Title: Phylotranscriptomics reveals convergent behavioral evolution associated with shared and unique mechanisms in cavity-nesting songbirds

Authors: Sara E Lipshutz[1,2,3], Mark S Hibbins[1,4], Alexandra B Bentz[1,5], Aaron M Buechlin[6], Tara A Empson[1,7], Elizabeth M George[1,8], Mark E Hauber[9,10,11], Doug B Rusch[6], Wendy M Schelsky[10], Quinn K. Thomas[2], Samuel J. Torneo[1], Abbigail M Turner[1,9], Sarah E Wolf[1,12], Mary J Woodruff[1], Matthew W Hahn[1,6,13], Kimberly A Rosvall[1]

*correspondence to: sara.lipshutz@duke.edu or krosvall@indiana.edu

Affiliations:
1: Department of Biology, Indiana University, Bloomington, IN, USA
2. Department of Biology, Loyola University Chicago, IL, USA
3: Department of Biology, Duke University, Durham, NC, USA
4: Department of Ecology & Evolutionary Biology, University of Toronto, Canada
5: School of Biological Sciences, University of Oklahoma, OK, USA
6: Center for Genomics and Bioinformatics, Indiana University, Bloomington, IN, USA
7: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, USA
8: Department of Evolution, Ecology, and Organismal Biology, The Ohio State University, OH, USA
9: Department of Evolution, Ecology, and Behavior, University of Illinois, Urbana-Champaign, IL, USA
10: Illinois Natural History Survey, Prairie Research Institute, University of Illinois, Urbana-Champaign, IL, USA
11: Advanced Science Research Center and Program in Psychology, Graduate Center of the City University of New York, New York, NY, USA
12: School of Biological Sciences, University of Edinburgh, Scotland
13. Department of Computer Science, Indiana University, Bloomington, IN, USA

Author Contributions:

S.E.L. and K.A.R. conceived of the study; S.E.L. led the fieldwork, supported by A.B.B., A.J.T., E.M.G, M.E.H., M.J.W., S.J.T, S.E.W., T.A.E., and W.M.S. S.E.L., M.S.H., A.M.B., D.B.R., Q.K.T., M.W.H. and K.A.R. analyzed and interpreted the data; S.E.L. M.S.H. and K.A.R. wrote the paper, and all authors contributed to revisions. The manuscript reflects the contributions and ideas of all authors.

Competing Interests: Authors declare no competing interests.

Classification: Biological Sciences, Evolution

Keywords: Competition, Behavioral Physiology, Molecular Evolution, Transcriptomics, Songbird, Aggression

This PDF file includes: Main Text, Figures 1 to 3

Abstract (241/250)

Uncovering the genomic bases of phenotypic adaptation is a major goal in biology, but this has been hard to achieve for complex behavioral traits. Here, we leverage the repeated, independent evolution of obligate cavity-nesting in birds to test the hypothesis that the shared pressure to compete for a limited breeding resource drives convergent behavioral evolution via convergent gene regulatory changes in the brain. Using behavioral assays in the field, hormonal measures of free-living subjects, and transcriptome-wide analyses of the brain in wild-captured males and females, we examined five species pairs across five avian families, each including one obligate cavity-nesting species and a related species with an open cup-nesting or otherwise more flexible nest strategy. Results support the hypothesis of behavioral convergence, with higher levels of territorial aggression in obligate cavity-nesters, particularly among females. Levels of testosterone in circulation were not associated with nest strategy or aggression for either sex, but phylogenetic analyses of individual genes and co-regulated gene networks revealed some shared patterns of gene expression, including a nest strategy-related gene network shared across families and two separate networks linked to aggression only in females. Though associated with convergent behavioral evolution, these genes were not significantly enriched for particular functional pathways, and the scope of convergent gene expression evolution was limited to a small percent of the genome. Together, these observations indicate that replicated evolutionary changes in complex behavior arise via a combination of convergent and lineage-specific evolution of gene regulation.

Significance (120/120)

Trait convergence has tremendous potential to reveal the proximate mechanisms by which natural selection generates evolutionary change. However, these investigations are often limited to traits with a simple and well-understood genetic basis, and less is known about the underpinnings of convergence in complex quantitative traits like many behaviors. Across five avian families with independent origins of obligate cavity-nesting, we find behavioral convergence, particularly among females, which display higher territorial aggression than their more flexibly-nesting relatives. We find both shared and unique patterns of brain gene expression, indicating that evolution follows diverse paths towards building an aggressive bird. These results highlight how complex behaviors can arise from a combination of repeated and independent molecular changes in response to shared ecological pressures.

Main Text

Introduction

Biologists have long been fascinated by striking phenotypic similarities among distantly related organisms exposed to similar ecological pressures (1). Research on the molecular basis of these convergent phenotypes has yielded critical insights into the predictability of evolution (2-5). Most comparative work on this issue has focused on morphological and physiological traits, such as independently derived coloration (6) or high-altitude tolerance (7). However, we know less about the molecular mechanisms underlying convergent behavioral traits (8-10), particularly complex behaviors that are likely to have a polygenic basis. Some studies of behavioral convergence have found that evolution repeatedly uses the same (or similar) changes in particular genes or pathways to produce phenotypic convergence (11, 12). Alternatively, the building blocks of phenotypic convergence may be unique, with lineage-specific mechanisms

across replicated evolutionary events (13-15). Understanding the contributions of shared vs unique genomic mechanisms of behavioral convergence requires a phylogenetic approach. However, efforts to connect behavioral evolution to transcriptomics rarely use phylogenetic methods (16-18), despite the potential for changes in gene expression to shape phenotypic diversification at the macroevolutionary scale.

Territorial aggression is a complex behavioral trait that is widespread across the animal kingdom and is therefore well-suited for evaluating potential molecular convergence. In general, we know much about the regulation of aggression, which in vertebrates is mediated by diverse neuroendocrine mechanisms acting across a constellation of hypothalamic and limbic brain regions (19, 20). A key focus has been the hormone testosterone (21) and its metabolite 17β-estradiol, which can act on sex steroid receptors in the brain to promote aggression (22, 23). Aggression is also linked to brain metabolic pathways (11) and G-protein coupled receptor signaling of dopamine, serotonin, and glutamate (24, 25). Many of these candidates are conserved across distantly related species (11, 24), indicating the potential for shared mechanisms in the independent evolution of territorial aggression. However, there is also interspecific variation in the abundance and distribution of these endocrine-molecular building blocks (26, 27), indicating that mechanisms of aggression can diverge over time.

Mechanisms of aggression can also differ within a species, including between males and females (28), both of whom exhibit territorial aggression in many species (29, 30). This provides a unique opportunity to include both sexes in analyses of behavioral evolution. After all, females and males share the majority of their genome, yet sex-specific selective pressures can alter gene expression (31) and hormone secretion (32), which together may shape aggression or the mechanisms that promote the expression of aggression, the latter of which is still remarkably understudied in females (33, 34).

Critically, in both sexes, aggression shapes access to resources and mates, and the degree of such competition varies among species (29, 30). Competition for breeding territories is thought to be especially intense for obligate secondary cavity-nesting birds, which cannot excavate holes for nesting and must secure a natural or previously excavated space inside a tree or other substrate to breed. Physical conflicts over limited cavities can lead to serious injury or death (35-37). Cavity-nesting has arisen independently over 30 times in birds (38), providing a strong foundation to evaluate the degree of mechanistic convergence in behavioral evolution. For many cavity-nesting species, both females and males compete to acquire and defend nesting territories (39-43), and our recent case study showed elevated aggression in two species of obligate cavity-nesters, though without quantitative phylogenetic methods (44).

Here, we test the hypothesis of evolutionary convergence using a phylogenetic approach that integrates behavioral, hormonal, and transcriptome-wide neurogenomic data. We compare species pairs from five avian families, which each represent independent origins of obligate secondary cavity-nesting (Figure 1A) (45). Each pair includes one obligate secondary cavity-nesting species, which regularly nest in artificial nest boxes, and one species with a more flexible nest strategy, either open cup-nesting or facultative cavity-nesting. Critically, each species pair is similar in other aspects of its ecology and life history (e.g. foraging ecology, degree of biparental care). Across species, we compared territorial aggression, testosterone in circulation, and brain gene expression.

As predicted, we find that the obligate cavity-nesters display elevated levels of territorial aggression, particularly in females (Figure 1B). This behavioral convergence is not associated with higher levels of testosterone. However, using multiple phylogenetically informed

approaches to compare brain gene expression, we find a combination of convergent and unique transcriptomic signatures of both aggression and cavity-nesting. Together, these results indicate that convergent behavioral evolution is associated with multiple regulatory routes.

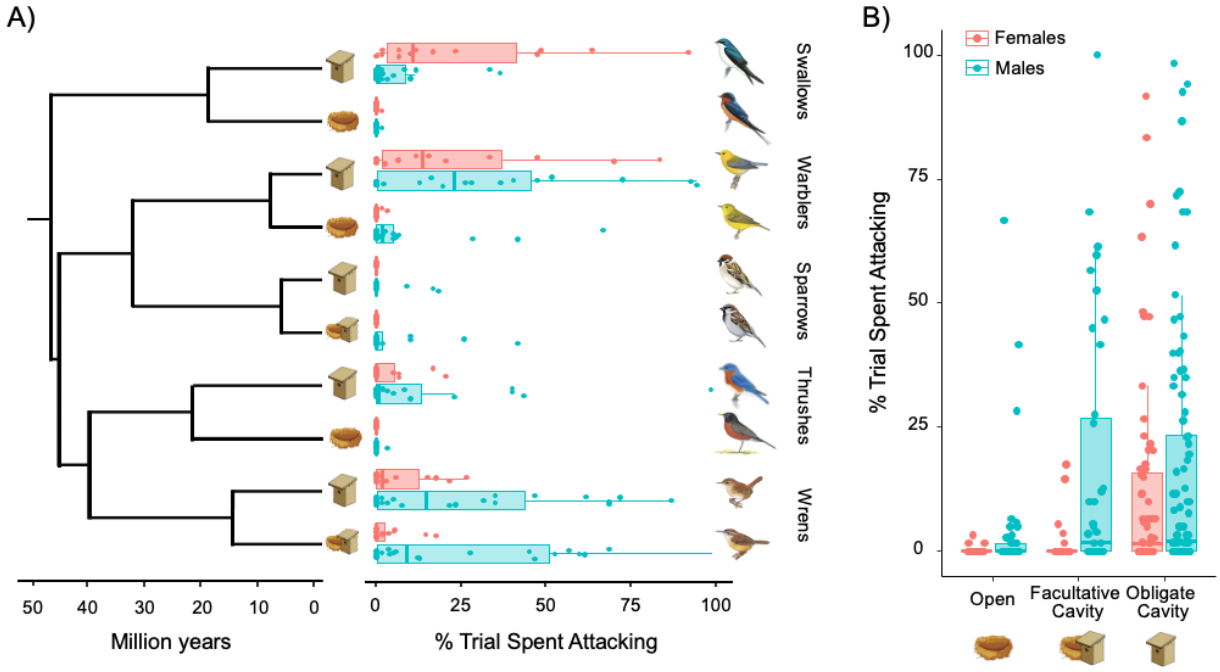

Figure 1: (A) *Left*: Consensus phylogeny of 10 species from five families, which diverged from a common ancestor ~44 mya (36-50 mya). Species pairs diverged ~9-20 mya (46). Nestboxes indicate obligate cavity-nesters, nests represent open nesters, and both together represent facultative cavity nesters. *Right*: Sex- and species-level aggression towards a conspecific decoy, measured by the proportion of 5-sec intervals that contained physical contact with the decoy during a 5-min aggression assay. Species listed in descending order: Swallows (Hirundinidae): tree swallows (*Tachycineta bicolor*), barn swallows (*Hirundo rustica*), Woodwarblers (Parulidae): prothonotary warblers (*Protonotaria citrea*), yellow warblers (*Setophaga petechia*), Sparrows (Passeridae): Eurasian tree sparrows (*Passer montanus*), house sparrows (*Passer domesticus*), Thrushes (Turdidae): Eastern bluebirds (*Sialia sialis*), American robins (*Turdus migratorius*), and Wrens (Troglodytidae): house wrens (*Troglodytes aedon*), Carolina wrens (*Thryothorus ludovicianus*). **(B)** Aggression, grouped by nest strategy. Each point is one assay on a unique free-living individual; box plots convey interquartile range. Obligate cavity-nesting females were significantly more aggressive than females with more flexible nest strategies. Illustrations reproduced with permission from Lynx Edicions.

Results and Discussion

Behavioral convergence among obligate cavity-nesters

Our study occurred during territorial establishment at the beginning of the breeding season, when we assayed aggression in 304 free-living female and male individuals in southern Indiana, Illinois, or Kentucky, USA (SI §1, Table S1). We focused on species pairs from five avian families – Hirundinidae (swallows), Parulidae (woodwarblers), Passeridae (sparrows), Turdidae (thrushes), and Troglodytidae (wrens); further details in SI §1. Each behavioral assay lasted for 5 minutes (min) and included a conspecific, taxidermic decoy of either female or male sex, coupled with playback of female or male vocalizations that occur during natural aggressive interactions (details in SI §2). We measured a variety of songbird aggressive behaviors and

focused on physical attacks (i.e. contact with the decoy). We calculated a maximum attack score of 60, based on the number of 5-second (sec) intervals that contained any physical contact. To visualize this behavior, we converted attack scores to a proportion of the trial spent attacking (# intervals including attack/ total # intervals x 100). We also measured average 'distance from decoy' to confirm each subject was present and engaged during a trial, regardless of whether they were aggressively attacking.

To evaluate the role of nest strategy as a driver of territorial aggression, we accounted for phylogenetic relationships among species using phylogenetic generalized linear mixed models (PGLMM). Briefly, these models correct for covariation in traits among species due to common ancestry by incorporating phylogenetic relatedness as a random effect. Along with sex of the territory holder and sex of the decoy, we included nest strategy as a fixed effect with two levels: obligate cavity-nesting vs. more flexible nest strategies (including both facultative cavity and open-nest). Alternative three-level models differentiating obligate cavity, facultative cavity, and open nest strategies are discussed in SI § 10-12). For downstream analyses, we include attack averages among species and sex categories as a fixed effect, referred to simply as 'aggression.' Unless otherwise noted, analyses used R version 4 (R-Core-Team 2019).

We found significant effects of nest strategy, sex of the territory holder, sex of the decoy, and their interaction on *aggression*. Obligate cavity-nesters spent more time attacking the decoy, compared to species with non-obligate strategies (PGLMM: $p = 0.0011$; Figure 1A; Table S2). Males attacked marginally more than females (PGLMM: $p = 0.058$), and sex interacted with nest strategy (PGLMM: $p = 0.0064$). This latter pattern was driven by higher aggression in obligate cavity-nesting females (Figure 1B), for whom aggressiveness is adaptive during competition for nesting territories (30, 40, 41). We also evaluated the effect of decoy sex, i.e., whether aggression was biased towards same vs. different sex intruders. We found a significant interaction between sex of the territory holder and sex of the decoy (PGLMM: $p = 0.011$), such that male territory holders attack male decoys significantly more, whereas female territory holders attack female and male decoys with similar levels of aggression (Figure S1; Table S2). Using these same methods, we evaluated *distance from the decoy*, which was not related to nest strategy (PGLMM: $p = 0.22$), sex ($p = 0.70$), nor their interaction ($p = 0.39$) (Table S3; Figure S2). All species were in audio-visual proximity to the simulated intruder (average distance = 5.0m), but obligate cavity-nesters spent more time attacking the decoy compared to their close relatives with more flexible nesting strategies (Figure 1, Table S2). These results are consistent with the hypothesis that an obligate strategy promotes aggression among conspecifics, particularly for females who cannot reproduce without a cavity (44). This finding also complements a recent study noting that cavity-nesters display more territorial behaviors against heterospecifics (47), underscoring multiple contexts linking territorial aggression with a cavity-nesting strategy.

Hormonal and neurogenomic mechanisms of behavioral convergence

Next, we sought to measure constitutive levels of testosterone and brain gene expression, as potential physiological drivers of behavioral convergence. We focused our analysis on baseline physiological and functional neurogenomic states that are representative of unprovoked, free-living animals (n = 196 total hormone samples, n = 121 brain-derived RNA libraries; sample sizes subset by species and sex in Table S1). See SI §3 for details on sample collection.

Males had significantly higher levels of testosterone in circulation than females (PGLMM: $p < 0.0001$), but testosterone was not related to nest strategy ($p = 0.80$) nor the interaction between

nest strategy and sex (p = 0.85; Figure S4, Table S4, SI §4). Testosterone-focused hypotheses have dominated aggression-related research for decades (21, 48), but here we find that testosterone does not explain species-level differences in female or male aggression, at least when birds are sampled at a baseline state.

Using RNA-seq, we measured mRNA abundance for 10,672 orthologous genes expressed in all 10 focal species in the ventromedial telencephalon, a conserved region of the vertebrate social behavior network associated with the regulation of aggression (19). PCA revealed clustering by sex, species, and family (Figure S5). An UpSet plot identified 221 genes consistently differentially expressed for all 5 families (Data S3, Figure S6). The number of differentially expressed genes generally increased with divergence time between species pairs within each family (p = 0.08, R^2 = 0.83; SI §6, Figure S7, Table S5), underscoring the need for phylogenetic methods. Using these expression data, we took two analytical approaches to understand the transcriptomic basis of convergent evolution associated with obligate cavity-nesting.

Gene expression is highly concordant between pairs of families, but involves different sets of genes across the phylogeny

First, we used the Rank Rank Hypergeometric Overlap (RRHO) approach to compare the magnitude and direction of gene expression differences among species pairs. RRHO revealed highly concordant patterns of gene expression across pairwise independent origins of obligate cavity-nesting (all p-values < 0.01, permutation test, Figure 2, Table S6). Genes that were more highly expressed in the obligate cavity-nester of one species pair were also more highly expressed in the obligate cavity-nester of another species pair, and vice-versa for genes with lower expression (Figure 2, Data S4). This concordant pattern applied to an average of 1390 ± 65 genes per contrast, amounting to ~13.0% of genes studied. To give one example, *GRIN1*, which encodes a glutamate receptor subunit, is more highly expressed in the cavity-nesting tree swallow compared to the open-nesting barn swallow. This same pattern applied for the cavity-nesting bluebird compared to the open-nesting robin. For each pairwise contrast, though, these concordant genes were not enriched for any specific functions (Data S4), based on Gene Ontology (GO). These results suggest that, while there is a significant degree of concordance in expression differences between pairs of species, these concordantly expressed genes are not repeatedly recruited from any particular set of functional pathways.

If the mechanisms that give rise to obligate cavity-nesting and associated aggressiveness converge alongside behavioral convergence, then concordance should also be strong among *multiple* pairs of species. Roughly 1,000 genes (~10% of orthologs) displayed concordant expression patterns among two or three families, and about three hundred (~3%) genes were shared by up to 6 family comparisons at a time (Figure 2A). Among this latter set, we noted dozens of genes with known links to behavioral neuroendocrinology, including those associated with steroid hormones (*PGR*) and monoamine neurotransmitters (*ADRB2, DRD5, TH, TPH1*). We also noted concordant genes associated with energetics, including insulin signaling (*IGF1, IGFBP3*), mitochondrial function (*APC2, CYCS, PC, TIMM10, TIMM22, TOM22*), and metabolism (lipid: *ABCA7, CPT1C, FITM2, SCD*; carbohydrate: *B3GALT1, B3GALT6, GALT*). However, GO analysis of genes shared by at least half of the RRHO comparisons (Data S5) again revealed no significant functional enrichment, echoing the conclusion from the paired comparison that no particular biological process is more or less likely to contribute to convergence.

Across all 10 possible RRHO contrasts, 11 genes were shared, meaning that 0.1% of 10,672 orthologs exhibited complete concordance across five independent origins of obligate cavity-nesting (Table S7). Given that the number of shared genes decreases rapidly with increasing inclusion of family comparisons, this result suggests a high degree of lineage specific gene expression evolution in association with cavity nesting (Figure 2A). Some of the globally concordant genes have established connections with behavior or brain function (49). For instance, *EMC3, NPTX1, and RGS19* are involved in neurotransmitter reception, *ATP6V0E1* and *DDX56* have ATPase activity, and most of the completely concordant genes have established connections with Alzheimer's (*ARRDC4, CTNND1, NPTX1, WIPF2*), addiction (*RGS19*) or other neurological disorders (*AFAP1L1*, *PQBP1*) (49).

We repeated these RRHO analyses with females and males separately, because sex-specific selection pressures could generate concordance that would be masked by analyzing both sexes together. The sexes differed in both the degree and direction of concordance. In females, we saw a higher degree of concordance among 2-4 families (Figure 2A). Female obligate cavity nesters generally had lower expression than non-obligate species (Figure S8), whereas male obligate cavity nesters generally had higher expression than non-obligate species, though this pattern was not consistent for every comparison (Figure S9). Overall, there were very few genes shared across all pairs of comparisons, highlighting again that global concordance is low.

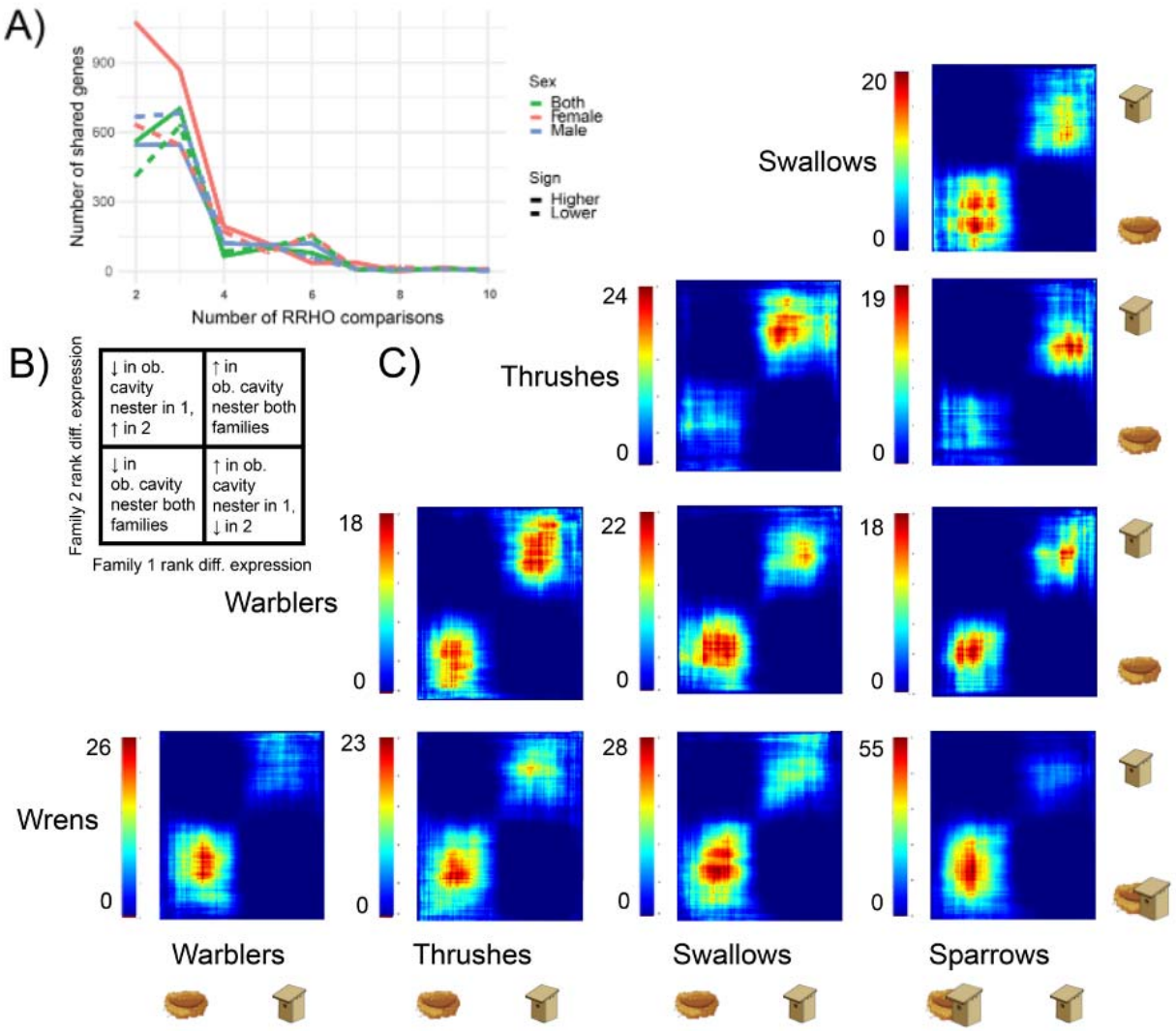

Figure 2: Concordance in differential expression based on rank-rank hypergeometric overlap (RRHO). Heatmap colors reflect adjusted hypergeometric -log(p-value). Larger values indicate more concordance in expression between families. Nestboxes indicate obligate cavity-nesters, nests represent open nesters, and both together represent facultative cavity nesters. **(A)** Number of concordantly expressed genes shared across family comparisons (i.e. across heatmaps) for males (blue), females (pink) and both sexes combined (green). Solid lines indicate higher expression in obligate cavity-nesters, dashed lines indicate lower expression **(B)** Key for interpreting individual heatmaps. Each quadrant of the key corresponds to a quadrant of an individual heatmap, shown in **(C)** Each pixel within the heatmaps contains two sets of approximately 100 genes being compared between the two families; the color legend indicates the log p-value of the overlap between these gene sets, with a higher value indicating stronger overlap.

Nest strategy and aggression are associated with convergently evolving genes

Next, we used PGLMMs to model expression of all ~10k orthologs as a function of nest strategy, sex, aggression, sex-by-nest-strategy interactions, and aggression-by-nest-strategy interactions. This approach extends the perspective gained from RRHO but has the added advantage that it explicitly (i) accounts for phylogenetic relationships and (ii) directly tests for effects of nest strategy, aggression, sex, the interaction of nest strategy with aggression, and the interaction of nest strategy with sex. This extension is critical because aggression is not the

only potentially evolving trait along obligate cavity-nesting lineages, and there is substantial among-species variation in the degree of territorial aggression, which is not explored in the RRHO approach.

PGLMMs identified hundreds of genes whose expression was associated with our model terms (Table S8, Data S6). Sex had the largest influence on expression variation and was associated with 510 genes (Table S8). Many more genes had higher expression in males than females. This pattern of sex-biased gene expression could stem from stronger sexual selection on males combined with reduced constraint on the evolution of higher expression, which would produce an upward expression bias in male-biased genes (50). Additionally, in birds, dosage compensation of the Z chromosome may be incomplete or absent (51), leading to higher expression of Z-linked genes in males than females, since males carry two copies of the Z chromosome. Genes with sex-biased expression were significantly enriched for cellular metabolic process.

Many genes showed some relationship with obligate cavity nesting, aggression, or a relevant interaction. Specifically, there were 234 genes associated with obligate cavity-nesting, 79 genes associated with aggression, and five genes overlapped among these: *ERN1, HIGD1A, RNASEH2B, TAF1B,* and *UGGT2*. An additional 62 genes exhibited an interaction between nest strategy and aggression. 76 genes exhibited an interaction between sex and obligate cavity-nesting, with approximately equal numbers of genes that had a stronger association in females vs. males.

Some nest-strategy and aggression-associated patterns appeared to be driven by expression in one or two species (Figure S11b,c), and so we developed additional criteria to pinpoint a robust set of convergently evolving genes. We retained only those that were (i) significantly different in expression (by *t*-test) between species in at least 3 out of 5 family comparisons; and (ii) different in a consistent direction (i.e. all higher or all lower expression) with respect to nest strategy in all families with a significant difference. We found more genes that met our criteria of convergence for nest strategy [168 genes out of 234 (72%)] in contrast to the interaction of nest strategy and aggression [8 out of 62 (12.9%)] or the interaction of nest strategy and sex [7 out of 76 (9.2%)] (Data S8). However, even among genes associated with nest strategy, this convergence only represents a small proportion, 1.6% of total orthologs. Thus, we find that the convergent evolution of complex behaviors is underlain by a relatively small number of convergently evolving genes, alongside a larger number of genes subject to lineage-specific evolution.

None of these convergent gene sets had significant GO enrichment (Data S8). However, several convergent genes associated with nest strategy have some connection to ATP and mitochondrial function (*ATP1B1, ATP5ME, HIGD1A, MRPL19, PITRM1, SLC25A24*), heat shock proteins (*DNAJC24, HSPA2, HSPB11*), and telomere dynamics (*POT1, HAT1*). Additional genes on this list have established connections to behavior or psychiatric risk (*TRMT1L, NT5C2, BCHE*) and neuron signaling (*GRM5, OMG, NLGN2, NPTN*) (49). Genes associated with aggression also have connections to neurodevelopmental and neurodegenerative disorders (e.g. *COX7C*, *EIF2B3*, *ERN1*, *NDUFA1, NPAS3*, *PLAA*) (49). Convergent genes associated with the interaction of nest strategy and aggression include *UNC13B*, which is involved with epilepsy (52) as well as *SREK1* and *TAF1B*, which are associated with Huntington's disease (53). Several convergent genes associated with the

interaction of nest strategy and sex relate to immune function, including *RELT* (54) and *SCRN1* (55).

Nest strategy and female aggression are associated with convergent gene modules

Complex phenotypes are often regulated by subtle but coordinated changes in gene networks (56), and approaches that analyze genes individually may miss important emergent properties. Therefore, we used weighted gene co-expression network analyses (WGCNA) to estimate co-expression across all 10 species (SI §9), yielding 93 modules of correlated genes (Data S10). Using each module's eigengene (akin to a principal component of expression), we ran PGLMM with fixed effects of nest strategy, aggression, sex, and their interactions. After FDR correction, two modules showed significant associations: (i) expression in the red module was associated with sex (Figure S11a), and (ii) expression in the tan4 module was associated with nest strategy (Table S9), with lower eigenvalues in each of the five independent origins of obligate cavity-nesting (Figure 3). As expected, the composition of the tan4 module's 44 genes recapitulated many of the individual convergent genes associated with nest strategy. Though not significantly enriched for any GO terms (Table S9), several genes in the tan4 module related to mitochondrial function and energy metabolism, cognition, and stress response.

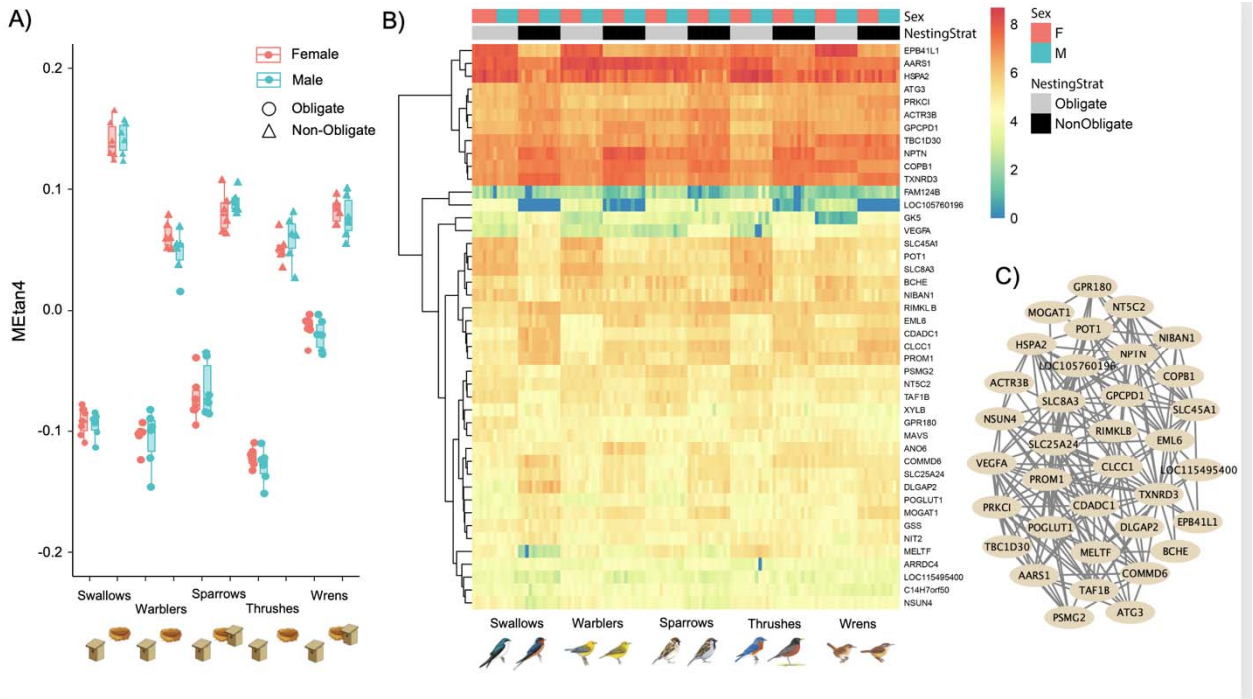

Figure 3. (A) Weighted gene co-expression network analysis (WGCNA) eigengene values for the tan4 module, which included 44 genes whose expression was significantly associated with obligate cavity-nesting relative to non-obligate nesting strategies. Individual points represent the eigengene value for a sampled individual, colored by sex. Nestboxes indicate obligate cavity-nesters, nests represent open nesters, and both together represent facultative cavity nesters. **(B)** Heatmap of log-scaled genes in tan4 module. Columns represent individuals, grouped by sex and nest strategy. Color scale indicates log of gene expression. **(C)** Network depicts genes with module membership |>| 0.6. Illustrations reproduced with permission from Lynx Edicions.

We repeated our phylogenetic analysis on WGCNA modules constructed for each sex individually (SI §11). In this sex-specific analysis, we identified two female-specific modules correlated with aggression – pink4 and sienna4 (Figure S13, Table S10). Though not enriched for any GO terms, these modules contained genes associated with the regulation of aggressive behavior, including DRD3, a dopamine receptor in pink4, and GRIA2, a glutamate receptor in sienna4. Finding aggression-associated modules specific to females further suggests that there are multiple gene regulatory routes to aggression, including those that may differ between sexes.

Gene expression provides many routes to convergent behavioral evolution

Integrating results across different analyses, we found 25 nest strategy-associated genes that overlapped across individual gene- and network level PGLMMs, six of which also overlapped with differentially expressed genes between species pairs: *BCHE, CDADC1, EML6, MOGAT1, POGLUT1,* and *TXNRD3* (Table S12). This set of 25 genes represents ~57% of genes in the tan4 module, a strong signal that this gene set is consistently associated with the convergent evolution of obligate cavity-nesting. Among these 25 genes, only *TAF1B* overlapped with the set of aggression-associated genes. Similarly, trait associated gene sets from PGLMMs did not overlap with gene sets identified in at least 5 out of 10 RRHO comparisons (Data S5). While RRHO revealed broad patterns of expression concordance across species pairs, it is not an explicit statistical test for associations with nest strategy or aggression, and it does not account for phylogenetic relationships. Consequently, many of the genes identified as concordantly evolving in RRHO may not be functionally related to nest strategy evolution. Thus, although our results are consistent with coevolution between obligate cavity nesting and aggressive behavior, the transcriptomic mechanisms associated with these traits are only partially overlapping. This lack of overlap is perhaps not surprising – the cavity-nesting strategy is likely associated with a suite of traits including but not limited to aggression, and aggressive behavior may be shaped by factors in addition to cavity-nesting.

To place our results in a broader context, we compared our list of aggression-associated genes to prior studies (Data S12), including those on the brain's response to an aggressive challenge (11, 57) and those that are known to mediate aggression in diverse species, including humans (25, 58). Using a permutation analysis, we found significant overlap between our aggression-associated genes and a module of genes in the tree swallow brain that was sensitive to experimental competition (SI §12, Table S13, Figure S18; (57)). Of the seven genes in common between these datasets (Table S14), *DPF3*, *PLXND1*, and *ZIC4* are involved in brain development, *CHN2* is associated with schizophrenia, *SPHKAP* is associated with neuroblastoma and apoptosis, and *GDF10* promotes neural repair after stroke (49), again linking aggression to key elements of brain function. The overlap of these two datasets also indicates concordance between short-term functional responses to competition and long-term evolutionary change in genes whose expression is sensitive to the social environment (59).

Conclusions

Biologists have long been interested in the molecular basis of convergent phenotypic evolution 2-4). One unresolved question is whether the repeated evolution of complex behavioral traits is

underlain by consistent or variable molecular mechanisms. In the case of the cavity-nesting in five families of birds, we find diverse patterns of gene expression in the brains of more aggressive obligate cavity nesting species, with a combination of convergent and lineage-specific shifts along branches towards obligate cavity nesting. This work highlights the power of comparative approaches to uncovering the mechanistic bases of complex, convergently evolving phenotypes (17, 18, 24). Notably, our study is among the first to use phylogenetically explicit models of gene expression to understand the evolution of such a complex and continuously varying behavioral trait. Our findings support the hypothesis that there are many routes to building a pro-aggression brain, layered atop a small set of convergently evolving genes.

Materials and Methods

This work was approved by Indiana University Bloomington IACUC #18-004 and #21-003, and relevant federal (MB59069B) and state permits (IL: W20.6355, KY: SC1911001, IN: 18-030, 19-272, 2545).

Assay of aggression

For cavity-nesting species, we placed the conspecific decoy on the nestbox and hung a Bluetooth speaker nearby. For non-cavity-nesters, we located nest sites or observed individuals for at least an hour to determine where they spent their time. Additional details on audio stimuli and decoys can be found in SI §2. We played a conspecific vocal lure to capture the attention of the focal individual and waited 30 s before beginning the 5-min aggression assay. We measured a suite of aggressive behaviors (see SI §2) and focused on the proportion of the trial spent physically attacking the decoy. We also measured distance from the decoy to confirm that all focal territory holders were present and engaged with the simulated intruder. We evaluated the effects of nest strategy, sex, decoy sex, and their interaction on physical attacks and distance from the decoy using phylogenetic linear mixed models (details below and in SI §10).

Brain dissection and RNA isolation

For 6 or 7 individuals per sex per species, we dissected whole brains into functional regions following (60). Gene expression analyses focused on the ventromedial telencephalon (VmT), which contains the avian medial amygdala, or TnA, and nearby limbic regions that are key nodes in the vertebrate social behavior network (19). We extracted total RNA using Trizol (Invitrogen), and resuspended RNA in UltraPure water. Tape Station Bioanalyzer (Agilent) on 121 samples showed RIN = 8.56 ± 0.05 (mean ± SE).

Sequencing, alignment, and mapping

We submitted total RNA to the Indiana University Center for Genomics and Bioinformatics, using paired end sequencing (Illumina NextSeq500, 75 cycle sequencing module) to generate an average of ~28.6 million reads per sample (Data S1). We used Trinity version 2.13.2 (61) to assemble transcriptomes per species, which we aligned to the zebra finch (*Taeniopygia guttata*) proteome National Center for Biotechnology Information (NCBI) Release (GCF_003957565.2_bTaeGut1.4.pri_protein.faa) (62) using tblastn v. 2.2.9. After alignment and mapping, downstream analyses identified 10,672 orthologous genes with high confidence in all 10 species (SI §5). We normalized counts (Data S2) and performed differential gene expression using DESeq2 (63), see SI§6.

Rank-Rank Hypergeometric Overlap

To identify shared transcriptomic patterns in obligate cavity-nesting species across pairs of families, we used the R package Rank-Rank Hypergeometric Overlap (RRHO) (64) (details in SI §8). RRHO uses normalized counts to rank genes based on the direction and magnitude of expression difference between groups, in our case each obligate cavity-nester and its species pair within the same family. Hypergeometric tests evaluate overlap among these ranked lists, visualized as heatmaps of p-values from 100 permutations. The pipeline also identifies the set of genes with the highest degree of concordance for each comparison.

Weighted Gene Co-Expression Network Analysis

We constructed gene networks for both sexes combined using WGCNA (65). We used the normalized counts from DESeq2 and filtered out genes with <15 norm counts in 90% of the samples. We generated a signed hybrid network by selecting a soft threshold power (β) = 6, in accordance with scale-free topology. We calculated a minimum module size of 30 and used a biweight midcorrelation (bicor) function. We merged modules in Dynamic Tree Cut using a threshold of 0.25. Genes with an absolute module membership value >0.6 were assessed for enrichment of biological processes in PantherGO (66), using our list of 10,672 orthologs as a reference set. Networks of interest were visualized in Cytoscape v3.10.1 (67). See SI §9 for gene network analyses run separately on females and males.

Phylogenetic Generalized Linear Mixed Models

Phylogenetic generalized linear mixed models (PGLMM) used MCMCglmm (68), with the phylogenetic covariance matrix from the consensus tree as a random effect (SI §10). Our consensus tree (Figure 1A) made use of the BirdTree.org (69) resource to obtain 1000 ultrametric trees from Ericson All Species. We constructed a majority rule consensus topology from this tree set using the *ape* (70) function 'consensus', and inferred branch lengths using the *phytools* (71) function 'consensus.edges' with the least squares method. For each PGLMM, we evaluated MCMC convergence using Heidelberger and Welch's convergence diagnostic as implemented by the heidel.diag function in the R package coda.

To evaluate whether nest strategy predicted aggression, average distance from the decoy, and testosterone, we fit models with nest type, sex, and their interaction as fixed effects. Our aggression dataset contained many individuals that never attacked, and so we used a zero-inflated binomial model, with random effect and residual variance priors fixed at 1, the default burn-in period of 3000, and 2,000,000 iterations. For this aggression analysis, technical limitations restricted our ability to specify obligate cavity-nesting as the baseline factor level; therefore, we combined facultative nesting and open nesting into a single category to contrast with obligate cavity-nesting. For all other PGLMMs, we fit models containing either two (obligate vs. non-obligate) or three (obligate vs. facultative vs. open) factor levels (SI §10). For these other models, we fit a standard Gaussian model with no priors, the default burn-in, and 100,000 iterations.

To predict nest-strategy and aggression associated genes from our set of ~10k orthologs, we included species/sex-level aggression sex, nest strategy, nest strategy by sex interaction and nest strategy by aggression interaction as fixed effects. After fitting a model for the log-expression of each gene, we applied a Benjamini-Hochberg false discovery rate correction in the *statsmodels* package (72) in Python version 3.7.11. Since our PGLMMs could not estimate p-values with precision lower than $1x10^{-4}$ (due to computational constraints), we set all raw p-values reported as less than this value to exactly $1x10^{-4}$. Since the Benjamini-Hochberg procedure relies on the relative ranks of raw p-values, rather than their magnitudes, this should not significantly affect our FDR correction.

To identify gene networks associated with nest strategy, sex, and their interaction we ran a similar phylogenetic analysis using the first eigengene value drawn from a weighted gene co-expression network analysis (SI §11).

Our PGLMM-related scripts are available at https://github.com/slipshut/CavityNesting/tree/main/PGLMM

Functional interpretation
We inferred gene function based on genecards.org (49) and GO analyses. We inferred GO terms from *Homo sapiens* because its ontologies are orthologous to, but more complete than avian references. We conducted an overrepresentation analysis of non-redundant biological process GO terms in PantherGO (66), using our list of orthologs as a reference set.

We also conducted a custom permutation analysis to evaluate the degree of overlap between our aggression-related genes and published genomic studies of aggression (11, 25, 57, 58) (Data S12; SI §12). Our enrichment scripts are available at https://github.com/slipshut/CavityNesting/tree/main/Enrichment

Acknowledgements
For facilitating fieldwork, we thank the Indiana Department of Natural Resources, IU's Research and Teaching Preserve, Thomas Rothfus and Chris Hagie at Therkilsden Field station, Auriel Fournier and Steve Havera at Forbes Field Station, Jeff Hoover, Nick Antonson, Hannah Scharf, and Angela Funk. For vocal stimuli, we thank Matt Wilkins, Dustin Reichard, Kara Krieg, and Elsa Chen. Thanks to Emmi Mueller for help with 3D printing. We thank Sumitha Nallu and Jie Huang for CGB TapeStation and library preparation; Emma Dossey Curole and Jace Kuske for assistance in the lab. We thank Becca Young for assistance with data analysis and Amanda Clark for assistance with data visualization.

Funding
US National Science Foundation (NSF) grant DBI-1907134 (to S.E.L.), DBI-2146866 (to M.W.H.), IOS-1953226 (to M.E.H.), and CAREER grant IOS-1942192 (to K.A.R.), as well as support from IU Biology and Duke Biology.

Data Accessibility
Raw sequence reads and count data can be obtained from the Gene Expression Omnibus database (GEO accession number GSE244480). Supplementary information, including R scripts and trait data can be obtained from Github: https://github.com/slipshut/CavityNesting.

References Cited

1. C. R. Darwin, *On the origin of species* (Harvard University Press, Cambridge, MA, 1859).
2. E. B. Rosenblum, C. E. Parent, E. E. Brandt, The Molecular Basis of Phenotypic Convergence. *Annual Review of Ecology, Evolution, and Systematics* **45**, 203-226 (2014).
3. J. F. Storz, Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* **17**, 239-250 (2016).

4. J. Arendt, D. Reznick, Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol* **23**, 26-32 (2008).
5. X. Chen *et al.*, Balanophora genomes display massively convergent evolution with other extreme holoparasites and provide novel insights into parasite–host interactions. *Nature Plants* **9**, 1627-1642 (2023).
6. L. Zhang, A. Mazo-Vargas, R. D. Reed, Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proc Natl Acad Sci U S A* **114**, 10707-10712 (2017).
7. X. Gou *et al.*, Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res* **24**, 1308-1315 (2014).
8. J. R. Gallant, L. A. O'Connell, Studying convergent evolution to relate genotype to behavioral phenotype. *J Exp Biol* **223** (2020).
9. N. Jourjine, H. E. Hoekstra, Expanding evolutionary neuroscience: insights from comparing variation in behavior. *Neuron* **109**, 1084-1099 (2021).
10. K. A. Rosvall, Evolutionary endocrinology and the problem of Darwin's tangled bank. *Horm Behav* **146**, 105246 (2022).
11. C. C. Rittschof *et al.*, Neuromolecular responses to social challenge: common mechanisms across mouse, stickleback fish, and honey bee. *Proc Natl Acad Sci U S A* **111**, 17929-17934 (2014).
12. A. L. Toth *et al.*, Brain transcriptomic analysis in paper wasps identifies genes associated with behaviour across social insect lineages. *Proc Biol Sci* **277**, 2139-2148 (2010).
13. A. D. Foote *et al.*, Convergent evolution of the genomes of marine mammals. *Nat Genet* **47**, 272-275 (2015).
14. A. S. Mikheyev, T. A. Linksvayer, Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *Elife* **4**, e04775 (2015).
15. M. R. Warner, L. Qiu, M. J. Holmes, A. S. Mikheyev, T. A. Linksvayer, Convergent eusocial evolution is based on a shared reproductive groundplan plus lineage-specific plastic genes. *Nat Commun* **10**, 2651 (2019).
16. M. C. Harrison *et al.*, Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol* **2**, 557-566 (2018).
17. J. B. Pease *et al.*, Layered evolution of gene expression in "superfast" muscles for courtship. *Proc Natl Acad Sci U S A* **119**, e2119671119 (2022).
18. R. L. Young *et al.*, Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc Natl Acad Sci U S A* **116**, 1331-1336 (2019).
19. J. L. Goodson, The vertebrate social behavior network: evolutionary themes and variations. *Horm Behav* **48**, 11-22 (2005).
20. R. J. Nelson, B. C. Trainor, Neural mechanisms of aggression. *Nat Rev Neurosci* **8**, 536-546 (2007).
21. J. C. Wingfield, R. E. Hegner, J. Dufty, Alfred M. , G. F. Ball, The "Challenge Hypothesis": Theoretical Implications for Patterns of Testosterone Secretion, Mating Systems, and Breeding Strategies. *The American Naturalist* **136**, 829-846 (1990).
22. B. M. Horton *et al.*, Estrogen receptor alpha polymorphism in a species with alternative behavioral phenotypes. *Proc Natl Acad Sci U S A* **111**, 1443-1448 (2014).

23. K. A. Rosvall *et al.*, Neural sensitivity to sex steroids predicts individual differences in aggression: implications for behavioural evolution. *Proc Biol Sci* **279**, 3547-3555 (2012).
24. M. C. Saul *et al.*, Cross-species systems analysis of evolutionary toolkits of neurogenomic response to social challenge. *Genes Brain Behav* **18**, e12502 (2019).
25. Y. Zhang-James *et al.*, An integrated analysis of genes and functional pathways for aggression in human and rodent models. *Mol Psychiatry* **24**, 1655-1667 (2019).
26. A. M. Kelly, L. C. Wilson, Aggression: Perspectives from social and systems neuroscience. *Horm Behav* **123**, 104523 (2020).
27. L. A. O'Connell, H. A. Hofmann, Evolution of a vertebrate social decision-making network. *Science* **336** (2012).
28. K. A. Rosvall, Proximate perspectives on the evolution of female aggression: good for the gander, good for the goose? *Philos Trans R Soc Lond B Biol Sci* **368**, 20130083 (2013).
29. T. Clutton-Brock, Sexual Selection in Males and Females. *Science* **318**, 1882-1885 (2007).
30. K. A. Rosvall, Intrasexual competition in females: evidence for sexual selection? *Behav Ecol* **22**, 1131-1140 (2011).
31. A. E. Wright, J. E. Mank, The scope and strength of sex-specific selection in genome evolution. *J Evol Biol* **26**, 1841-1853 (2013).
32. W. Goymann, J. C. Wingfield, Male-to-female testosterone ratios, dimorphism, and life history—what does it really tell us? *Behavioral Ecology* **25**, 685-699 (2014).
33. K. A. Rosvall, A. B. Bentz, E. M. George, How research on female vertebrates contributes to an expanded challenge hypothesis. *Horm Behav* **123**, 104565 (2020).
34. K. O. Smiley *et al.*, Beyond a biased binary: A perspective on the misconceptions, challenges, and implications of studying females in avian behavioral endocrinology. *Front Physiol* **13**, 970603 (2022).
35. R. A. Duckworth, Adaptive dispersal strategies and the dynamics of a range expansion. *Am Nat* **172 Suppl 1**, S4-17 (2008).
36. D. Leffelaar, R. J. Robertson, Nest usurpation and female competition for breeding opportunities by Tree Swallows. *The Wilson Bulletin* **97**, 221-224 (1985).
37. J. Merilä , D. A. Wiggins, Interspecific Competition for Nest Holes Causes Adult Mortality in the Collared Flycatcher. *The Condor* **97**, 445-450 (1995).
38. G. L. Davidson, A. Thornton, N. S. Clayton, Evolution of iris colour in relation to cavity nesting and parental care in passerine birds. *Biol Lett* **13** (2017).
39. A. N. Albers, J. A. Jones, L. Siefferman, Behavioral Differences among Eastern Bluebird Populations Could Be a Consequence of Tree Swallow Presence: A Pilot Study. *Frontiers in Ecology and Evolution* **5** (2017).
40. C. A. Krieg, T. Getty, Fitness benefits to intrasexual aggression in female house wrens, Troglodytes aedon. *Animal Behaviour* **160**, 79-90 (2020).
41. K. A. Rosvall, Sexual selection on aggressiveness in females: evidence from an experimental test with tree swallows. *Animal Behaviour* **75**, 1603-1610 (2008).
42. D. R. Rubenstein, I. J. Lovette, Reproductive skew and selection on female ornamentation in social species. *Nature* **462**, 786-789 (2009).
43. E. R. Schuppe, G. D. Sanin, M. J. Fuxjager, The social context of a territorial dispute differentially influences the way individuals in breeding pairs coordinate their aggressive tactics. *Behavioral Ecology and Sociobiology* **70**, 673-682 (2016).

44. S. E. Lipshutz, K. A. Rosvall, Nesting strategy shapes territorial aggression but not testosterone: A comparative approach in female and male birds. *Horm Behav* **133**, 104995 (2021).
45. R. Zenil-Ferguson, J. P. McEntee, J. G. Burleigh, R. A. Duckworth, Linking Ecological Specialization to Its Macroevolutionary Consequences: An Example with Passerine Nest Type. *Syst Biol* **72**, 294-306 (2023).
46. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol Biol Evol* **34**, 1812-1819 (2017).
47. J. P. Drury, M. C. Cowen, G. F. Grether, Competition and hybridization drive interspecific territoriality in birds. *Proc Natl Acad Sci U S A* **117**, 12923-12930 (2020).
48. J. F. Husak *et al.*, Life history and environment predict variation in testosterone across vertebrates. *Evolution* **75**, 1003-1010 (2021).
49. G. Stelzer *et al.*, The GeneCards Suite: From Gene Data Mining to Disease Genome Sequence Analyses. *Current Protocols in Bioinformatics* **54**, 1.30.31-31.30.33 (2016).
50. S. Grath, J. Parsch, Sex-Biased Gene Expression. *Annu Rev Genet* **50**, 29-44 (2016).
51. C. Mullon, A. E. Wright, M. Reuter, A. Pomiankowski, J. E. Mank, Evolution of dosage compensation under sexual selection differs between X and Z chromosomes. *Nat Commun* **6**, 7720 (2015).
52. J. Wang *et al.*, UNC13B variants associated with partial epilepsy with favourable outcome. *Brain* **144**, 3050-3060 (2021).
53. I. H. Hernández *et al.*, Pathogenic SREK1 decrease in Huntington's disease lowers TAF1 mimicking X-linked dystonia parkinsonism. *Brain* **143**, 2207-2219 (2020).
54. G. L. Sica *et al.*, RELT, a new member of the tumor necrosis factor receptor superfamily, is selectively expressed in hematopoietic tissues and activates transcription factor NF-κB. *Blood* **97**, 2702-2707 (2001).
55. G. Way, N. Morrice, C. Smythe, A. J. O'Sullivan, Purification and Identification of Secernin, a Novel Cytosolic Protein that Regulates Exocytosis in Mast Cells. *Molecular Biology of the Cell* **13**, 3344-3354 (2002).
56. N. M. Baran, P. T. McGrath, J. T. Streelman, Applying gene regulatory network logic to the evolution of social behavior. *Proc Natl Acad Sci U S A* **114**, 5886-5893 (2017).
57. A. B. Bentz *et al.*, Experimental competition induces immediate and lasting effects on the neurogenome in free-living female birds. *Proc Natl Acad Sci U S A* **118** (2021).
58. A. L. Filby, G. C. Paull, T. F. Hickmore, C. R. Tyler, Unravelling the neurophysiological basis of aggression in a fish model. *BMC Genomics* **11**, 498 (2010).
59. C. Alaux *et al.*, Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proc Natl Acad Sci U S A* **106**, 15400-15405 (2009).
60. K. K. Soma, B. A. Schlinger, J. C. Wingfield, C. J. Saldanha, Brain aromatase, 5 alpha-reductase, and 5 beta-reductase change seasonally in wild male song sparrows: relationship to aggressive and sexual behavior. *J Neurobiol* **56**, 209-221 (2003).
61. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).
62. A. Rhie *et al.*, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737-746 (2021).

63. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
64. S. B. Plaisier, R. Taschereau, J. A. Wong, T. G. Graeber, Rank–rank hypergeometric overlap: identification of statistically significant overlap between gene-expression signatures. *Nucleic Acids Research* **38**, e169-e169 (2010).
65. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
66. H. Mi, A. Muruganujan, D. Ebert, X. Huang, P. D. Thomas, PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Research* **47**, D419-D426 (2018).
67. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
68. J. D. Hadfield, MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**, 1-22 (2010).
69. W. Jetz, G. H. Thomas, J. B. Joy, K. Hartmann, A. O. Mooers, The global diversity of birds in space and time. *Nature* **491**, 444-448 (2012).
70. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
71. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223 (2011).
72. S. Seabold, J. Perktold (2010) Statsmodels: Econometric and Statistical Modeling with Python. in *SciPy Conferences*.