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 Buechlein[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[1,13], Kimberly A Rosvall[1]

*correspondence to: sara.lipshutz@duke.edu or krosvall@iu.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 data collection, 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.B.B., 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 4

Abstract

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 pressure to compete for a limited breeding resource has facilitated convergent evolution in behavior, hormones, and gene expression. Using aggression assays in the field, testosterone measures, and transcriptome-wide analyses of the brain in wild-captured females and males, we examined species pairs across five avian families, each including one obligate cavity-nesting species and a related species with a more flexible nest strategy. We find 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, nor aggression. Phylogenetic analyses of individual genes and co-regulated gene networks revealed shared patterns of brain gene expression, but 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 aggression arise via a combination of both repeated and independent evolution of gene regulation.

Main

Biologists have long been fascinated by phenotypic convergence as a window into the predictability of evolution (1-3). However, our understanding of the molecular mechanisms of behavioral convergence lags behind that of morphological or physiological traits (4-6). Some studies of behavioral convergence find that evolution repeatedly uses the same (or similar) changes in particular genes or pathways (7-9). Alternatively, the building blocks of behavioral convergence may be independent, with lineage-specific mechanisms across replicated evolutionary events (10, 11). Understanding the relative contributions of these two processes requires a comparative approach that embraces the likely polygenic nature of complex behaviors (12-15). However, efforts to connect behavioral evolution to transcriptomics rarely use phylogenetic methods (6), despite the potential for changes in gene expression to shape phenotypic diversification.

Territorial aggression is a widespread behavioral trait that is well-suited for evaluating potential molecular convergence. In vertebrates, aggression is mediated by diverse neuroendocrine mechanisms (16), including the hormone testosterone (17) and its metabolite 17β-estradiol, both of which can act on sex steroid receptors in the brain to promote aggression (18, 19). Aggression is also linked to brain metabolic pathways (7) and G-protein coupled receptor signaling of dopamine, serotonin, and glutamate (20, 21). Many of these candidates are shared across distantly related species (7, 20), demonstrating the potential for conserved or repeatedly evolved mechanisms in the evolution of territorial aggression. However, there is also interspecific variation in the abundance and distribution of these endocrine-molecular building blocks (22, 23), indicating that mechanisms of aggression can diverge over time.

Mechanisms of aggression can also differ within a species, including between males and females (24), both of whom exhibit territorial aggression in many species (25, 26). 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 (27) and hormone secretion (28), which together may shape aggression or the mechanisms that promote the expression of aggression, the latter of which is still remarkably understudied in females (29, 30). Critically, in both sexes, aggression shapes access to resources and mates, and the degree of such competition varies among species (25, 26).

Competition for breeding territories is thought to be especially intense for obligate secondary cavity-nesting birds, which must secure an excavated space inside a tree or other substrate. They need a cavity to breed, they cannot excavate one themselves, and physical conflicts over cavities can lead to serious injury or death (31-33). For many cavity-nesting species, both females and males compete to acquire and defend nesting territories (34-38), and our recent case study showed elevated aggression in two species of obligate cavity-nesters, though without quantitative phylogenetic methods (39). Cavity-nesting has arisen independently over 30 times in birds (40), providing a strong foundation to evaluate the degree of mechanistic convergence in behavioral evolution.

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 - Hirundinidae (Swallows), Parulidae (Woodwarblers), Passeridae (Sparrows), Turdidae (Thrushes), and Troglodytidae (Wrens) - which each represent independent origins of obligate secondary cavity-nesting (Figure 1A; further details in SI §1) (41). Each pair included 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.

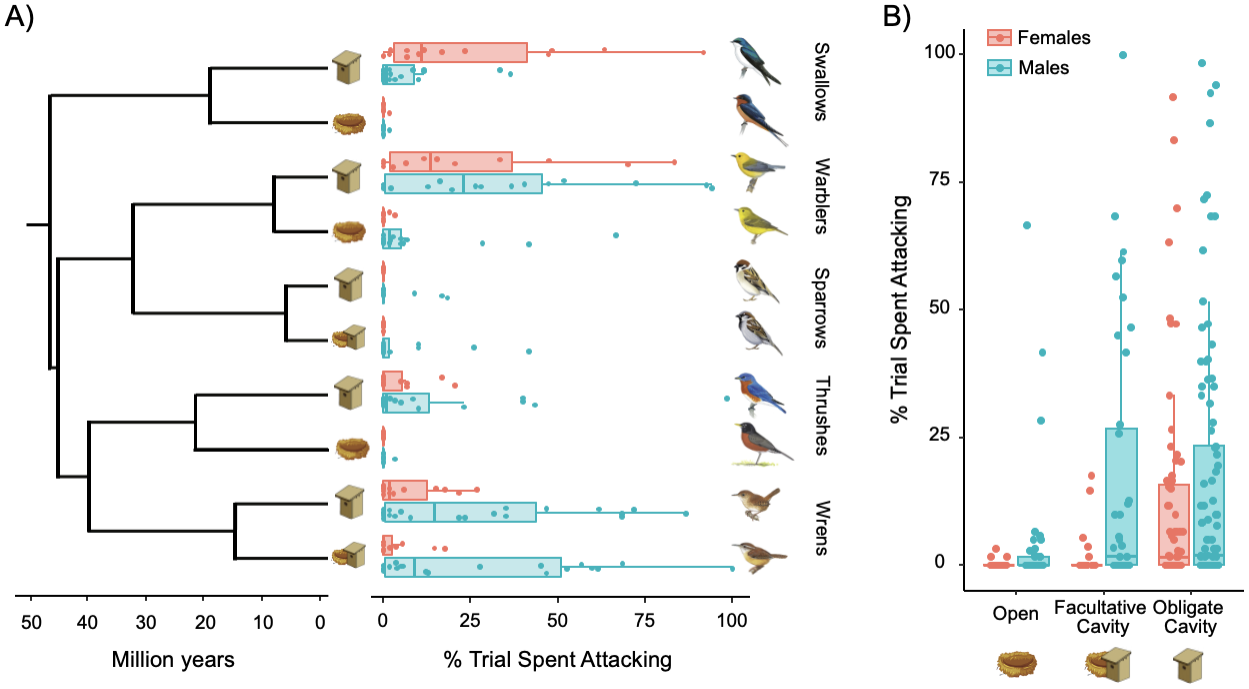

Figure 1: (A) *Left*: Consensus phylogeny of 10 species from five families, which diverged from a common ancestor ~44 mya (36-50 mya). Species pairs diverged ~9-20 mya (42). Nestboxes indicate obligate cavity-nesters, nests represent open nesters, and both symbols together represent facultative cavity-nesters. *Right*: Sex- and species-level aggression towards a conspecific decoy, measured by the proportion of 5-sec intervals that contained physical contact during a 5-min aggression assay. Species listed in descending order: Swallows: tree swallows (*Tachycineta bicolor*), barn swallows (*Hirundo rustica*), Woodwarblers: prothonotary warblers (*Protonotaria citrea*), yellow warblers (*Setophaga petechia*), Sparrows: Eurasian tree sparrows (*Passer montanus*), house sparrows (*Passer domesticus*), Thrushes: Eastern bluebirds (*Sialia sialis*), American robins (*Turdus migratorius*), and Wrens: 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

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 birds in Indiana, Illinois, or Kentucky, USA (SI §1, Table S1). Each aggression assay lasted 5 minutes 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 aggressive behaviors and focused on physical attacks (i.e. contact with the decoy). To evaluate the role of nest strategy as a driver of territorial aggression, we used 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 strategies (including both facultative cavity- and open-nesting). We also ran three-level models differentiating obligate cavity, facultative cavity, and open nest strategies (SI § 10-12).

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 more flexible nest strategies (PGLMM: $p = 0.0011$; Figure 1A; Table S2). Males attacked marginally more than females (PGLMM: $p = 0.058$). Notably, sex interacted with nest strategy (PGLMM: $p = 0.0064$), a pattern driven by elevated levels of territorial aggression in obligate cavity-nesting females (Figure 1B). We also found a significant interaction between sex of the territory holder and sex of the decoy (PGLMM: $p = 0.011$), such that male territory holders were more aggressive towards male decoys, whereas female territory holders attacked female and male decoys with similar levels of aggression (Figure S1; Table S2).

Additionally, we measured the average distance from the decoy to confirm each subject was present and engaged during a trial, regardless of whether they were aggressively attacking. *Distance from the decoy* 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).

Testosterone levels are not associated with nest strategy, nor aggression, across species

Next, we sought to measure constitutive levels of testosterone, as a potential physiological driver of behavioral convergence. We focused our analyses on baseline states that are representative of unprovoked, free-living animals (n = 196 total hormone samples; Table S1). See SI §3 and §4 for details on sample collection and enzyme immunoassays. Males had significantly higher levels of testosterone in circulation than females (PGLMM: $p < 0.0001$). Testosterone was not related to nest strategy (PGLMM: $p = 0.80$) nor the interaction between nest strategy and sex (PGLMM: $p = 0.85$; Figure S4, Table S4).

Neurogenomic mechanisms of behavioral convergence

Finally, we examined convergent evolution in brain gene expression. We focused our analysis on baseline neurogenomic states (n = 121 brain-derived RNA libraries; Table S1; SI §3). Using

RNA-seq, we measured mRNA abundance for 10,672 orthologous genes expressed in all 10 focal species in the ventromedial telencephalon; this region contains core nodes of the vertebrate social behavior network, which regulates social behavior (43). 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 three analytical approaches to understand the mechanisms underlying convergent behavioral evolution.

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

We used the Rank Rank Hypergeometric Overlap (RRHO) approach to compare the magnitude and direction of gene expression differences among species pairs. If the same genes are repeatedly targeted by selection for the obligate cavity-nesting strategy, genes should be differentially expressed in the same direction for cavity-nesters across multiple family comparisons. Indeed, RRHO revealed concordant patterns of gene expression (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, suggesting a large number of genes evolving in a similar direction of expression *within each family.*

However, this concordance did not scale *among multiple families*, indicating a lack of convergence in the identity of genes differentially expressed between species pairs. The number of shared genes decreased rapidly with increasing inclusion of family comparisons (Figure 2A). Roughly 1,000 genes (~10% of orthologs) displayed concordant expression patterns among two or three families, and 497 genes (~5%) were shared by at least 5 out of 10 family comparisons at a time (Figure 2A). However, Gene Ontology (GO) analyses of these 497 genes (Data S4, S5) revealed no significant functional enrichment. This pattern indicates that – despite a significant degree of concordance between pairs of species – concordantly expressed genes are not repeatedly recruited from any particular set of biological processes.

Across all ten family comparisons, only 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). Some of these 11 globally concordant genes have established connections with behavior or brain function (44). 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*) (44). These results indicate a high degree of independent gene expression evolution in association with the convergent evolution of obligate cavity nesting.

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. We found similarly high concordance in these sex-specific analyses, though the degree and direction of concordance differed between sexes (Figure 2A, Figure S8, S9).

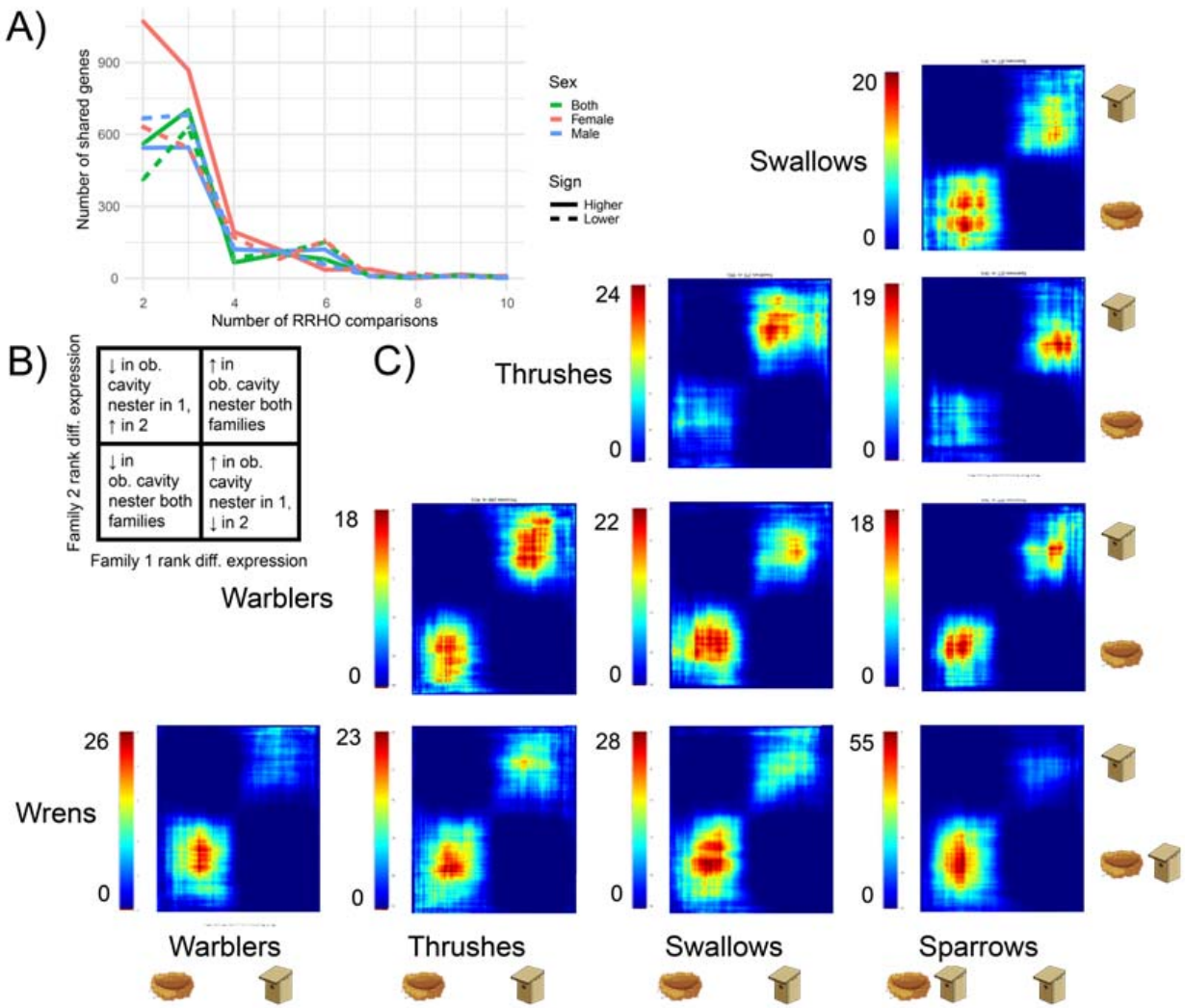

Figure 2: Concordance in differential expression based on rank-rank hypergeometric overlap (RRHO). Heatmap colors reflect adjusted hypergeometric -log(p-value). Larger values indicate more concordance in expression between families. Nestboxes indicate obligate cavity-nesters, nests represent open nesters, 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. Sex had the largest influence on expression variation and was associated with 510 genes (Table S8), the majority of which had higher expression in males than females. 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 (Table S8, Data S6).

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. 168 of 234 genes met these criteria of convergence for nest strategy (Data S8). Further restricting our analyses to all 5 family comparisons, we identified 40 genes (0.4% of orthologues) associated with nest strategy. 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, PITRM1, SLC25A24*), as well as behavior or psychiatric risk (*TRMT1L, NT5C2, BCHE*).

We also used PGLMMs to examine the expression of a subset of candidate genes related to testosterone and steroid hormone signaling in our transcriptome dataset, but we did not find a relationship with obligate cavity nesting (see §10).

Nest strategy and female aggression are associated with convergent gene networks

Complex phenotypes are often regulated by subtle but coordinated changes in gene networks (45), 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 networks of correlated genes (Data S10). Using each network'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 networks showed significant associations. Expression in the red network was higher in males than in females (Figure S11a), and 147 of 193 genes (76%) were on the zebra finch Z chromosome. Expression in the tan4 network was associated with nest strategy (Data S9, Table S9), with lower eigenvalues in each of the five independent origins of obligate cavity-nesting (Figure 3A). Though not significantly enriched for any GO terms (Table S9), several genes in the tan4 network related to mitochondrial function and energy metabolism, cognition, and stress response (Figure 3C).

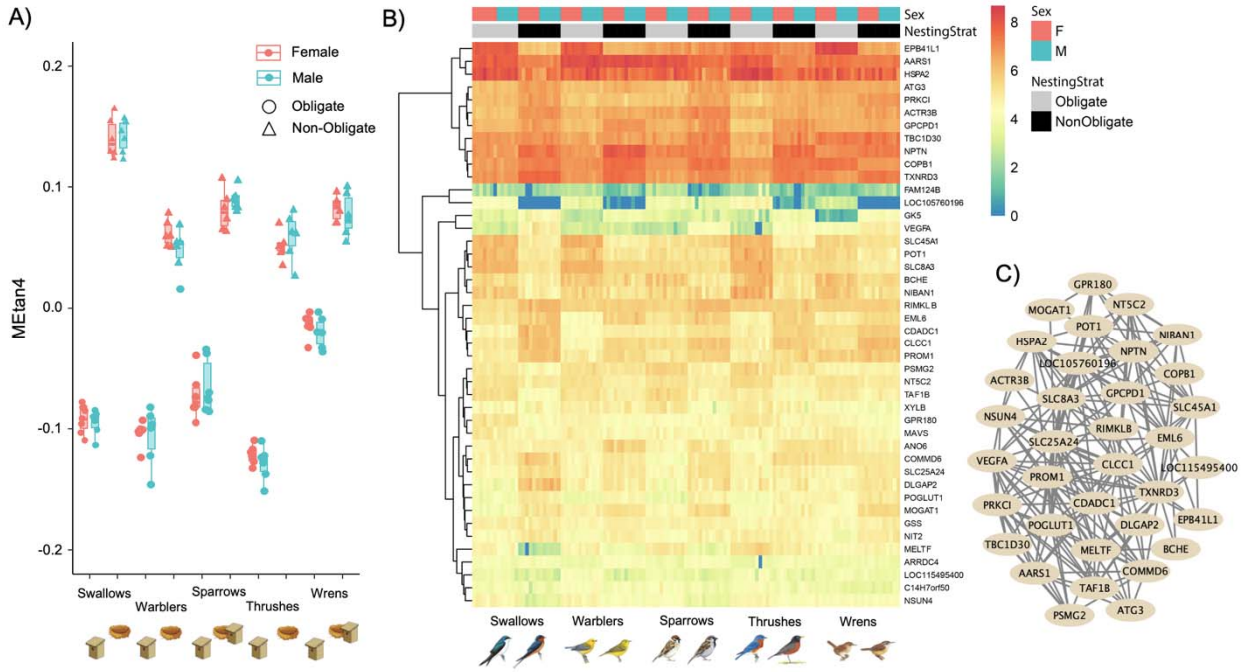

Figure 3. (A) Weighted gene co-expression network analysis (WGCNA) eigengene values for the tan4 network, which included 44 genes whose expression was significantly associated with obligate cavity-nesting relative to non-obligate nesting strategies. Individual points represent the eigengene value for 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 network. Columns represent individuals, grouped by sex and nest strategy. Color scale indicates log of gene expression. **(C)** Network depicts genes with network membership |>| 0.6. Illustrations reproduced with permission from Lynx Edicions.

Because sexes have the potential to differ in mechanisms of behavior, we repeated this analysis for each sex separately (SI §11). In this sex-specific analysis, we identified two female-specific gene networks correlated with aggression – pink4 and sienna4 (Pearson's correlation: r |>| 0.78, $p < 0.0001$, Figure 4, Table S10). Though not enriched for any GO terms, these networks contained genes with established connections to aggression, including DRD3, a dopamine receptor in the pink4 network, and GRIA2, a glutamate receptor in the sienna4 network. Finding aggression-associated networks specific to females further suggests that there are multiple gene regulatory routes to aggression, including those that may differ between sexes.

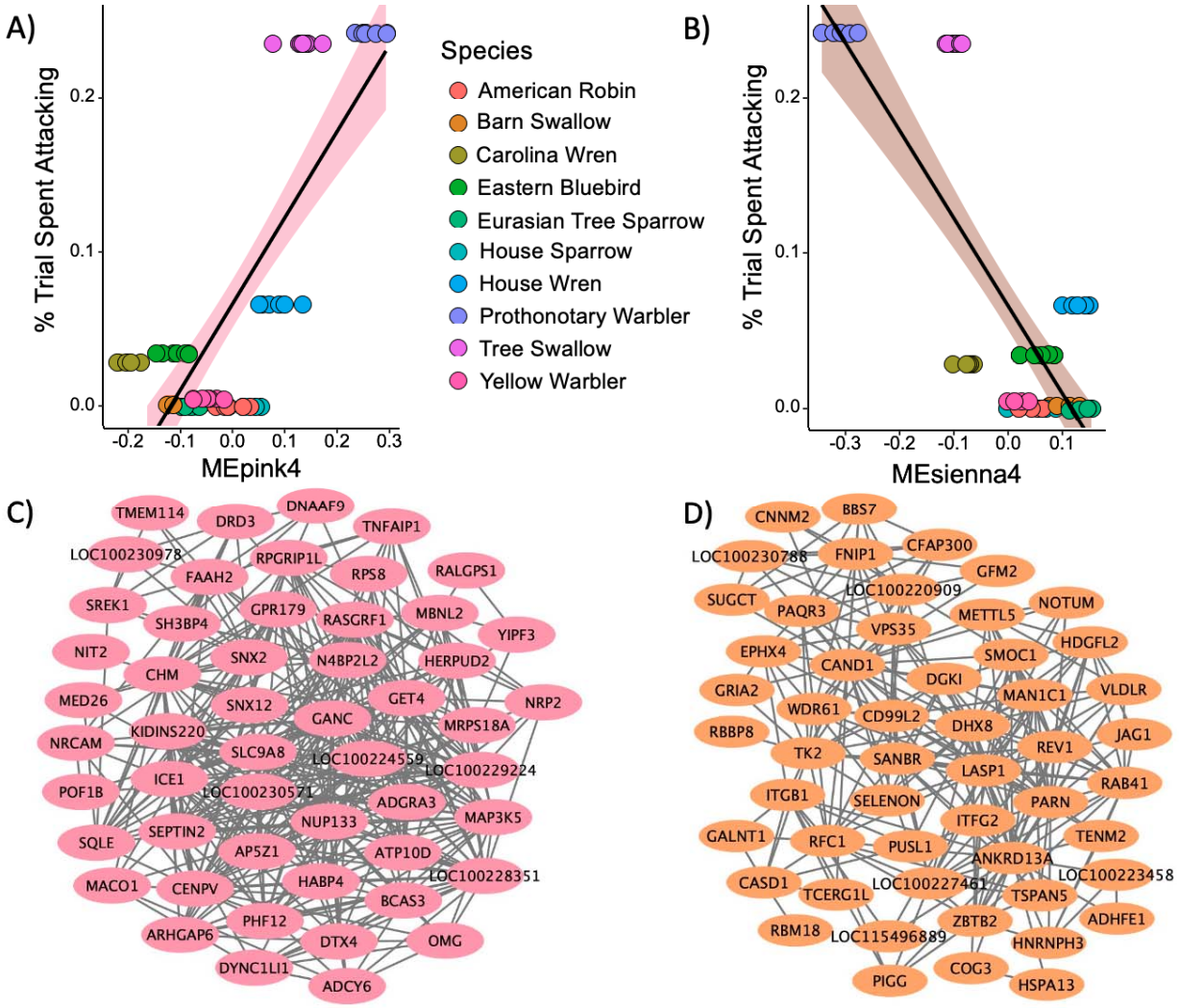

Figure 4. For females only, WGCNA network eigengenes for the **(A)** pink4 network and **(B)** sienna4 network were associated with species-level aggression towards a conspecific decoy, measured by the proportion of 5-sec intervals that contained physical contact during a 5-min aggression assay. Networks for **(C)** pink4 and **(D)** sienna4 depict genes with network membership |>| 0.6.

Gene expression provides many unique and some shared routes to convergent behavioral evolution

To identify candidate genes that are robustly (i.e. independently across multiple analyses) associated with associated with nest strategy and/or aggression, we integrated results across differentially expressed genes, individual gene- and network level PGLMMs, and concordant genes from at least 5 out of 10 RRHO comparisons (Table S12). For nest strategy-associated candidates, 27 out of 44 genes (61%) in the tan4 network recapitulated convergent individual genes in the PGLMM analysis, a strong signal that this gene set is consistently associated with the convergent evolution of obligate cavity-nesting. Out of the 221 differentially expressed genes shared among species pairs, 13 genes overlapped with nest-strategy associated individual PGLMM analyses and the tan4 network. We also identified 6 aggression-associated genes that overlapped with nest-strategy associated analyses, including 3 with concordant RRHO genes, 2 with individual PGLMM analyses, and 1 gene, *TAF1B*, which overlapped with both the tan4 network and individual PGLMM analyses. Altogether, these results suggest that

repeated selection on a small core set of genes may have contributed to the evolution of obligate cavity nesting and associated changes in behavior.

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 (7, 46) and those that are known to mediate aggression in diverse species, including humans (21, 47). Using a permutation analysis, we found significant overlap between our 79 aggression-associated genes and another study's network of 290 genes that were sensitive to experimental competition in the tree swallow brain (SI §12, Data S12, Table S13, Figure S17; (46)). 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 (44), again linking aggression to key elements of brain function.

Discussion

Striking phenotypic similarities among distantly related organisms exposed to similar ecological pressures yields critical insights into the predictability of evolution. However, investigations of the proximate mechanisms by which natural selection generates evolutionary change are often limited to traits with a simple and well-understood genetic basis (2, 3, 48), and less is known about the underpinnings of convergence in complex behavioral traits. A particularly unresolved question is the relative contributions of repeated vs. independent molecular changes to convergent behavioral evolution. Across five avian families with independent origins of obligate secondary cavity-nesting, we find behavioral convergence in territorial aggression. This convergent behavioral evolution was associated with convergent expression of a core set of genes, but not testosterone levels. These results highlight how complex behaviors can arise from a combination of repeated and independent molecular changes in response to shared ecological pressures.

Across ten species, those with obligate cavity-nesting strategies exhibited greater physical aggression, though all species were present and responsive to the simulated intruder. Our results complement a macroevolutionary study finding that cavity-nesters display more territorial behaviors against heterospecifics (49), underscoring multiple contexts linking territorial aggression with a cavity-nesting strategy. Among females, aggression was particularly high for obligate cavity nesters, consistent with the hypothesis that female aggression is adaptive during competition for nesting territories (26, 36, 39, 50). Females were aggressive towards decoys of both sexes, whereas males were more aggressive towards their own sex. Conspecific aggression may serve multiple functions, most likely nest site defense or mate guarding (26, 51), considering the timing of stimulated intrusions during the early spring period of territorial establishment before egg laying.

Despite behavioral convergence, we did not observe higher levels of testosterone in obligate cavity-nesters. Testosterone-focused hypotheses have dominated aggression-related research for decades (17), but we find that testosterone does not explain species-level differences in female or male territorial aggression, at least when birds are sampled without a prolonged social challenge at a baseline state. Recent phylogenetic analyses have found that breeding season length and mating system predict testosterone levels across species (52), but these life history traits are similar across our focal species. We previously identified a positive correlation

between territorial aggression and testosterone within female tree swallows (39), suggesting that microevolutionary processes shaping behavior and testosterone within a species do not necessarily scale to the macroevolutionary processes acting across species.

Using multiple phylogenetically informed approaches, we find that the convergent evolution of obligate cavity-nesting is associated with a small set convergently expressed genes, alongside a larger set of lineage-specific genes shared only by some families or species. These patterns could be driven by expression evolution in either the obligate-cavity or open nesters, but the latter possibility is less parsimonious given open nesting is the ancestral state among our study species. RRHO analyses revealed striking patterns of expression concordance between family comparisons, but only 0.1% of orthologs (11 genes) were associated with obligate cavity nesting across all comparisons. Single-gene and network PGLM analyses likewise revealed a small set of convergently evolving genes (0.4%; ~40 genes). These proportions are similar to a recent phylogenetic study on muscle gene expression and convergent courtship behavior in birds (14). However, they are far less convergent than most other studies of brain gene expression and behavioral evolution (9, 13, 53), which report that 4-6% of expressed orthologs are convergently evolving alongside behavior. Though our study differs by the number of taxa and/or degree of evolutionary divergence, another key difference is our explicit phylogenetic approach, which statistically eliminates shared patterns of gene expression that stem from common ancestry and more directly tests for convergence. Altogether, these results suggest that the convergent evolution of complex behavioral phenotypes can be underlain by mostly independent, lineage-specific changes in gene expression, with some convergent expression in a small, core set of genes shared across species.

This constellation of changes in gene expression may be associated with obligate cavity-nesting or aggression, and our PGLMMs disentangle these effects directly. Notably, we found limited overlap between these two gene sets. Coupled with a larger set of genes showing an interaction between aggression and cavity-nesting, these patterns suggest that there are many neuromolecular correlates of aggression, including many that differ among species. Expanding beyond our study, we found significant gene overlap with our aggression-associated genes and a previously published socially-sensitive gene network (46), indicating concordance between short-term functional responses to competition and long-term evolutionary change in competitive environments. We also identified two female-specific gene networks that correlated with aggression. These sex-specific transcriptomic correlations with aggression, in tandem with our finding that female obligate cavity nesters are particularly aggressive, suggest that female aggression has evolved differently from male aggression in obligate cavity nesting species, again underscoring multiple regulatory routes to behavioral convergence.

We also find a pattern of sex-biased gene expression, for which males had higher expression that females. In birds, males carry two copies of the Z chromosome, and dosage compensation of the Z chromosome may be incomplete or absent (54), leading to higher expression of Z-linked genes in males than females. This mechanism likely explains the WGCNA patterns for the red network, which had higher expression in males than in females. An alternative, but not mutually exclusive, hypothesis is that the upward expression bias in male-biased genes could stem from stronger sexual selection on males, combined with reduced constraint on the evolution of higher expression (55).

Convergent evolution provides a natural experiment to identify the genomic underpinnings of complex phenotypes. Though future experiments are needed to functionally validate the aggression-associated patterns we identify, our work nevertheless highlights the power of comparative approaches to uncover the mechanistic bases of convergence (13, 14, 20). Notably, our study is among the first to use phylogenetically explicit models of gene expression to understand the evolution of a complex and continuously varying behavioral trait. Our findings support the hypothesis that there are diverse evolutionary routes to building an aggressive bird, layered atop a small set of convergently evolving genes.

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 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 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). 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).

Brain dissection and RNA isolation

For 6 to 7 individuals per sex per species, we dissected whole brains into functional regions following (56). We removed the cerebellum, hindbrain, optic chiasm, optic tecta, and hypothalamus to the depth of the anterior commissure. Gene expression analyses focused on the ventromedial telencephalon (VmT), which contains nodes in the vertebrate social behavior network (43) including the extended medial amygdala (TnA), bed nucleus of the stria terminalis (BNST), and lateral septum (LS). We collected this region by removing ~1mm of the ventromedial portion of the caudal telencephalon. 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 (57) 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) (58) 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 (59), 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) (60) (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 (61). 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 network 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 network membership value >0.6 were assessed for enrichment of biological processes in PantherGO (62), using our list of 10,672 orthologs as a reference set. Networks of interest were visualized in Cytoscape v3.10.1 (63). 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 (64), 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 (65) resource to obtain 1000 ultrametric trees from Ericson All Species. We constructed a majority rule consensus topology from this tree set using the *ape* (66) function 'consensus', and inferred branch lengths using the *phytools* (67) 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 (68) 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 (44) 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 (62), 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 (7, 21, 46, 47) (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, Cara 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, Loyola 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. Arendt, D. Reznick, Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? *Trends Ecol Evol* **23**, 26-32 (2008).
4. J. R. Gallant, L. A. O'Connell, Studying convergent evolution to relate genotype to behavioral phenotype. *J Exp Biol* **223** (2020).
5. N. Jourjine, H. E. Hoekstra, Expanding evolutionary neuroscience: insights from comparing variation in behavior. *Neuron* **109**, 1084-1099 (2021).
6. K. A. Rosvall, Evolutionary endocrinology and the problem of Darwin's tangled bank. *Horm Behav* **146**, 105246 (2022).
7. 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).
8. 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).
9. A. R. Pfenning *et al.*, Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* **346**, 1256846 (2014).
10. A. S. Mikheyev, T. A. Linksvayer, Genes associated with ant social behavior show distinct transcriptional and evolutionary patterns. *Elife* **4**, e04775 (2015).
11. 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).
12. E. A. Boyle, Y. I. Li, J. K. Pritchard, An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177-1186 (2017).
13. R. L. Young *et al.*, Conserved transcriptomic profiles underpin monogamy across vertebrates. *Proc Natl Acad Sci U S A* **116**, 1331-1336 (2019).
14. 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).
15. M. C. Harrison *et al.*, Hemimetabolous genomes reveal molecular basis of termite eusociality. *Nat Ecol Evol* **2**, 557-566 (2018).
16. R. J. Nelson, B. C. Trainor, Neural mechanisms of aggression. *Nat Rev Neurosci* **8**, 536-546 (2007).
17. 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).
18. 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).
19. 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).
20. M. C. Saul *et al.*, Cross-species systems analysis of evolutionary toolkits of neurogenomic response to social challenge. *Genes Brain Behav* **18**, e12502 (2019).
21. 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).
22. A. M. Kelly, L. C. Wilson, Aggression: Perspectives from social and systems neuroscience. *Horm Behav* **123**, 104523 (2020).
23. L. A. O'Connell, H. A. Hofmann, Evolution of a vertebrate social decision-making network. *Science* **336** (2012).

24. 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).
25. T. Clutton-Brock, Sexual Selection in Males and Females. *Science* **318**, 1882-1885 (2007).
26. K. A. Rosvall, Intrasexual competition in females: evidence for sexual selection? *Behav Ecol* **22**, 1131-1140 (2011).
27. A. E. Wright, J. E. Mank, The scope and strength of sex-specific selection in genome evolution. *J Evol Biol* **26**, 1841-1853 (2013).
28. 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).
29. 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).
30. 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).
31. R. A. Duckworth, Adaptive dispersal strategies and the dynamics of a range expansion. *Am Nat* **172 Suppl 1**, S4-17 (2008).
32. D. Leffelaar, R. J. Robertson, Nest usurpation and female competition for breeding opportunities by Tree Swallows. *The Wilson Bulletin* **97**, 221-224 (1985).
33. J. Merilä , D. A. Wiggins, Interspecific Competition for Nest Holes Causes Adult Mortality in the Collared Flycatcher. *The Condor* **97**, 445-450 (1995).
34. 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).
35. C. A. Krieg, T. Getty, Fitness benefits to intrasexual aggression in female house wrens, Troglodytes aedon. *Animal Behaviour* **160**, 79-90 (2020).
36. K. A. Rosvall, Sexual selection on aggressiveness in females: evidence from an experimental test with tree swallows. *Animal Behaviour* **75**, 1603-1610 (2008).
37. D. R. Rubenstein, I. J. Lovette, Reproductive skew and selection on female ornamentation in social species. *Nature* **462**, 786-789 (2009).
38. 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).
39. 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).
40. Anonymous, (!!! INVALID CITATION !!! (34)).
41. 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).
42. 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).
43. J. L. Goodson, The vertebrate social behavior network: evolutionary themes and variations. *Horm Behav* **48**, 11-22 (2005).
44. 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).
45. 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).
46. 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).

47. 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).
48. J. F. Storz, Causes of molecular convergence and parallelism in protein evolution. *Nat Rev Genet* **17**, 239-250 (2016).
49. 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).
50. M. J. West-Eberhard, Sexual selection, social competition, and speciation. *The Quarterly Review of Biology* **58**, 155-183 (1983).
51. A. M. Harts, I. Booksmythe, M. D. Jennions, Mate guarding and frequent copulation in birds: A meta-analysis of their relationship to paternity and male phenotype. *Evolution* **70**, 2789-2808 (2016).
52. J. F. Husak *et al.*, Life history and environment predict variation in testosterone across vertebrates. *Evolution* **75**, 1003-1010 (2021).
53. S. C. P. Renn *et al.*, Gene expression signatures of mating system evolution. *Genome* **61**, 287-297 (2018).
54. 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).
55. S. Grath, J. Parsch, Sex-Biased Gene Expression. *Annu Rev Genet* **50**, 29-44 (2016).
56. 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).
57. M. G. Grabherr *et al.*, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology* **29**, 644-652 (2011).
58. A. Rhie *et al.*, Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737-746 (2021).
59. 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).
60. 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).
61. P. Langfelder, S. Horvath, WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
62. 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).
63. P. Shannon *et al.*, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).
64. J. D. Hadfield, MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software* **33**, 1-22 (2010).
65. 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).
66. E. Paradis, K. Schliep, ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528 (2019).
67. L. J. Revell, phytools: an R package for phylogenetic comparative biology (and other things). *Methods in Ecology and Evolution* **3**, 217-223 (2011).
68. S. Seabold, J. Perktold (2010) Statsmodels: Econometric and Statistical Modeling with Python. in *SciPy Conferences*.