *RAG-1* gene. Ardea. 2009;97(4):581–591.

Wittkopp PJ, Kalay G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. Nat Rev Genet. 2012;13(1):59–69.

Wray GA. The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. 2007;8(3):206–216.

Yang Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. Mol Biol Evol. 1998; 15(5):568–573.

Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–1591.

Yang Z, dos Reis M. Statistical properties of the branch-site test of positive selection. Mol Biol Evol. 2011;28(3):1217–1228.

Yang Z, Nielsen R. Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. Mol Biol Evol. 2002;19(6):908–917.

Yusuf L, Heatley MC, Palmer JPG, Barton HJ, Cooney CR, Gossmann TI. Noncoding regions underpin avian bill shape diversification at macroevolutionary scales. Genome Res. 2020;30(4):553–565.

Zhang G, Li C, Li Q, Li B, Larkin DM, Lee C, Storz JF, Antunes A, Greenwold MJ, Meredith RW, *et al.*; Avian Genome Consortium. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014;346(6215):1311–1320.

Zhang J, Nielsen R, Yang Z. Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. Mol Biol Evol. 2005;22(12):2472–2479.

Communicating editor: A. Rokas