Gene Expression Signatures of Mating System Evolution

Journal:	<i>Genome</i>
Manuscript ID	gen-2017-0075.R1
Manuscript Type:	Article
Date Submitted by the Author:	18-Jul-2017
Complete List of Authors:	Renn, Suzy; Reed College, Biology Machado, Heather; Wellcome Trust Sanger Institute, Wellcome Genome Campus Unterwaditzer, Nina; University of Texas, Austin, Department of Integrative Biology Harris, Rayna; University of Texas, Austin, Department of Integrative Biology; University of Texas at Austin, Institute for Cellular and Molecular Biology Sessa, Anna; University of Texas, Austin, Department of Integrative Biology Hofmann, Hans; University of Texas at Austin, Department of Integrative Biology; University of Texas at Austin, Institute for Cellular and Molecular Biology; University of Texas at Austin, Center for Computational Biology and Bioinformatics
Is the invited manuscript for consideration in a Special Issue? :	Ecological Genomics
Keyword:	mating system, sociogenomics, transcriptomics, epigenetics
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
gen-2017-0075.R1Suppla.xls	

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35

Gene Expression Signatures of Mating System Evolution

Susan C.P. Renn¹, Heather E. Machado^{1,2}, Nina Duftner³, Anna K. Sessa³, Rayna M.
Harris^{3,4}, Hans A. Hofmann^{3,4,5}

¹Department of Biology, Reed College

²Department of Biology, Stanford University

³Department of Integrative Biology, The University of Texas at Austin

⁴Institute for Cellular and Molecular Biology, The University of Texas at Austin

⁵Center for Computational Biology and Bioinformatics, Institute for Neuroscience, The
University of Texas at Austin

Corresponding authors:

Dr. Hans A. Hofmann	Dr. Susan C.P. Renn
The University of Texas at Austin	Department of Biology
Department of Integrative Biology	Reed College
2415 Speedway – C0990	3203 SE Woodstock Blvd.
Austin, TX 78712, USA	Portland, OR, 97202
Phone: 512-475-6754	Phone: 503-517-7967
Email: hans@utexas.edu	Email: renns@reed.edu

Figures: 5
Tables: 2
Supplementary Figures: 3
Supplementary Tables: 6

ABSTRACT:

The diversity of mating systems among animals is astounding. Importantly, similar mating systems have evolved even across distantly related taxa. However, our understanding of the mechanisms underlying these convergently evolved phenotypes is limited. Here, we examine on a genomic scale the neuromolecular basis of social organization in Ectodini cichlids from Lake Tanganyika. Using field collected males and females of four closely related species representing two independent evolutionary transitions from polygyny to monogamy, we take a comparative transcriptomic approach to test the hypothesis that these independent transitions have recruited similar gene sets. Our results demonstrate that while lineage and species exert a strong influence on neural gene expression profiles, social phenotype can also drive gene expression evolution. Specifically, 331 genes (~6% of those assayed) were associated with monogamous mating systems independent of species or sex. Among these genes, we find a strong bias (4:1 ratio) toward genes with increased expression in monogamous individuals. A highly conserved nonapeptide system known to be involved in the regulation of social behavior across animals was not associated with mating system in our analysis. Overall, our findings suggest deep molecular homologies underlying the convergent or parallel evolution of monogamy in different lineages of Ectodini cichlids.

Keywords: cichlid, gene expression, monogamy, polygyny, deep homology

INTRODUCTION

Social behavior can vary greatly across species, and such diversity lends itself to comparative analyses of the underlying proximate mechanisms (Crews & Moore 1986). The evolution of mating systems is an excellent example of how modification of an animal's reproductive behavior in response to environmental conditions can maximize fitness (Emlen & Oring 1977). Throughout the animal kingdom, monogamy, along with biparental care, has evolved repeatedly in response to ecological factors such as predation pressure or the availability and distribution of mates and resources (Grant 1993). At the mechanistic level, the role of the nonapeptide arginine vasopressin (AVP) and its V1a receptor in pair bond formation has been well established in the prairie vole, *Microtus ochrogaster* (Ophir et al 2008) as well as *Peromyscus* mice (Bendesky et al 2017). Importantly, AVP, and its non-mammalian homolog arginine vasotocin (AVT) have been shown to be associated with mating system variation in a range of vertebrates (reviewed by Oldfield et al 2013, Oldfield et al 2015), suggesting deep molecular homologies underlying the independent evolution of sociality in different vertebrate clades. If we are to understand the neural and molecular building blocks of social behavior, and how these are shaped over developmental and evolutionary time, we need to examine closely related species with different mating strategies that differ in few ecological and morphological aspects. The extent to which patterns of deep homology more broadly explain repeated evolution of social phenotypes requires genome-wide studies to complement the analysis of candidate genes.

The cichlid family of fishes offers unique opportunities for genome scale comparative studies examining the evolution of complex behavior patterns involved in behavioral decision-making and motor output, as the rapid radiation of species with diverse social phenotypes allows for comparison across closely related species (Brawand et al 2014, Hofmann 2003, Kocher 2004). Cichlids are a classic model system for studying the evolution of social behavior because of their extraordinary behavioral diversity (Kocher 2004, Pollen et al 2007). The abundant literature on cichlid ecology, evolution, behavior, and genomics puts cichlids in an ideal position for studies on the molecular basis of ecological and evolutionary processes. The behavioral diversity and genomic resources of the cichlid system makes it uniquely suited to identify the molecular substrates underlying the evolution of mating systems across a range of carefully selected species. The cichlid lineages in Lake Tanganyika, in particular, display a wide range of mating strategies (e.g., monogamous and polygynous) and provide parental care in a maternal, biparental, or cooperative manner either via buccal incubation (mouth brooding) or substrate guarding (see (Barlow 2000) for review).

Within these lineages, the monophyletic Ectodini tribe of cichlids exhibits striking differences in mating and parental care strategy, thus this clade is particularly well suited for a genomic exploration of the evolution of mating system. Habitat and social organization in this clade correlate with differences in gross morphology of the brain (Pollen et al 2007) as well as visual behavior (Dobberfuhl et al 2005). Importantly, the Ectodini tribe comprises species that display either monogamous (with biparental offspring care) or polygynous (with maternal only care) behavior and thus offers a unique

103 opportunity to study the evolution of distinct mating systems within a phylogenetic
104 framework. Kidd et al (2012) reconstructed the evolutionary relationships among 16
105 species of this lineage and found that during 2.5 to 3 million years of evolution within
106 this clade there have been 3–5 independent transitions from polygyny to monogamy (see
107 also (Koblmuller et al 2004). These results suggest that the evolution of parental care
108 strategies is not only remarkably fast, but also much more labile than previously
109 expected.

110
111 Over the last decade, a multitude of genomic resources have been developed for cichlids
112 (Aubin-Horth et al 2007, Salzburger et al in prep, Watanabe et al 2004), including draft
113 genomes and numerous transcriptomes for five African cichlid species (Brawand et al
114 2014). Despite the amazing diversity in ecology and behavior, there is very little
115 divergence in coding sequence (Loh et al 2008, Salzburger et al in prep, Watanabe et al
116 2004). Importantly, a cDNA microarray constructed for Burton's Mouthbrooder cichlid,
117 *Astatotilapia burtoni*, has been used to identify neural gene expression modules
118 associated with complex social traits in this model system for social neuroscience
119 (Duftner et al 2008, O'Connell & Hofmann 2012, Renn et al 2008, Renn & O'Rourke
120 2015). This array platform has also been applied successfully in a range of other teleost
121 fishes for which little or no genomic information was available (Aubin-Horth et al 2009,
122 Cummings et al 2008, Machado et al 2009, Renn et al 2004). Heterologous array
123 hybridization (whereby a single array platform is used to study a group of closely related
124 species) has been an effective tool for such comparative studies in several systems (e.g.
125 (Buckley 2007, Kassahn et al 2007). This is particularly the case when competitive

genomic DNA hybridizations between species are used to provide a means for assessing the degree to which sequence divergence biases the results of expression studies. We previously introduced a masking technique that corrects interspecific gene expression differences with genomic DNA hybridization ratios for the same set of species and determines a threshold sequence divergence (Machado et al 2009). This masking technique was first used to compare neural gene expression profiles between individual males and females from two closely related ectodine species, the polygynous *Enantiopus melanogenys* and the monogamous *Xenotilapia flavipinnis* (Machado et al 2009). In that study, the authors asked whether sex-specific neural gene expression is more closely associated with mating system than with gonadal sex and found that the gene expression profiles are largely species-specific, as relatively few genes showed conserved expression patterns associated with either sex (Machado et al 2009). However, in order to test whether this pattern is generalizable to other independent (parallel) transitions from polygyny to monogamy in this system (Kidd et al 2012) more species pairs have to be examined that represent independent transitions.

In the current study, we test further this hypothesis by examining the molecular basis of interspecific variation in social organization, specifically mating strategy, in four closely related ectodine cichlids. We aim to determine whether the independent evolutionary transitions from polygyny to monogamy have co-opted similar sets of genes, or whether the mechanisms that underlie seemingly similar mating strategies are wholly independent. We hypothesize that while some gene sets will be associated with lineage (phylogeny) and/or gonadal sex, at least a subset of genes associated with mating system

will be shared across the two contrasts. Such a correspondence in mating system specific gene sets across these parallel transitions from polygyny to monogamy would suggest broader patterns of deep homology.

METHODS

Choice of Species

We chose four closely related species from the monophyletic Ectodine clade differing in mating and parental care strategy (Figure 1). From the primarily sand-dwelling *Xenotilapia* genus we selected *X. ochrogenys*, a polygynous species in which only the females provide maternal care, and the closely related *X. flavipinnis*, a monogamous species with the male and female forming a pair bond and providing parental care. Similarly, the rock-dwelling *Asprotilapia leptura* is monogamous and biparental but the closely related *Microdontochromis tenuidentata* (which lives in intermediate habitats) is a polygynous species in which only the females provide maternal care (Kidd et al 2012, Pollen et al 2007). In the following, we refer to *X. ochrogenys* and *X. flavipinnis* as the “X-lineage” and to *A. leptura* and *M. tenuidentata* as the “nonX-lineage”. Together, these two contrasts represent two independent transitions from the ancestral polygynous to a monogamous mating system (Kidd et al 2012, Koblmuller et al 2004).

Sample collection

Unless indicated below, specimens were obtained by SCUBA (*X. flavipinnis*, *A. leptura*, *M. tenuidentata*) or netted by beach seine (*X. ochrogenys*) at field sites surrounding Kigoma (Tanzania) in July 2004 or Mpulungu (Zambia) in April 2005. Note that all

species reproduce year-round, although during the wet seasons' adverse weather events are more likely and can disrupt breeding. The animals were collected at the following locations: *X. flavipinnis*: Katonga Beach (4°55'22.5"S 29°40'20.5"E); *A. leptura*: Hilltop South Cliffs (4°53'24.50"S 29°36'45.00"E); *X. ochrogenys*: TAFIRI beach (4°53'10.11"S; 29°37'12.69"E); *M. tenuidentata*: near Mpulungu (8°44'47.5"S 31°07'30.5"E). The animals used in the present study ranged in standard length as follows. *A. leptura*: 9.3 - 10.0 cm (males) and 7.0 - 9.4 cm (females); *M. tenuidentata*: 6.63 - 7.4 cm (males) and 6.43 - 7.08 cm (females); *X. flavipinnis*: 5.8 - 6.5 cm (males) and 5.9 - 7.4 cm (females); and *X. ochrogenys*: 7.6 - 8.5 cm (males) and 6.5 - 7.4 cm (females). For each species, we collected adult and sexually mature individuals (five of either sex). The males of all species possessed large gonads with distinct and mature sperm packages. The females of both monogamous species were gravid and most were still holding fry in their mouths after capture, as were several of the monogamous males. Four of the five *X. ochrogenys* females were also mouth-brooding. All *M. tenuidentata* females had large ovaries with eggs at various stages of maturation; however, by the time of capture none were holding fry anymore. Animals were euthanized as quickly as possible after capture and brains were rapidly removed and stored in RNAlater solution (Ambion) within three minutes of death. Total RNA was extracted from brains using the TRIzol protocol (Invitrogen) following homogenization of brain tissue. RNA quality and concentration was determined using the Bioanalyzer (Agilent).

Microarray Platform

We used a second generation 19K *A. burtoni* microarray that contains features from a brain-specific library and a mixed-tissue library representing 1248 tentative contigs (TCs), 6572 singletons (sequences not belonging to a TC) and 6840 features without sequence information, for a total of 14,660 potential genes (Renn et al 2004, Salzburger et al in prep) (GEO platform GPL6416). After hybridization (see below for details) arrays were scanned (Axon 4000B: Axon Instruments) using Genepix 5.0 software (Axon Instruments).

Heterologous Array-based Comparative Genomic Hybridization

When heterologous hybridization is used to assess expression difference between two or more species, hybridization of genomic DNA from the two species onto the microarray can help identify genes with sequence divergence or other genomic characteristics that differ between the two species (Buckley 2007, Kassahn et al 2007, Machado et al 2009, Renn et al 2010). Six heterologous array-based comparative genomic hybridizations (aCGH) incorporating dye-swaps and all pairwise comparisons were used to identify array features with genomic bias among the four test species (Supplemental Figure 1A). We extracted genomic DNA from ethanol-stored fin clips of five to ten individuals per species using a standard Proteinase K/Phenol protocol, pooled by species, and sheared to roughly 1.5 Kb (Hydroshear, Genome Solutions/Digilab). Each pooled DNA sample (3 µg) was fluorescently labeled with Alexa-Fluor conjugated dCTP by Klenow reaction (Invitrogen, Bio-Prime), and quantified (Nanodrop 3300) in order to match Cy3 and Cy5 samples for between species competitive genomic hybridizations that proceeded for ~16 hours at 48 °C in Ambion Hyb Buffer 1 (Ambion) blocked by Cot-1DNA (Invitrogen)

using the 19K *A. burtoni* cDNA microarray platform. Scanning, background correction and normalization were also conducted as described below for expression arrays.

Genomic Hybridization Masking Procedure

Sequence divergence between the four Ectodini species and the array platform species (*A. burtoni*) will influence the observed gene expression differences and thus confound our results (Cummings et al 2008, Machado et al 2009, Renn et al 2010, Schumer et al 2011).

In order to determine which sequences have diverged enough to potentially bias hybridization efficacy, we examined the correlation between aCGH ratios and expression ratios (see below for hybridization protocol) for each pairwise species comparison. Both of these measures of relative hybridization to the array were calculated independently using the “lmFit” function in LIMMA and extracting each pairwise contrast between-species using a modified “contrasts.fit” function (Machado et al 2009), which calculates the correct (rather than the approximate) standard errors even in the presence of missing data due to quality filters. For these masking steps, we treated features of similar sequence (belonging to the same tentative contigs, TCs) as independent.

For each pairwise comparison, we used an iterative, stepwise process (scripts available on GitHub https://github.com/zrenn/Renn_etal_2017_Ectodini_expression-profiling). At each step the two additional array features that showed the greatest magnitude aCGH ratio, were added to the mask (*i.e.*, eliminated from further analysis). This process proceeded until a sufficient number of microarray features had been masked such that for an appropriately sized “test group” the correlation between aCGH ratio and gene

expression ratio was no longer significant ($P > 0.05$). The test group size was set for each pairwise comparison as the number of microarray features (greater than 50 to avoid spurious correlation) with the greatest magnitude of aCGH ratio that resulted in the largest positive and statistically significant ($P < 0.05$) correlation between aCGH ratio and expression ratio, when only features with genomic bias in the same direction of expression bias were masked. This method produced six independent genomic masks, one for each pairwise species comparison, resulting in 399 features masked for *A. leptura* vs. *M. tenuidentata*, 194 for *A. leptura* vs. *X. flavipinnis*, 384 for *A. leptura* vs. *X. ochrogenys*, 547 for *M. tenuidentata* vs. *X. flavipinnis*, 460 for *M. tenuidentata* vs. *X. ochrogenys* and 36 for *X. flavipinnis* vs. *X. ochrogenys*. The union of these six genomic masks, a total of 1319 array features, was masked out for all analyses of differential gene expression presented here.

Expression Analysis using Heterologous Array Hybridization

We used the nested loop design with dye-swap (Chruchill 2002) to emphasize within-lineage comparison between species of different mating strategies, analyzing each sex independently (Supplemental Figure 1B). As two-color microarray technology provides relative measures of gene expression for those samples that are compared directly or indirectly through competitive hybridization, the nested male loop and nested female loop produce independent gene lists that can then be compared. The nested female loops included nine X-lineage and nine nonX-lineage array comparisons between monogamous and polygynous species. The nested male loops included ten X-lineage and nine nonX-lineage comparisons. For both sexes, eight between-lineage hybridizations allowed direct

comparison between the X- and nonX-lineages, both within- and between- mating strategy. For each individual sample, 2 μ g of total RNA from whole brain was reverse transcribed and labeled according to Renn et al (2004). Briefly, amino-allyl dUTP (Sigma) was incorporated to the cDNA using oligo-dT(12 -18) with SuperScript II (Invitrogen) according to the manufacturer's protocol, and dye-coupling with Cy3 or Cy5 (CyDye Post-Labeling Reactive Dye Pack, Amersham) was followed by RNA hydrolyzation and purification. The neutralized color reaction was purified and combined with the appropriate competitive sample in hybridization buffer containing SSC and HEPES buffer with poly(dA)-poly(dT) (Sigma) for blocking and 0.1% for overnight hybridization at 65° C. Whole brain samples were used in order to allow for the use of wild caught samples.

Analysis of Differential Expression

In addition to masking out features determined to exhibit hybridization bias due to sequence divergence (see above), raw expression data were filtered such that all faint (average intensity < 2 standard deviations above background) and/or small (diameter < 60 pixels) were removed prior to background correction and normalization as described above. Here, to test for differential regulation at the level of the gene, the intensities of features of similar sequence (i.e., members of tentative contigs, TCs) were averaged to a single value prior to fitting a linear model to the data (lmFit) and estimating expression differences (contrasts.fit) (Machado et al 2009) and statistical significance (eBayes) (Smyth et al 2005). Eight contrasts of interest were assessed to identify gene expression patterns associated with mating system (1 contrast), lineage (1 contrast), and species (6

pair-wise contrasts). Determination of significant differential gene expression bias was based on a threshold of $P < 0.005$.

Functional Annotation and Gene Ontology Analysis

Microarray features for which sequence information was available were annotated according to the available cichlid genome sequences (Brawand et al 2014) by BLAST comparison to the available five species tissue-specific transcriptomes. For features with no transcript alignments with bit score greater than 200 and for features with available sequence length greater than that of the available transcript, the initial sequence was retained. We then compared this set of sequences to the set of predicted genes for the five cichlid species. The best hit with a bit score greater than 200 was used to annotate the microarray feature by transferring available gene names.

Gene Ontology (GO) terms were used to give functional annotation to the array features and to test for significant overrepresentation of functional groups among the genes showing mating-system-biased and lineage-biased expression. We used the Generic GO slims (Mundodi & Ireland 2002) previously developed to provide a useful summary of GO annotation for comparison of genomes and transcriptomes when a broad overview of the ontology content is required. Of the 7820 unique *A. burtoni* sequences available for array features, 4824 have been annotated to GO Slim terms. Note that any single assembled sequence may be annotated in all three ontologies and according to multiple ontology terms. A total of 20138 annotations have been applied: 8175 biological process annotations (including 66 GO Slim terms applied to 3591 array sequences); 7871

molecular function annotations (including 41 GO Slim terms applied to 4432 array sequences); and 4992 cellular component annotations (including 30 GO Slim terms applied to 3253 sequences).

Over-representation of biological process and molecular function GO terms for a regulated set of genes was determined in Cytoscape (Shannon et al 2003) using the Biological Network Gene Ontology tool, BiNGO (Maere et al 2005), which relies upon statistical significance of a hypergeometric distribution. Owing to the small number of genes for each ontology term and the relatively small number of genes that are regulated, there is little statistical power to identify significantly under-represented GO terms. Also, due to the highly non-independent nature of GO categories we use GO analysis as a hypothesis-generating tool and report only uncorrected hypergeometric P-values for over-representation.

RESULTS AND DISCUSSION

Our parallel nested loop designs consisted of 26 female and 28 male within-sex comparisons for analysis of gene expression patterns associated with lineage, species, and mating system using five males and females each from each species. The 1319 features that showed significant hybridization bias due to sequence variation were masked (i.e., removed) from the analysis (see methods), and quality filters were applied to all individual array features as described above prior to estimating expression coefficients. This resulted in 9888 genes available for the analysis of female gene expression, and only 5696 genes available for the analysis of males (which contained a few arrays of poorer

quality). Here, we report results for the intersection of 5232 genes that were available for analysis in both sexes (GEO: GSE 97082). We first confirm the effectiveness of this masking procedure before examining gene expression variation. We then report gene expression variation associated with lineage in males and in females. Next, we identify genes whose expression varies with mating system regardless of lineage and examine concordance between the sexes. Not surprisingly, we also identify genes that are regulated in a species-specific manner both within and across X- and nonX-lineages in both sexes. Finally, we use the Gene Ontology framework to provide further functional insights.

Ascertaining Array Features without Sequence Bias

In order to demonstrate the effectiveness of the masking procedure and confirm that these results represent true differences in gene expression levels, we asked whether genes identified as differentially expressed showed any (even non-significant) genomic hybridization bias. Importantly, for the 1284 genes that we found to be differentially expressed in any pair-wise species comparison, gene expression bias and genomic hybridization bias (aCGH ratio) were not correlated (Supplemental Table 1). Graphically, the array features that showed statistically significant expression variation in pair-wise species comparisons are randomly distributed with respect to aCGH ratios (Figure 2). This was true for both lineages and both sexes. As previously demonstrated by Machado et al (2010), these results confirm the suitability of heterologous array hybridization for the comparative analysis of gene expression patterns across species.

Gene Expression Varies According to Phylogeny

In order to identify gene expression regulation related to lineage we directly compared gene expression patterns of the two closely related species from the X-lineage to those from the nonX-lineage, for each sex separately. In the analysis of males, we found 92 genes with increased expression in the X-lineage and 69 genes with increased expression in the nonX-lineage ($P < 0.005$). In the analysis of females, 274 genes were found to have increased expression in the X-lineage and 157 genes were found to have increased expression in the nonX-lineage ($P < 0.005$) (Supplementary Table 2). Our experimental design with a nested loop design for each sex did not allow calculation of direct contrasts between the sexes. Instead, we compared the two lineage-bias gene lists (derived separately for males and for females) in order to identify the set of genes that exhibit concordant expression in the X-lineage compared to the nonX-lineage, independent of sex. We found 46 genes that were X-lineage biased and 15 genes that were nonX-lineage biased in both analyses. We found no genes that showed discordant lineage-biased expression between the male and female analyses.

We can visualize the role played by phylogenetic relatedness in regulating gene activity, independent of sex or mating system, by projecting the significant results from one sex on a volcano plot of the results from the analysis in the other sex (Figure 3). Specifically, the majority of the genes whose mRNA levels we identified as lineage-biased in males (Figure 3 A1) showed the same (though not always statistically significant) trend when projected onto the female analysis (Figure 3 A2). In the reverse projection, lineage-biased genes identified in the female analysis (Figure 3 B1) followed the same trend when

projected onto the male analysis (Figure 3 B2). This high level of concordance demonstrates the strong influence of lineage.

Our study is one of only a handful that have examined lineage-specific gene expression on a genomic scale using more than one species pair or multiple subspecies from the same lineage. For example, Oldham et al (Oldham et al 2006) compared the brain transcriptomes of humans and chimpanzees and showed that gene expression profiles vary greatly across discrete brain regions, with those in subcortical brain regions more conserved across species than in the cerebral cortex. However, the inference is weakened by the limited number of species and lineages in the analysis. Other groups used comparative transcriptomics to determine whether a molecular correlate of a phylotypic stage can be identified during animal development (six species of *Drosophila spec.*: (Kalinka et al 2010); four vertebrate species: (Irie & Kuratani 2011)). Maybe the most comprehensive study to date was carried out by Brawand et al (2011), who examined the dynamics of mammalian transcriptome evolution by comparing six organs across ten species of mammals (with a single representative from each lineage, plus an outgroup), demonstrating that the rate of gene expression evolution varies considerably among organs, lineages and chromosomes. Together with the present study, these examples support the notion that comparative transcriptomics can provide novel and fundamental insights into biological problems in a way that is complementary to experimental studies.

Gene Expression Profiles are Associated with Mating System

Our experimental design (Supplemental Figure 1B) emphasized direct comparisons of gene expression profiles of the two polygynous species (*X. ochrogenys* and *M. tenuidentata*, respectively) with the two monogamous species (*X. flavipinnis* and *A. leptura*, respectively). We chose this design as it provides considerably more power to discover genes whose expression patterns are associated with mating system compared with the species-centered analyses presented above. In the female analysis, 331 (~6%) of the genes analyzed here showed expression variation according to mating system, including 260 that were up-regulated in the monogamous species and 71 that were up-regulated in the polygynous species ($P < 0.005$). Similarly, in the male analysis a total of 247 genes showed mating system-associated expression variation, including 131 up-regulated in the monogamous species and 46 up-regulated in the polygynous species (Supplemental Table 3). Importantly, not only were both the female and the male results biased ~4:1 toward genes more highly expressed in the monogamous species, but these gene lists were also highly concordant. Specifically, 81 genes were concordantly regulated according to mating system; the vast majority of those (66 vs. 15) were up-regulated in monogamous individuals (Figure 4). Of note, only one gene showed discordant gene expression bias in this analysis. This strong pattern is also evident when we investigate trends regardless of statistical significance. Most of the genes identified as differentially expressed between monogamous and polygynous males (Figure 4 A1) showed the same direction of expression bias when projected onto the female analysis (Figure 4 A2) and *vice versa* (Figure 4B1 & 2).

The concordance between the evolution of neural gene expression in males and females during the transition from polygyny to monogamy is remarkable in light of the behavioral roles of the sexes, which have evolved in somewhat opposing directions in the monogamous species. Specifically, the transition to monogamy resulted in increased parental care by males and decreased or shared parental care by females. Our results suggest that changes in gene expression have been convergent or parallel, while changes at the behavioral level have been divergent (males evolved to increase care behavior while females reduced care). It should be noted, however, that gene expression studies aim at identifying differences. Statistical tests of equivalence are hampered by large data sets, the need to set arbitrary thresholds, and noisy data (but see(Qiu & Cui 2010)). Alternatively, the observed concordant evolution of gene expression may underlie parallel evolution of other behavior patterns such as pair-bonding, which is common in monogamous but not in polygynous mating systems.

With regard to social evolution, research on hymenoptera has been important to our understanding of evolution of genome regulation on multiple timescales associated with social phenotypes. On fairly rapid timescales, the same genes appear to be associated with plasticity and adaptation for aggression levels (Alaux et al 2009). Over longer timescales, genome regulation associated with provisioning and foraging behaviors appears to be largely conserved across bees (Sen Sarma et al 2007) and even between wasps and bees, which are separated by ~100 million years of evolution (Toth et al 2010). Similar attempts to address the evolution of gene expression associated with complex traits in animals have focused on physiological traits such as cold tolerance

(Makinen et al 2016), hypoxia tolerance (Kozak et al 2014), salinity tolerance (Latta et al 2012), predator avoidance (Fischer et al 2014, Ghalambor et al 2015), generation of a weakly electromagnetic field (Gallant et al 2014), sulfur tolerance (Kelley et al 2016) and eco-toxin resistance (Whitehead et al 2012). One recent study found similar gene expression modules in brain regions involved in vocal learning in songbirds, parakeet, hummingbirds, and humans (Pfenning et al 2014). Similar to our identification of common gene expression patterns associated with mating strategy, the emerging picture from these studies suggests deep molecular homologies underlying the independent evolution of monogamy in different lineages of ectodine cichlids.

Are Nonapeptide Systems Associated with Mating System Variation?

Nonapeptide systems play important, albeit complex, roles in the regulation of social behavior across vertebrates (and beyond). The nonapeptide arginine vasopressin (AVP), and its non-mammalian homolog arginine vasotocin (AVT), have been examined especially in relation to pair-bonding and social affiliation (e.g., prairie vole, *Microtus ochrogaster*: (Winslow et al 1993, Young et al 2008); zebrafish, *Taeniopygia guttata*: (Goodson & Adkins-Regan 1999); convict cichlid, *Amatitlania nigrofasciata*: (O'Connell et al 2012, Oldfield et al 2013, Oldfield & Hofmann 2011)). Recently, Oldfield et al (2015) integrated the known diversity in AVP/AVT function in affiliative behavior with resource defence theory in order to explain variation in territory-based mating systems across vertebrates. By carefully reviewing the literature, these authors found that expression of AVP/AVT (and its V1a receptor) in one particular neural circuit involving the lateral septum of the forebrain is associated with territorial behavior in males of

diverse species, likely due to effects of this system on social cognition. We therefore examined whether the expression levels of AVT was associated with mating system in either male or female Ectodini. (Note that probes for the nonapeptide oxytocin/mesotocin/isotocin as well as any of the nonapeptide receptors did not pass quality thresholds for inclusion in final analysis.) We found that expression did not vary significantly according to mating system (nor sex or species). It is, of course, possible that profiling gene expression in whole brains simply masked any variation in the AVT cell population that might be relevant to mating system. There are at least three functionally distinct AVT neuron groups in the teleost preoptic area alone (e.g., (Greenwood et al 2008)), and possibly also in the telencephalon ((Rodriguez-Santiago et al 2017)), and they play distinct and sometimes even opposing roles in the regulation of social behavior (Greenwood et al 2008). Alternatively, it may well be that at least at this level of analysis AVT is not involved in establishing the difference between monogamous and polygynous ectodine cichlids. Our finding that sequence variation in the AVT promoter across the entire Ectodini clade is not correlated with mating system (Duftner & Hofmann, unpublished results) supports this interpretation. Similarly, the number of preoptic AVT immunoreactive neurons (as a measure of AVT system activity) appears to be independent of mating system when examined in four Ectodini species (including three of the four analyzed in the present study): *A. leptura*, *X. flavipinnis* (both monogamous); and *X. ochrogenys*, *Enantiopus melanogenys* (both polygynous) (Shumway & Hofmann, unpublished results). Clearly, nonapeptide systems (along with numerous other neuroendocrine and neuromodulatory pathways) are deeply homologous in that they are critically involved in the regulation of social behavior across animals (for

review see (Weitckamp & Hofmann 2016)). It may, however, be naïve to expect their regulatory logic to be faithfully replicated in independent transitions to monogamy (or other complex social traits). Rather, a systems-level view facilitated by comparative transcriptomics may be much better suited to discover deeply homologous molecular substrates underlying convergent or parallel evolution of behavioral phenotypes and social systems.

Gene Expression Profiles associated with Specific Species

In order to identify gene expression differences that are likely to be involved in species-specific phenotypes (e.g. habitat, food preference, inter-specific aggression etc.), we next identified species-specific gene expression biases. Specifically, we identified genes that showed either significantly higher or significantly lower expression ($P < 0.005$) in one species contrasted against all three other species (see Table 1 for gene numbers and Supplementary Table 4 gene annotations). Interestingly, expression patterns of the polygynous species *M. tenuidentata* appeared to be the most divergent compared to the other three species, especially in females. Similarly, *X. ochrogenys* females, the other polygynous species, showed ten times more genes with species-specific biased expression compared to the males of this species. We also found significant concordance (*i.e.*, species-specific in both sexes: Fisher exact test $P < 0.05$) of gene sets (Table 1). This high level of concordance strongly demonstrates that species, regardless of sex, is a strong determinant for gene expression profile. Even though we cannot entirely rule out the possibility that, despite the masking procedure, species differences in genomic sequence might have affected these results, any such effects would likely be minor, such

that the species-specific expression observed here primarily relates to genes sufficiently conserved to be detected as orthologs. Of course, taxon-specific genes, which comprise 10 – 20 % of protein-coding genes in many sequenced genomes, likely also underlie lineage-specific traits (Dai et al 2008, Johnson & Tsutsui 2011, Khalturin et al 2009), such as mating and parental care behavior.

Species-specific Gene Expression Profiles Associated with Mating System within and across Lineages

We next conducted pairwise species comparisons in order to identify gene expression regulation associated with mating system that may be specific to a given species pair. First, we interrogated our dataset within lineage (X and nonX), comparing the expression profiles from each monogamous species to its closely related polygynous species (Table 2; gene annotations are provided in Supplemental Table 5). Notably for the analyses of both males and females, many more genes showed mating system-dependent expression within the nonX-lineage than within the X-lineage, likely due to the very divergent *M. tenuidentata* expression profiles (see above). With regard to the direction of gene expression associated with mating system, there were many more genes up-regulated than down-regulated in the monogamous species relative to the polygynous species (Table 2). This was true for both sexes in the nonX-lineage and for the females in the X-lineage.

Though we have less statistical power in this analysis, the genes identified reflect species-specific expression correlated with mating system. In order to focus on gene most likely related to mating system, we looked for concordance between the pairwise comparisons

within sex and within lineage. We found that the mating system-biases in gene expression in the X-lineage were highly concordant between males and females (109 concordant genes, 0 discordant genes), which was not the case in the nonX-Lineage (8 concordant genes, 8 discordant genes) (Supplemental figure 2; Table 2). Next, we looked for concordance between the pairwise comparisons within sex but across lineage for gene expression regulation associated with mating system. Interestingly, very few genes appeared to be concordantly regulated at a statistically significant level, and similar numbers were discordantly regulated (Supplemental Figure 3; Table 2). In summary, we found strong bias toward increased gene expression in monogamous individuals and this bias is concordant between sexes within the X-lineage but not concordant between sexes in the nonX-lineage nor between lineages regardless of sex. While we found no genes associated with mating system that showed concordant expression variation across both sexes and both lineages, there were genes that showed concordant regulation according to mating system in a) both sexes for one lineage or, more importantly, b) in one sex or the other in both lineages. This suggests that the changes in gene expression regulation that have evolved in males and in females are not entirely parallel or may be obscured by evolution of other phenotypic traits.

When we analyzed the data at the species level, at reduced statistical power, we found a lack of overall concordance when considering both male and female comparisons between X- and nonX-lineages. This suggests that the most extreme and robust differences in brain gene expression between species (i.e., those that can be detected with smaller sample sizes) do not reflect any shared molecular mechanisms underlying similar

mating system phenotypes. Instead, much of the species-specific differences in gene expression may be associated with species-specific phenotypes, or they may represent alternate mechanisms underlying similar phenotypes. Whatever the case, this result underscores the importance of including individuals from different lineages in the analysis, in addition to multiple comparisons of monogamy and polygyny. Because we see a strong phylogenetic signal, at the level of both lineage and species, the inclusion of two lineages and two independent instances of a transition to monogamy is critical for detecting gene expression variation associated with mating strategy.

Functional Analysis Using Gene Ontology

In order to infer functional information about the genes that show biased expression for mating strategy or lineage we applied generic GO-Slim annotations to the array features (Figure 5; Supplemental Table 6). Among genes up-regulated in both males and females of polygynous species, the GO categories translation (biological process), structural molecule activity (molecular function), and ribosome (cellular component) were over-represented ($P < 0.05$). Identification of these GO categories was driven largely by numerous genes annotated to ribosomal proteins. Conversely, the GO categories that were significantly associated with monogamy in the analysis of the males (response to stress) or the females (external encapsulating structure) were unique to each sex. Interestingly, the GO categories translation (biological process), structural molecule activity (molecular function), and ribosome (cellular component) were also significantly associated with nonX-lineage in both males and females as were biosynthetic process and cell cycle (biological process), structural molecule activity (molecular function), and

cytoplasm, intracellular and organelle (cellular component). These GO categories identified for the nonX-lineage were again largely driven by genes annotated to ribosomal proteins, while the GO term categories identified as enriched in the X-lineage (protein modification process and plasma membrane) included genes representing more neural specific pathways. The biological interpretation of these results remains uncertain, as is often the case with Gene Ontology analyses in transcriptome studies.

CONCLUSION

In the present study, we tested the hypothesis that independent evolutionary transitions from polygyny to monogamy in Ectodini cichlids have co-opted similar brain gene expression profiles, independent of phylogeny and/or sex. Male and female gene expression profiles were analyzed independently for lineage, species, and mating system. We identified signatures of gene expression that correspond to sex, species, lineage, and mating system. Importantly, we found high concordance between the male and female analyses for genes associated with either lineage or mating system, demonstrating that both phylogenetic relatedness and social phenotype can drive gene expression evolution. Further, among the genes associated with mating system independent of lineage, we find a strong bias (at a ratio of 4:1) toward genes that show increased expression in monogamous individuals, possibly due to the requirement for both males and females to change their affiliative and parental care behavior as monogamy evolved. A highly conserved nonapeptide system known to be critically involved in the regulation of social behavior across animals was not associated with mating system in our analysis. Our findings thus support the hypothesis that, on a genomic scale, independent evolutionary

transitions from polygyny to monogamy were accompanied by similar changes in brain gene expression patterns, suggesting deep molecular homologies underlying the independent evolution of monogamy in different lineages of ectodine cichlids. The extent to which these results hold more broadly across vertebrates is currently unknown, motivating a fascinating area of future research.

ACKNOWLEDGEMENTS

We thank George and Wilbroad Kazumbe, Sarah Bahan, and Alex Pollen for assistance in the field, and Lin Huffman and Celeste Kidd for laboratory assistance. We also thank Dr. Rebecca Young Brim for reading earlier versions of this manuscript. We are grateful to the Tanzania Fisheries Research Institute (TAFIRI), the Tanzania Commission on Science and Technology (COSTECH), and Alfeo Nikundiwe (University of Dar es Salaam) for their kind support of our research. Finally, we extend sincere thanks to Andy Cohen and Ellinor Michel from the Nyanza Project, John Fitzpatrick, Saskia Marijnissen and the Vaitha brothers for providing materials and support to our fieldwork. Sara Natale and Andrew Winterman, among other NSF-REU funded undergraduate students, contributed improvements to the LIMMA code, under advice of Dr. Albyn Jones (Reed College). This research was supported by an Erwin-Schrödinger postdoctoral fellowship provided by the FWF Austrian Science Fund (N.D.); NIH National Research Service Award MH070180-02, the Jordan Award of the American Cichlid Association, a Murdock Life Trust Foundation award, and National Science Foundation (NSF) grants IOS 0818957 and DEB 1021582 (S.C.P.R.); and by NSF grants IOS 0217915 and IOS 0843712, the Alfred P. Sloan Foundation, and a Dwight W. and

629 Blanche Faye Reeder Centennial Fellowship in Systematic and Evolutionary Biology

630 (H.A.H.).

631

Draft

REFERENCES

- Alaux C, Sinha S, Hasadsri L, Hunt GJ, Guzman-Novoa E, et al. 2009. Honey bee aggression supports a link between gene regulation and behavioral evolution. *Proceedings of the National Academy of Sciences of the United States of America* 106(36): 15400-05. doi: 10.1073/pnas.0907043106
- Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA. 2007. Masculinized dominant females in a cooperatively breeding species. *Mol. Ecol.* 16(7): 1349-58. doi: 10.1111/j.1365-294X.2007.03249.x
- Aubin-Horth N, Letcher BH, Hofmann HA. 2009. Gene-expression signatures of Atlantic salmon's plastic life cycle. *General and Comparative Endocrinology* 163(3): 278-84. doi: 10.1016/j.ygcen.2009.04.021
- Barlow GW. 2000. *Cichlid Fishes: Nature's Grand Experiment in Evolution*. Harper Collins Publishers.
- Bendesky A, Kwon YM, Lassance JM, Lewarch CL, Yao SQ, et al. 2017. The genetic basis of parental care evolution in monogamous mice. *Nature* 544(7651): 434-+. doi: 10.1038/nature22074
- Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513(7518): 375-+. doi: 10.1038/nature13726
- Buckley BA. 2007. Comparative environmental genomics in non-model species: using heterologous hybridization to DNA-based microarrays. *Journal of Experimental Biology* 210(9): 1602-06. doi: 10.1242/jeb.002402
- Chruchill GA. 2002. Fundamentals of experimental design for cDNA microarrays. *Nature Genetics* 32(Suppl.): 490-95. doi:
- Crews D, Moore MC. 1986. Evolution of Mechanisms Controlling Mating-Behavior. *Science* 231(4734): 121-25. doi: 10.1126/science.3941893
- Cummings ME, Larkins-Ford J, Reilly CRL, Wong RY, Ramsey M, Hofmann HA. 2008. Sexual and social stimuli elicit rapid and contrasting genomic responses. *Proceedings of the Royal Society B-Biological Sciences* 275(1633): 393-402. doi: 10.1098/rspb.2007.1454
- Dai HZ, Chen Y, Chen SD, Mao QY, Kennedy D, et al. 2008. The evolution of courtship behaviors through the origination of a new gene in *Drasophila*. *Proceedings of the National Academy of Sciences of the United States of America* 105(21): 7478-83. doi: 10.1073/pnas.0800693105
- Dobberfuhr AP, Ullmann JFP, Shumway CA, Ne. 2005. Visual acuity, environmental complexity, and social organization in African cichlid fishes. *Behavioral Neuroscience* 119(6): 1648-55. doi: 10.1037/0735-7044.119.6.1648
- Duftner N, Larkins-Ford J, Legendre M, Hofmann HA. 2008. Efficacy of RNA amplification is dependent on sequence characteristics: Implications for gene expression profiling using a cDNA microarray. *Genomics* 91(1): 108-17. doi: 10.1016/j.ygeno.2007.09.004
- Emlen ST, Oring LW. 1977. Ecology, Sexual Selection, and the Evolution of Mating Systems. *Science* 197(4300): 215-223 doi: 10.1126/science.327542
- Fischer EK, Harris RM, Hofmann HA, Hoke KL. 2014. Predator exposure alters stress physiology in guppies across timescales. *Hormones and Behavior* 65(2): 165-72. doi: 10.1016/j.yhbeh.2013.12.010

- Gallant JR, Traeger LL, Volkening JD, Moffett H, Chen PH, et al. 2014. Genomic basis for the convergent evolution of electric organs. *Science* 344(6191): 1522-25. doi: 10.1126/science.1254432
- Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* 525(7569): 372-+. doi: 10.1038/nature15256
- Goodson JL, Adkins-Regan E. 1999. Effect of intraseptal vasotocin and vasoactive intestinal polypeptide infusions on courtship song and aggression in the male zebra finch (*Taeniopygia guttata*). *Journal of Neuroendocrinology* 11(1): 19-25.
- Grant JWA. 1993. Whether or not to defend? The influence of resource distribution. *Marine Behavior and Physiology* 23:137-53. doi: 10.1080/10236249309378862
- Greenwood AK, Wark AR, Fernald RD, Hofmann HA. 2008. Expression of arginine vasotocin in distinct preoptic regions is associated with dominant and subordinate behaviour in an African cichlid fish. *Proceedings of the Royal Society B-Biological Sciences* 275(1649): 2393-402. doi: 10.1098/rspb.2008.0622
- Hofmann HA. 2003. Functional genomics of neural and behavioral plasticity. *Journal of Neurobiology* 54(1): 272-282. doi: 10.1002/neu.10172
- Irie N, Kuratani S. 2011. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. *Nature Communications* 2 doi: 10.1038/ncomms1248
- Johnson BR, Tsutsui ND. 2011. Taxonomically restricted genes are associated with the evolution of sociality in the honey bee. *BMC Genomics* 12 doi: 10.1186/1471-2164-12-164
- Kalinka AT, Varga KM, Gerrard DT, Preibisch S, Corcoran DL, et al. 2010. Gene expression divergence recapitulates the developmental hourglass model. *Nature* 468(7325): 811-U102. doi: 10.1038/nature09634
- Kassahn KS, Caley MJ, Ward AC, Connolly AR, Stone G, Crozier RH. 2007. Heterologous microarray experiments used to identify the early gene response to heat stress in a coral reef fish. *Molecular Ecology* 16(8): 1749-63. doi: 10.1111/j.1365-294X.2006.03178.x
- Kelley JL, Arias-Rodriguez L, Martin DP, Yee MC, Bustamante CD, Tobler M. 2016. Mechanisms Underlying Adaptation to Life in Hydrogen Sulfide-Rich Environments. *Molecular Biology and Evolution* 33(6): 1419-34. doi: 10.1093/molbev/msw020
- Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. 2009. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends in Genetics* 25(9): 404-13. doi: 10.1016/j.tig.2009.07.006
- Kidd MR, Duftner N, Koblmüller S, Sturmbauer C, Hofmann HA. 2012. Repeated Parallel Evolution of Parental Care Strategies within *Xenotilapia*, a Genus of Cichlid Fishes from Lake Tanganyika. *Plos One* 7(2): doi: 10.1371/journal.pone.0031236
- Koblmüller S, Salzburger W, Sturmbauer C. 2004. Evolutionary relationships in the sand-dwelling cichlid lineage of lake Tanganyika suggest multiple colonization of rocky habitats and convergent origin of biparental mouthbrooding. *J. Mol. Evol.* 58(1): 79-96. doi: 10.1007/s00239-003-2527-1

- 723 Kocher TD. 2004. Adaptive evolution and explosive speciation: The cichlid fish model.
724 Nat. Rev. Genet. 5(4): 288-98. doi: doi:10.1038/nrg1316
- 725 Kozak GM, Brennan RS, Berdan EL, Fuller RC, Whitehead A. 2014. Functional and
726 Population Genomic Divergence within and between Two Species of Killifish
727 Adapted to Different Osmotic Niches. *Evolution* 68(1): 63-80. doi:
728 10.1111/evo.12265
- 729 Latta LC, Weider LJ, Colbourne JK, Pfrender ME. 2012. The evolution of salinity
730 tolerance in *Daphnia*: a functional genomics approach. *Ecology Letters* 15(8):
731 794-802. doi: 10.1111/j.1461-0248.2012.01799.x
- 732 Loh Y-HE, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT. 2008. Comparative
733 analysis reveals signatures of differentiation amid genomic polymorphism in Lake
734 Malawi cichlids. *Genome Biology* 9(7): Unpaginated. doi: 10.1186/gb-2008-9-7-
735 r113
- 736 Machado HE, Pollen AA, Hofmann HA, Renn SC. 2009. Interspecific profiling of gene
737 expression informed by comparative genomic hybridization: A review and a novel
738 approach in African cichlid fishes. *Integrative and Comparative Biology* 49(6):
739 644-59. doi: 10.1093/icb/icp080
- 740 Makinen H, Papakostas S, Vollestad LA, Leder EH, Primmer CR. 2016. Plastic and
741 Evolutionary Gene Expression Responses Are Correlated in European Grayling
742 (*Thymallus thymallus*) Subpopulations Adapted to Different Thermal
743 Environments. *Journal of Heredity* 107(1): 82-89. doi: 10.1093/jhered/esv069
- 744 Mundodi S, Ireland A. 2002. Generic GO Slim. WWW Document:
745 <http://www.geneontology.org/GO.slims.shtml>
- 746 O'Connell LA, Hofmann HA. 2012. Social Status Predicts How Sex Steroid Receptors
747 Regulate Complex Behavior across Levels of Biological Organization.
748 *Endocrinology* 153(3): 1341-51. doi: 10.1210/en.2011-1663
- 749 O'Connell LA, Matthews BJ, Hofmann HA. 2012. Isotocin regulates paternal care in a
750 monogamous cichlid fish. *Hormones and Behavior* 61(5): 725-33. doi:
751 10.1016/j.yhbeh.2012.03.009
- 752 Oldfield RG, Harris RM, Hendrickson DA, Hofmann HA. 2013. Arginine vasotocin and
753 androgen pathways are associated with mating system variation in North
754 American cichlid fishes. *Hormones and Behavior* 64(1): 44-52. doi:
755 10.1016/j.yhbeh.2013.04.006
- 756 Oldfield RG, Harris RM, Hofmann HA. 2015. Integrating resource defence theory with a
757 neural nonapeptide pathway to explain territory-based mating systems. *Frontiers*
758 in *Zoology* 12 doi: 10.1186/1742-9994-12-S1-S16
- 759 Oldfield RG, Hofmann HA. 2011. Neuropeptide regulation of social behavior in a
760 monogamous cichlid fish. *Physiology & Behavior* 102(3-4): 296-303. doi:
761 10.1016/j.physbeh.2010.11.022
- 762 Oldham MC, Horvath S, Geschwind DH. 2006. Conservation and evolution of gene
763 coexpression networks in human and chimpanzee brains. *Proceedings of the*
764 *National Academy of Sciences of the United States of America* 103(47): 17973-
765 78. doi: 10.1073/pnas.0605938103
- 766 Ophir AG, Wolff JO, Phelps SM. 2008. Variation in neural V1aR predicts sexual fidelity
767 and space use among male prairie voles in semi-natural settings. *Proceedings of*

- the National Academy of Sciences of the United States of America 105(4): 1249-54. doi: 10.1073/pnas.0709116105
- Pfenning AR, Hara E, Whitney O, Rivas MV, Wang R, et al. 2014. Convergent transcriptional specializations in the brains of humans and song-learning birds. *Science* 346(6215): 1333-+. doi: 10.1126/science.1256846
- Pollen AA, Dobberfuhl AP, Scace J, Igulu MM, Renn SCP, et al. 2007. Environmental Complexity and Social Organization Sculpt the Brain in Lake Tanganyikan Cichlid Fish. *Brain Behavior and Evolution* 70(1): 21-39. doi:
- Qiu J, Cui XQ. 2010. Evaluation of a Statistical Equivalence Test Applied to Microarray Data. *Journal of Biopharmaceutical Statistics* 20(2): 240-66. doi: 10.1080/10543400903572738
- Renn SCP, Aubin-Horth N, Hofmann HA. 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. *BMC Genomics* 5doi: 10.1186/1471-2164-5-42
- Renn SCP, Aubin-Horth N, Hofmann HA. 2008. Fish and chips: functional genomics of social plasticity in an African cichlid fish. *Journal of Experimental Biology* 211(18): 3041-56. doi: 10.1242/jeb.018242
- Renn SCP, Machado HE, Jones A, Soneji K, Kulathinal RJ, Hofmann HA. 2010. Using comparative genomic hybridization to survey genomic sequence divergence across species: a proof-of-concept from *Drosophila*. *BMC Genomics* 11:271. doi:
- Renn SCP, O'Rourke C. 2015. Integrating adaptive trade-offs between parental care and feeding regulation. *Current Opinion in Behavioral Sciences* 6:160-67. doi: 10.1016/j.cobeha.2015.11.010
- Rodriguez-Santiago M, Nyuren M, Winton LS, Weickkamp CA, Hofmann HA. 2017. Arginine Vasotocin is expressed in surprising regions of the teleost forebrain. *Frontiers in Endocrinology* 8(195) doi:10.3389/fendo.2017.00195
- Salzburger W, Renn SCP, Steinke D, Hofmann HA, Fernald RD, et al. (2008) Annotation of expressed sequence tags for the East African cichlid fish species *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. *BMC Genomics* 9:96 doi: 10.1186/1471-2164-9-96
- Schumer M, Krishnakant K, Renn SCP. 2011. Comparative gene expression profiles for highly similar aggressive phenotypes in male and female cichlid fishes (*Julidochromis*). *Journal of Experimental Biology* 214(19): 3269-78. doi: 10.1242/jeb.055467
- Sen Sarma M, Whitfield CW, Robinson GE. 2007. Species differences in brain gene expression profiles associated with adult behavioral maturation in honey bees. *BMC Genomics* 8 doi: 10.1186/1471-2164-8-202
- Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, et al. 2003. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11): 2498-504. doi: 10.1101/gr.1239303
- Smyth GK, Michaud J, Scott HS. 2005. Use of within-array replicate spots for assessing differential expression in microarray experiments. *Bioinformatics* 21(9): 2067-75. doi: 10.1093/bioinformatics/bti270
- Toth AL, Varala K, Henshaw MT, Rodriguez-Zas SL, Hudson ME, Robinson GE. 2010. Brain transcriptomic analysis in paper wasps identifies genes associated with

behaviour across social insect lineages. *Proceedings of the Royal Society B-Biological Sciences* 277(1691): 2139-48. doi: 10.1098/rspb.2010.0090

Watanabe M, Kobayashi N, Shin-I T, Horiike T, Tateno Y, et al. 2004. Extensive analysis of ORF sequences from two different cichlid species in Lake Victoria provides molecular evidence for a recent radiation event of the Victoria species flock - Identity of EST sequences between *Haplochromis chilotes* and *Haplochromis* sp "Redtailsheller". *Gene* 343(2): 263-69. doi: 10.1016/j.gene.2004.09.013

Weitckamp CA, Hofmann HA. 2016. Brain Systems Underlying Social Behavior In *Evolution of Nervous Systems*, ed. GF Striedter, JH Kaas, pp. 327-34

Whitehead A, Pilcher W, Champlin D, Nacci D. 2012. Common mechanism underlies repeated evolution of extreme pollution tolerance. *Proceedings of the Royal Society B-Biological Sciences* 279(1728): 427-33. doi: 10.1098/rspb.2011.0847

Winslow JT, Hastings N, Carter CS, Harbaugh CR, Insel TR. 1993. A Role for Central Vasopressin in Pair Bonding in Monogamous Prairie Voles. *Nature* 365(6446): 545-48. doi: 10.1038/365545a0

Young KA, Liu Y, Wang ZX. 2008. The neurobiology of social attachment: A comparative approach to behavioral, neuroanatomical, and neurochemical studies. *Comparative Biochemistry and Physiology C-Toxicology & Pharmacology* 148(4): 401-10. doi: 10.1016/j.cbpc.2008.02.004

Supplementary data file (gen-2017-0075.R1Suppla.xls): Microsoft Excel data file containing Supplementary Tables 1-5.

Supplemental Table 1: Validating correlation statistics for pairwise species comparisons between genomic hybridization ratio and gene expression ratio from significantly regulated genes for both male and female analysis. Lack of correlation validates genomic masking technique.

Supplemental Table 2: Gene expression regulation associated with lineage for male and female analyses of X-lineage versus nonX-lineage indicating concordant and non-concordant regulation.

Supplemental Table 3: Gene expression regulation associated with mating strategy for male and female analyses of monogamy versus polygyny indicating concordant and non-concordant regulation.

Supplemental Table 4: Gene expression regulation associated with specific species for male and female analyses.

Supplemental Table 5: Gene expression regulation associated with mating system, monogamy versus polygyny, within X-lineage and within the nonX-lineage for male and female analyses.

Supplemental Table 6: Gene Ontology categories including gene annotations

Supplemental Figure 1 (gen-2017-0075.R1Supplb.jpg): Experimental design for heterologous array hybridization experiments. (A) Experimental loop design for

assessing hybridization bias due to sequence divergence. Arrows indicate single array hybridizations, with tails indicating the genomic DNA samples labeled with Cy3 and the arrowheads indicating the genomic DNA sample labeled with Cy5. Purple shades represent polygynous species and orange shades represent monogamous species as in main figures. (B) For the gene expression analyses a nested loop design was used that emphasized within-sex and within-lineage analyses across mating systems, while also including between lineage hybridizations for statistical analysis of differences between independent evolutionary transitions. Each arrow indicates single array hybridizations, with the tails indicating the samples labeled with Cy3 and the arrowheads indicating the samples labeled with Cy5. *X. flavipinnis* (Xf), *A. leptura* (Al), *M. tenuidentata* (Mt) and *X. ochrogenys* (Xo). Colors as in (A). Some pairwise comparisons were omitted (e.x. Al₃vsMt₃) due to technical errors in microarray hybridization and insufficient DNA for replicates.

Supplemental Figure 2(gen-2017-0075.R1Supplc.jpg): representation of within-sex gene expression variation according to mating system. The graphs show log 2 fold-differences in gene expression between monogamous and polygynous females (*A. leptura*, Al.F, vs *M. tenuidentata*, Mt.F) plotted against those of males (*A. leptura*, Al.M, vs *M. tenuidentata*, Mt.M). Positive values indicate monogamy-bias in gene expression, negative values indicate polygyny bias. Highlighted features indicate statistical significance ($P < 0.005$) with polygyny-bias (deep purple and magenta) or monogamy bias (orange and gold) for (A) X-lineage analysis (*X. flavipinnis* vs *X. ochrogenys*) and (B) nonX-lineage analysis (*A. leptura* vs *M. tenuidentata*). Numbers in quadrants represent the number of genes with statistically significant concordant or discordant expression (see also Table 2).

Supplemental Figure 3(gen-2017-0075.R1Suppld.jpg): Representation of within-lineage gene expression variation according to mating system. The graphs show log 2 fold-differences in gene expression between monogamous and polygynous species of the nonX-lineage (*A. leptura*, Al, vs *M. tenuidentata*, Mt) plotted against those of the X-lineage (*X. flavipinnis*, Xf, vs *X. ochrogenys*, Xo). Positive values indicate monogamy-

890 bias in gene expression, negative values indicate polygyny bias. Highlighted features
891 indicate statistical significance ($P < 0.005$) with polygyny-bias (orange and gold) or
892 monogamy bias (deep purple and magenta) for (A) the female analysis and (B) the male
893 analysis. Numbers in quadrants represent the number of genes with statistically
894 significant concordant or discordant expression (see also Supplemental Table 2).
895

Draft

FIGURE LEGENDS

Figure 1. Convergent evolution of mating strategies within the Ectodini and Xenotilapia clades from Lake Tanganyika. An ancestral character state reconstruction by maximum parsimony reveals multiple transitions from polygyny (purple) to monogamy (orange) adapted from Kidd et al. 2012.

Figure 2. Effectiveness of masking procedure in reducing error in gene expression analyses. Volcano plots showing statistical significance in expression variation (log odds) against log 2 genomic DNA hybridization ratios comparing (A) the X-lineage species *X. flavipinnis* with *X. ochrogenys* males (A1) and females (A2), and (B) the nonX-lineage species *A. leptura* with *M. tenuidentata* males (B1) and females (B2), respectively. Colored arrows indicate direction of species bias for genomic hybridization ratios. Highlighted features indicate genes that showed statistically significant gene expression bias in the pair-wise species comparisons. The lack of color segregation for highlighted features demonstrates that the genomic masking procedure is effective in preventing bias by species-specific genomic DNA sequence characteristics in heterologous array hybridization experiments.

Figure 3. Graphical representation of gene expression variation according to lineage. Volcano plots showing expression variation differences between the X- and nonX-lineages (log 2 fold-differences) plotted in grey against statistical significance (log odds) for male (A1 & B2) and female (B1 & A2) analyses, respectively. Genes whose expression pattern was identified as significantly lineage-biased (highlighted in black) in

the male analysis (A1) showed the same lineage-bias (though not always significant) projected onto the female analysis (A2). Similarly, genes whose expression pattern was significantly lineage-biased (highlighted in black) in the female analysis (B1) showed the same lineage-bias (though not always significant) when projected onto the male analysis (B2).

Figure 4. Graphical representation of gene expression variation according to mating system. Volcano plots showing log 2 fold-differences in gene expression according to mating system plotted against statistical significance (log odds). Genes identified as differentially expressed between monogamous (orange) and polygynous (purple) individuals in the analysis of males (A1) showed the same (though not always significant) trend when projected onto the analysis of females (A2). Conversely, mating system-associated genes identified in the analysis of females (B1) follow the same trend when projected onto the analysis of males (B2).

Figure 5. Functional analysis using Gene Ontology. The generic GO-Slim was used to analyze gene function. Red shading indicates degree of statistical significance for over-represented categories. Gene expression biases were more prevalent across clades (X- vs. nonX-lineage). Translation and structural molecule activity, and ribosome are highly significant for mating system in polygynous individuals, but were also significant in the nonX-lineage. Cell cycle, cytoplasm, and organelle categories are significant in nonX-individuals only.

942

Draft

Tables

Table 1: Species-specific gene expression variation in males and females of the four study species. Shown are numbers of genes significantly ($P < 0.005$) different in expression in any given species compared to the three other species for males and females as well as those that are shared (overlap) between the sexes. Numbers in parentheses indicate the number of genes significantly up- or down-regulated in the named species.

	Male (up/down)	Female (up/down)	Concordant (up/down)
<i>X. flavipinnis</i>	230 (133/97)	218 (156/62)	50 (35/15)
<i>X. ochrogenys</i>	41 (22/19)	339 (166/173)	17 (7/10)
<i>A. leptura</i>	69 (38/31)	109 (67/42)	14 (11/3)
<i>M. tenuidentata</i>	254 (88/166)	892 (327/565)	193 (66/127)

Table 2: Lineage-specific gene expression variation in males and females of the four study species. For each lineage the numbers of genes are shown that are significantly ($P < 0.005$) different in expression according to mating strategy as well as those that are shared (overlap) between the sexes. Numbers in parentheses indicate the number of genes significantly up-regulated in monogamous or polygynous species, respectively.

	Male (monogamy/polygyny)	Female (monogamy/polygyny)	Concordant	Discordant
X-lineage	91 (47/44)	103 (70/33)	8 (1/7)	8 (6/2)
NonX-lineage	207 (146/61)	593 (412/181)	109 (81/28)	0 (0/0)
Concordant	13 (7 / 6)	9 (7/2)	0 (0 / 0)	
Discordant	10 (8/2)	15 (7/8)

Figure 1. Convergent evolution of mating strategies within the Ectodini and Xenotilapia clades from Lake Tanganyika. An ancestral character state reconstruction by maximum parsimony reveals multiple transitions from polygyny (purple) to monogamy (orange) adapted from Kidd et al. 2012.

190x254mm (96 x 96 DPI)

Figure 2. Effectiveness of masking procedure in reducing error in gene expression analyses. Volcano plots showing statistical significance in expression variation (log odds) against log 2 genomic DNA hybridization ratios comparing (A) the X-lineage species *X. flavipinnis* with *X. ochrogenys* males (A1) and females (A2), and (B) the nonX-lineage species *A. leptura* with *M. tenuidentata* males (B1) and females (B2), respectively. Colored arrows indicate direction of species bias for genomic hybridization ratios. Highlighted features indicate genes that showed statistically significant gene expression bias in the pair-wise species comparisons. The lack of color segregation for highlighted features demonstrates that the genomic masking procedure is effective in preventing bias by species-specific genomic DNA sequence characteristics in heterologous array hybridization experiments.

190x254mm (96 x 96 DPI)

Draft

Figure 3. Graphical representation of gene expression variation according to lineage. Volcano plots showing expression variation differences between the X- and nonX-lineages (log 2 fold-differences) plotted in grey against statistical significance (log odds) for male (A1 & B2) and female (B1 & A2) analyses, respectively. Genes whose expression pattern was identified as significantly lineage-biased (highlighted in black) in the male analysis (A1) showed the same lineage-bias (though not always significant) projected onto the female analysis (A2). Similarly, genes whose expression pattern was significantly lineage-biased (highlighted in black) in the female analysis (B1) showed the same lineage-bias (though not always significant) when projected onto the male analysis (B2).

190x254mm (96 x 96 DPI)

Figure 4. Graphical representation of gene expression variation according to mating system. Volcano plots showing log₂ fold-differences in gene expression according to mating system plotted against statistical significance (log odds). Genes identified as differentially expressed between monogamous (orange) and polygynous (purple) individuals in the analysis of males (A1) showed the same (though not always significant) trend when projected onto the analysis of females (A2). Conversely, mating system-associated genes identified in the analysis of females (B1) follow the same trend when projected onto the analysis of males (B2).

190x254mm (96 x 96 DPI)

Figure 5. Functional analysis using Gene Ontology. The generic GO-Slim was used to analyze gene function. Red shading indicates degree of statistical significance for over-represented categories. Gene expression biases were more prevalent across clades (X- vs. nonX-lineage). Translation and structural molecule activity, and ribosome are highly significant for mating system in polygynous individuals, but were also significant in the nonX-lineage. Cell cycle, cytoplasm, and organelle categories are significant in nonX-individuals only.

190x254mm (96 x 96 DPI)

Supplemental Figure 1: Experimental design for heterologous array hybridization experiments. (A) Experimental loop design for assessing hybridization bias due to sequence divergence. Arrows indicate single array hybridizations, with tails indicating the genomic DNA samples labeled with Cy3 and the arrowheads indicating the genomic DNA sample labeled with Cy5. Purple shades represent polygynous species and orange shades represent monogamous species as in main figures. (B) For the gene expression analyses a nested loop design was used that emphasized within-sex and within-lineage analyses across mating systems, while also including between lineage hybridizations for statistical analysis of differences between independent evolutionary transitions. Each arrow indicates single array hybridizations, with the tails indicating the samples labeled with Cy3 and the arrowheads indicating the samples labeled with Cy5. *X. flavipinnis* (Xf), *A. leptura* (Al), *M. tenuidentata* (Mt) and *X. ochrogenys* (Xo). Colors as in (A). Some pairwise comparisons were omitted (e.g. Al₃vsMt₃) due to technical errors in microarray hybridization and insufficient DNA for replicates.

190x254mm (96 x 96 DPI)

Supplemental Figure 2: representation of within-sex gene expression variation according to mating system. The graphs show log 2 fold-differences in gene expression between monogamous and polygynous species of the nonX lineage (*A. leptura*, *Al.F*, vs *M. tenidentata*, *Mt.F*) plotted against those the X lineage (*X. flavipinnis* vs *X. ochrogenys*). Positive values indicate monogamy-bias in gene expression, negative values indicate polygyny bias. Highlighted features indicate statistical significance ($P < 0.005$) with polygyny-bias (deep purple and magenta) or monogamy bias (orange and gold) for (A) males and (B) females. Numbers in quadrants represent the number of genes with statistically significant concordant or discordant expression (see also Table 2).

190x254mm (96 x 96 DPI)

Supplemental Figure 3: Representation of within-lineage gene expression variation according to mating system. The graphs show log₂ fold-differences in gene expression between monogamous and polygynous species of the non-X-lineage (A. *leptura*, Al, vs *M. tenidentata*, Mt) plotted against those of the X-lineage (*X. flavipinnis*, Xf, vs *X. ochrogenys*, Xo). Positive values indicate monogamy-bias in gene expression, negative values indicate polygyny bias. Highlighted features indicate statistical significance ($P < 0.005$) with polygyny-bias (orange and gold) or monogamy bias (deep purple and magenta) for (A) the female analysis and (B) the male analysis. Numbers in quadrants represent the number of genes with statistically significant concordant or discordant expression (see also Table 2).

190x254mm (96 x 96 DPI)