Conserved transcriptomic profiles underpin monogamy across vertebrates

Rebecca L. Young1,2,3, Michael H. Ferkin4, Nina F. Ockendon-Powell5, Veronica N. Orr5, Steven M. Phelps6,7, Ákos Pogány3, Corinne L. Richards-Zawacki1, Kyle Summers1, Tamás Székely8,9, Brian C. Trainor8, Araxi O. Urrutia10,11, Gergely Zachar3, Lauren A. O’Connell12, and Hans A. Hofmann1,2,3,13

1Department of Integrative Biology, The University of Texas at Austin, Austin, TX 78712; 2Center for Computational Biology and Bioinformatics, The University of Texas at Austin, Austin, TX 78712; 3Department of Biological Sciences, The University of Memphis, Memphis, TN 38111; 4Life Sciences, University of Bristol, Bristol BS8 1TQ, United Kingdom; 5Department of Psychology, University of California, Davis, CA 95616; 6Institute for Cellular and Molecular Biology, The University of Texas at Austin, Austin, TX 78712; 7Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712; 8Department of Ethology, Eötvös Loránd University, Budapest H-1117, Hungary; 9Department of Biological Sciences, University of Pittsburgh, Pittsburgh, PA 15260; 10Department of Biology, East Carolina University, Greenville, NC 27858; 11Milner Centre for Evolution, University of Bath, Bath BA2 7AY, United Kingdom; 12Centre for Networks and Collective Behaviour, University of Bath, Bath BA2 7AY, United Kingdom; 13Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad de México 04510, Mexico; 14Department of Anatomy, Histology, and Embryology, Semmelweis University, Budapest H-1094, Hungary; and 15Department of Biology, Stanford University, Stanford, CA 94305

Edited by Gene E. Robinson, University of Illinois at Urbana–Champaign, Urbana, IL, and approved November 26, 2018 (received for review August 14, 2018)

Social monogamy, typically characterized by the formation of a pair bond, increased territorial defense, and often biparental care, has independently evolved multiple times in animals. Despite the independent evolutionary origins of monogamous mating systems, several homologous brain regions and neuroendocrine systems, several homologous brain regions and neuropeptides and their receptors have been shown to play a conserved role in regulating social affiliation and parental care, but little is known about the neuromolecular mechanisms underlying monogamy on a genomic scale. Here, we compare neural transcriptomes of reproductive males in monogamous and nonmonogamous species pairs of Peromyscus mice, Microtus voles, parid songbirds, dendrobatid frogs, and Xenotilapia species of cichlid fishes. We find that, while evolutionary divergence time between species or clades did not explain gene expression similarity, characteristics of the mating system correlated with neural gene expression patterns, and neural gene expression varied concordantly across vertebrates when species transition to monogamy. Our study provides evidence of a universal transcriptomic mechanism underlying the evolution of monogamy in vertebrates.


The diversity of animal social behavior has motivated a wealth of studies that explore variation in behavioral repertoires, sensory and cognitive specializations, and the ecological contexts in which they have evolved. Despite this extensive variation, the action of hormones, specifically sex steroids and neuropeptides (1), and other candidate pathways appears to be remarkably conserved in the regulation of social behavior (e.g., refs. 2 and 3). Moreover, recent studies support the intriguing hypothesis that coordinated activity of conserved gene sets underlies independent evolutionary transitions to similar behavioral phenotypes (4–8). It should thus not be a surprise that behavioral phenotypes may share molecular mechanisms regardless of their evolutionary history. Like extant animals, the most recent common ancestor had to meet challenges imposed by fluctuating internal and external conditions. The mechanisms used by these ancestral organisms to maintain homeostasis serve as the building blocks for the evolution of more derived behavioral responses as evidenced by the conserved role of homologous brain regions in processing social signals (9–11). At the molecular level, a “toolkit” of molecular pathways and gene networks can be preserved for hundreds of millions of years (12), and phenotypic novelty often can be attributed to new uses of such conserved gene sets (13, 14). The pervasiveness of conserved gene modules is highlighted by phenologs—functionally and physiologically unrelated phenotypes in different species with a statistical overrepresentation of shared sets of underlying orthologous genes (15). Finally, recent progress resolving evolutionary relationships among metazoan animals indicates that homoplasy is much more common than previously appreciated (16), even among phenotypes with overlapping molecular mechanisms [e.g., the nervous systems (17, 18)]. These discoveries have transformed our thinking about the origins and evolution of morphological and developmental phenotypes, but are rarely applied to investigations of the evolution of behavior. Uncovering universal mechanisms of similar phenotypes requires a broadly comparative approach (19–21). Here, we ask to what extent similar neural transcriptomic profiles are associated with variation in social behavior across vertebrates, using mating system evolution as an example, and discuss the importance of...
our discovery for understanding the origins of behavioral diversity. While a more narrow focus within a clade may reveal more candidate genes with similar expression (6, 22), these results cannot be generalized across clades, limiting their broader implications. Monogamous species with nonmonogamous close relatives can be found in at least four major vertebrate clades (teleosts, amphibians, birds, mammals), providing an unparalleled opportunity to examine whether repeated transitions to a particular mating system evolve via shared transcriptomic mechanisms. We compare neural transcriptomes of reproductive males in monogamous and nonmonogamous species pairs of Peromyscus mice, Microtus voles, passeroid songbirds, dendrobatid frogs, and ectodine cichlid fishes. We characterize similarity in neural gene expression patterns among species in relation to similarity in mating system, ecological attributes, and evolutionary divergence time. While neither similarity in ecology nor divergence time between clades explained gene expression similarity, neural gene expression varied concordantly across vertebrates between males of monogamous and nonmonogamous species. Genes with highly increased or decreased expression in monogamous species of one clade are likely to also have highly increased or decreased expression in the monogamous species of another clade. Our study provides evidence of a universal transcriptomic mechanism underlying monogamy in vertebrates.

Animal mating systems can be characterized as a suite of reproductive, parental, and agonistic phenotypes each can be highly variable among closely related (and even within) species depending on sex ratio as well as ecological factors such as predation risk, resource distribution, and extent of competition (23–25). Despite this potential diversity, similar mating systems have been described in numerous distantly related species. Social monogamy, for example, is typically characterized by the formation of a pair bond, increased territorial defense, and often biparental care; this suite of social behaviors has evolved independently numerous times (23, 26). Studies of pair bonding and parental care in mammals, birds, and fishes reveal a conserved, albeit complex, role for arginine vasopressin (AVP) and oxytocin (OT) as well as their receptors in regulating social affiliation (27–31). Recent studies have advanced our understanding of variation and the evolution of such pathways, for example, by illustrating how life history trade-offs underlie molecular, transcriptomic, and epigenetic variation among individuals (32). Less attention has been given to characterizing the complexities of neuromolecular mechanisms underlying monogamy on a genomic scale, including identifying novel candidate genes and pathways.

**Results and Discussion**

**Shared and Unique Patterns Across Clades.** Within each clade we compared expression of orthologous gene groups (OGGs) between species pairs (SI Appendix, Fig. S3). We found that the mean difference in expression between species pairs was near zero across all 1,979 OGGs (SI Appendix, Fig. S3A). Intriguingly, gene expression was least variable between the two bird species (SI Appendix, Fig. S3B), which may be explained by the fact that the mating systems of these species pairs are considerably more similar than those of the other clades (Fig. 2). Specifically, both Anthus spinolleta and Prunella modularis can form pair bonds (although less frequently in P. modularis) and exhibit direct paternal care (Fig. 2 and SI Appendix, Fig. S1 and Table S1). Despite these differences in OGG expression variation across clades, clade-specific Gene Ontology (GO) analysis revealed high conservation of GO term enrichment highlighting cell communication, signaling receptor activity, and membrane proteins as consistently associated with monogamy-related expression across clades (SI Appendix, Fig. S4).

**Across Vertebrates, Gene Expression Varies Concordantly Between Males of Monogamous and Nonmonogamous Species.** To assess concordance of OGG expression variation between monogamous and nonmonogamous species across vertebrate clades, we used the differential expression analysis software package DESeq2 (33). We assessed differential OGG expression multiple times with distinct evolutionary groupings from mammals to all clades where monogamous species of distinct clades were entered as interspecific replicates of monogamy (Fig. 3). We find that differences in OGG expression are generally concordant, particularly among OGGs that exhibit larger fold-differences (Fig. 3). In addition, most OGGs maintain directional concordance. For example, many of the OGGs that show increased expression (positive log2 fold-difference) in one evolutionary group (e.g., mammals) show increased expression in other evolutionary groups (Fig. 3). These results indicate that across vertebrates monogamous species recruit a common set of OGGs despite evolutionarily independent transitions to similar mating systems. We find that, as the evolutionary frame of reference is expanded and more distantly related clades are added to the analysis, fewer OGGs retain significance. In particular, we find a large decrease in above-threshold OGGs when amniote species pairs (i.e., mamalian and avian species pairs) are included compared with only the mammalian species pairs (Fig. 3). This finding likely reflects the decreased expression variation notable in our bird species pair (SI Appendix, Fig. S3). While the observed effect of the mating system for some OGGs may be smaller at broader taxonomic scales, adding species pairs increases the statistical power. Thus, with the exception of the comparison between mammals and birds described above, the decrease in number of OGGs meeting our threshold cutoff is quite small (Fig. 3).

When we included all species pairs, we identified 123 OGGs (6.2%) associated with monogamy across vertebrates (SI Appendix, Fig. S5). We find a number of OGGs significant at one level of analysis failing to meet the significance threshold at another (Fig. 3). Many differential expression analysis approaches, including DESeq2, rely on expression variation among biological replicates to determine differential expression. This approach is limiting when biological replicates are highly variable (34) as is the case here, where species of different clades are included as interspecific replicates of monogamy. In addition to evolutionary distance, a number of biological and technical features likely generate noise in our analysis (e.g., ecological differences among species, course-grained tissue sampling, and technical variation during sequencing). To identify conserved transcriptomic patterns across monogamous species and assess whether the degree of overlap in expression between monogamous species is statistically significant, we utilize the Rank-Rank Hypergeometric Overlap (RRHO) approach (34). Comparing ranked fold-differences enables discovery of OGGs that share patterns of expression among monogamous species without requiring that expression values be similar across evolutionarily distant clades. Additionally, rather than adhering to discrete thresholds to identify candidates with similar expression, RRHO identifies candidates with coordinated directional shifts in expression using a sliding threshold (34) (Fig. 4).

Using RRHO including all 1,979 OGGs, we find an enrichment of OOG overlap in the on-diagonal extremes (i.e., at high log2 fold-differences in expression between monogamous and nonmonogamous species in both clades being compared; Fig. 4D). Most notably, we find enriched overlap of OGGs exhibiting decreased expression in monogamous species of all clades. In all comparisons, the concordant down-regulated (i.e., bottom left) quadrant of the RRHO plots has the greatest overall significance (i.e., most significant quadrant mean $-\log_{10} P$ value), strongly indicative of a universal signature of expression among monogamous vertebrates (Fig. 4 C and D). We generally do not find enrichment at small log2 fold-differences with a few exceptions (e.g., frogs and voles). This observation is likely due to the fact that the majority of OGGs exhibit small fold-differences (although they may be important for monogamy-related behavior). Subtle variation in expression of genes involved in clade-specific
or basic cellular functions could mask potentially important similarities at this level. A consequence of this limitation of the RRHO approach is that, for OGGs with small fold-differences, a large overlap in OGGs is required to reach significant enrichment. Interestingly, clade comparisons including frogs exhibit the most enrichment in the off-diagonal (i.e., discordant) regions. Specifically, many orthologous gene groups that show decreased expression in other groups increase in the monogamous frogs, a finding consistent with the general trend of increased gene expression in the monogamous frogs (SI Appendix, Fig. S3A). This pattern is clear in all pairwise comparisons involving frogs, suggesting that the monogamy-related expression patterns in frogs are the least similar to the other clades. Comparisons including birds show the least enrichment in OGG overlap, possibly because the mating systems of the bird species in our study are relatively similar to each other, as discussed above. Importantly, however, birds still share the same overall pattern of OGG overlap in the on-diagonal extremes and show strong similarities in expression patterns in OGGs with high differential expression (e.g., Bottom Left and Top Right quadrants of birds vs. mice and birds vs. fish, Fig. 4D). Variation among clades in overall and specific patterns of OGG enrichment may reflect a difference among clades in species-specific features of monogamy. For example, male parental care is ancestral in the poison frog clade (35). Thus, biparental care associated with monogamy in our focal species evolved from male-biased care in our frog clade rather than from female-biased care as in the other clades examined here. Future analyses—including species exhibiting some but not all reproductive, parental, and agonistic phenotypes of monogamy as well as a more anatomically focused analysis—may shed light on this variation, enabling decoupling and association of gene expression patterns with more specific behavioral phenotypes.

**Discovery of Candidates Underlying Monogamy.** We identified 24 candidate genes that are robustly (i.e., independently in both differential expression and RRHO analyses) associated with monogamous mating systems across vertebrates (Fig. 5, SI Appendix, Table S7 and Dataset S1). Interestingly, across lineages, genes involved in neural development, cell–cell signaling, synaptic activity (e.g., GRM6), learning and memory (e.g., DSCAM), and cognitive function are up-regulated in monogamous males. Conversely, genes involved in gene transcription (e.g., genes associated with RNA Polymerase II) and AMPA receptor regulation are down-regulated. These differences might indicate increased neural plasticity in the face of tighter transcriptional regulation in monogamous males. A more detailed discussion of these candidate genes, including a description of neurolocalization, is provided in SI Appendix, Table S7 and Dataset S1. While the majority of the candidate OGGs identified by both approaches to differential expression analysis (DESeq2 and RRHO) show increased expression in monogamous males, the greatest overlap enrichment occurs in OGGs with decreased expression in monogamous males (Fig. 4). Many of these OGGs do not survive threshold-dependent DESeq2 analysis and thus are eliminated from our list of robust candidates. Traditional candidates associated with monogamy-related behavior (e.g., AVP and OT as well as their receptors) were not identified among those consistently expressed in monogamous males, and in some species were not identified in the RNA-seq analysis (SI Appendix, Fig. S6). This finding does not indicate that these traditional candidates are unimportant in regulating monogamy-related behavior across these clades, rather that their roles (if they are conserved across monogamous species) are not reflected in a consistent expression pattern at the coarse neuro-anatomical scale that we sampled.

**Phylogenetic and Ecological Correlates.** The combined discovery of monogamy-related OGG expression by differential expression analysis (Fig. 3) and enrichment of gene expression rank correlation among clades (Fig. 4) supports the hypothesis that shared
gene expression patterns underlie the behavioral expression of monogamous mating systems across vertebrate clades. However, it is possible that historical and ecological features shared between clades influence similarity in neural gene expression patterns. By design, species pairs were quite similar in their ecological attributes, differing primarily in specific characteristics of their mating systems (Fig. 2 and SI Appendix, Fig. S1 and Tables S1 and S2). Thus, the selection of species pairs by clade should minimize confounding ecological factors; however, several other factors may play a role in gene expression similarity. First, divergence times between species pairs vary between ~2.5 and 34 million years (36). Historical contingency can bias the path of evolution such that more closely related species may be more similar due to shared evolutionary history. Second, elaboration of mating systems varies among species such that comparisons between monogamous and nonmonogamous mating systems are not equivalent across the clades. For example, the bird species included in this study share a number of mating system characteristics (Fig. 2). To assess the role of evolutionary history and mating system on gene expression divergence, we compared evolutionary and mating system distances to OGG expression distance for all species pairs. Neither evolutionary distance nor mating system distance correlated with OGG expression divergence between species pairs (Fig. 6 A and B). Notably, however, we find that the birds and the frogs are the most similar in their respective transcriptomes and are also the most similar in characteristics of mating system, while at the same time being the most distantly related of all species pairs (birds: 29 Mya; frogs: 34.2 Mya; Fig. 1 and SI Appendix, Fig. S3). When expression and mating system variation attributable to phylogeny is removed (using phylogenetic independent contrasts), we find a significant relationship between neural gene expression and mating system (Fig. 6C). Even though phylogenetic relatedness and ecological attributes affect neural transcriptome similarity across species in complex ways, together these observations indicate a critical role for mating system in driving gene expression similarity in the brain.

Conclusions. Using a comparative transcriptomics approach, we asked whether independent transitions to a monogamous mating system across four major clades of vertebrates are associated with shared neural gene expression patterns. A shared mechanistic basis of social behavior across distantly related clades has been documented at the level of neural circuitry where brain-region-specific expression of neurochemical genes is remarkably conserved in the Social Decision Making Network of the vertebrate fore- and midbrain (37). Further neural gene expression comparisons of aggressive behavior in bees, stickleback, and mice (4) provide
and some mechanisms evolve de novo (38–40). We argue that a conceptual shift is critical for achieving progress toward testing mechanistic hypotheses of the evolution of complex social behavior using comparative ‘omic approaches (as has been suggested for morphological traits; cf. ref. 41). Because the preponderance of evidence supports ancient, deeply homologous molecular and developmental pathways in evolutionary diversification, similar phenotypes should be expected to have mechanistic similarities regardless of evolutionary origins. Under this model, independent evolutionary transitions to similar behavioral phenotypes are expected to show similar mechanisms as a consequence of (i) integration of new traits with shared, ancestral physiological processes and their associated developmental and genetic pathways; and (ii) similar challenges to basic physiology that organisms experience in the evolution of these similar phenotypes. Whereas previous accounts of transcriptomic similarities among evolutionarily distinct complex behavior have focused on gene expression similarities, this conceptual approach provides a context to leverage both similarities and differences in gene expression to understand the mechanistic basis of complex trait evolution. Overall, our results indicate that independent evolutionary transitions to a monogamous mating system across vertebrate clades are accompanied by similar changes in gene expression in the male fore- and midbrain. In combination with the discovery of shared transcriptomic mechanisms associated with aggression (4), learned vocalizations in birds and humans support for the hypothesis that broad similarities in social behavior across species may result from independent recruitment of ancient gene modules. However, inference from this analysis was limited by the complexity of homology inference among brains in bees and brain regions sampled in mice and stickleback. We expand on this previous comparative neural gene expression study of social behavior by comparing equivalent brain sections and providing a phylogenetically informed assessment of transcriptome similarity in monogamous species of distinct and distantly related clades. Furthermore, by carefully scoring mating system characteristics and ecological attributes and by accounting for phylogenetic nonindependence, we test alternative hypotheses underlying the origins of transcriptomic similarities that originated in independent evolutionary transitions to monogamy across vertebrates. While our sampling of whole fore- and mid-brains provides a coarse-grained view of the neuromolecular correlates of mating system, this approach circumvents potential problems with the sometimes tenuous homology inference of brain regions, ensuring consistent sampling across clades.

Our analysis reveals that monogamous species of distantly related vertebrate clades share more gene expression similarities than would be expected by chance, particularly in OGGs that are highly differentially expressed in monogamous species. Shared OGG expression across these clades provides evidence of deeply homologous mechanisms underlying the suite of reproductive, parental, and agonistic behavioral phenotypes that accompany monogamous mating systems. Our discovery adds to the growing body of research highlighting shared mechanistic bases of independently evolved morphological and physiological phenotypes (12). The present work illustrates that the independent evolution of complex behavioral phenotypes, like monogamy, is better conceptualized as the product of both parallel and convergent processes where many components of the underlying mechanism arise via parallel recruitment of deeply shared gene networks and some mechanisms evolve de novo (38–40). We argue that a conceptual shift is critical for achieving progress toward testing mechanistic hypotheses of the evolution of complex social behavior using comparative ‘omic approaches (as has been suggested for morphological traits; cf. ref. 41). Because the preponderance of evidence supports ancient, deeply homologous molecular and developmental pathways in evolutionary diversification, similar phenotypes should be expected to have mechanistic similarities regardless of evolutionary origins. Under this model, independent evolutionary transitions to similar behavioral phenotypes are expected to show similar mechanisms as a consequence of (i) integration of new traits with shared, ancestral physiological processes and their associated developmental and genetic pathways; and (ii) similar challenges to basic physiology that organisms experience in the evolution of these similar phenotypes. Whereas previous accounts of transcriptomic similarities among evolutionarily distinct complex behavior have focused on gene expression similarities, this conceptual approach provides a context to leverage both similarities and differences in gene expression to understand the mechanistic basis of complex trait evolution. Overall, our results indicate that independent evolutionary transitions to a monogamous mating system across vertebrate clades are accompanied by similar changes in gene expression in the male fore- and midbrain. In combination with the discovery of shared transcriptomic mechanisms associated with aggression (4), learned vocalizations in birds and humans
Expression Omnibus (42). Procedures for sample collection are detailed in SI Appendix. These selected species pairs differ in mating system characteristics, but are similar in other ecological attributes (Fig. 2; SI Appendix, Fig. S1 and Tables S1 and S2; and ref. 43).

One challenge associated with comparative analysis of gene expression patterns across distantly related species is identifying homologous tissues and comparable orthologous genes. To limit the requirement of brain region homology inference across distantly related clades, we extracted RNA from the combined fore- and midbrain tissues after hindbrain removal. To improve rather than individual genes. Across our 10 species we identified 1,979 OGGs when an OGG contained more than one gene (as in ref. 4). Genes in the same OGGs were generally concordant in the selection of the most differentially expressed paralog did not obscure the overall similarity in expression pattern and allowed for downstream analysis of candidate genes. Thus, for each OGG and each clade the gene with the largest expression difference between the monogamous and nonmonogamous species was selected as the representative gene (Dataset S2 and ref. 45). For brevity, we refer to expression of this representative gene as OGG expression.

ACKNOWLEDGMENTS. We thank A. Ball, R. Harris, and R. Kar for assistance with the research; A. Battenhouse, B. Goetz, and E. Ortego (Center for Computational Biology and Bioinformatics, University of Texas at Austin), C. Jordan, and the Texas Advanced Computing Center (University of Texas Austin) for technical support and D. Greer and members of the H.A.H. laboratory for discussion and helpful comments on earlier versions of this manuscript. This work was supported by the Alfred P. Sloan Foundation (BR-4900); NSF Grants IOS-1354942, IOS-1501704, and IOS-1601734 (to H.A.H.); NIH Grant R01 MH58692-52 (to B.C.T.); and the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (to A.P.).

5. Pfenning AR, et al. (2016) Comparative transcriptomics reveals the conserved building blocks of complex brain development: Insights from reproductive males of closely related monogamous and nonmonogamous species pairs used for the remainder of the analysis (as in ref. 4). Genes in the same OGGs were generally concordant in directionality of expression difference (SI Appendix, Fig. S2 and Table S4). Thus, the selection of the most differentially expressed paralog did not obscure the overall similarity in expression pattern and allowed for downstream analysis of candidate genes. Thus, for each OGG and each clade the gene with the largest expression difference between the monogamous and nonmonogamous species was selected as the representative gene (Dataset S2 and ref. 45).