Interspecific profiling of gene expression informed by comparative genomic hybridization: A review and a novel approach in African cichlid fishes

Heather E. Machado,* Alexander A. Pollen,† Hans A. Hofmann‡ and Suzy C. P. Renn[1,*]

*Department of Biology, Reed College, Portland, OR 97202, USA; †Program in Neuroscience, Stanford University, Stanford, CA 94305, USA; ‡Section of Integrative Biology, Institute for Molecular and Cellular Biology, Institute for Neuroscience, The University of Texas at Austin, Austin, TX 78712, USA

Synopsis Modern genomic approaches have facilitated great progress in our understanding of the molecular and genetic underpinnings of ecological and evolutionary processes. Analysis of gene expression through heterologous hybridization in particular has enabled genome-scale studies in many ecologically and evolutionarily interesting species. However, these studies have been hampered by the difficulty of comparing—on a common array platform—gene-expression profiles across species due to sequence divergence altering the dynamics of hybridization. All too often, comparisons of expression profiles across species were limited to contrasting lists of gene or even of just functional categories. Here we review these issues and propose a novel solution. Exploiting the diverse cichlid lineages of East Africa as our model-system, we then present results from an experimental case study that compares the neural gene-expression profiles of males and females of two species that differ in mating system. Using a single microarray platform that contains genes from one species, *Astatotilapia burtoni*, we conducted a total of 16 direct comparisons for neural gene-expression level between individual males and females from a pair of sister species, the polygynous *Enantiopus melanogenys* and the monogamous *Xenotilapia flavipinnis*. Next, we conducted a meta-analysis with previously published data from two different intra-specific expression studies to determine whether sex-specific neural gene expression is more closely associated with behavioral phenotype than it is with gonadal sex. Our results indicate that the gene expression profiles are species-specific to a large extent, as relatively few genes show conserved expression patterns associated with either sex. Finally, we describe how competitive genomic DNA hybridizations between the two focal species allow us to assess the degree to which divergence of sequences biases the results. We propose a masking technique that correlates interspecific expression ratios obtained with cDNA with hybridization ratios obtained with genomic DNA for the same set of species and determines threshold sequence divergence to reduce false positives. Our approach should be applicable to a wide range of interesting questions related to the evolution and ecology of gene expression.

Introduction

The unparalleled diversity of cichlid fishes from the African rift lakes resembles a 'natural mutant' screen sculpted by natural selection over short evolutionary timescales (Kocher 2004). Cichlid fishes provide a singular opportunity to understand how genomic variation (regulatory and structural) sculpts adaptive variation. Sequence on a genomic scale is now available for nine cichlid species. Four species have been the focus of EST projects: *Astatotilapia burtoni*, a haplochromine cichlid from Lake Tanganyika (Renn et al. 2004; Salzburger et al. 2008), which has become a major model system for the study of behavioral and neural plasticity (Hofmann, 2003); two haplochromines from Lake Victoria (*Haplochromis chilotes*, *H. sp.* '*Redtailsheller*') (Watanabe et al. 2004); and most recently the Nile tilapia, *Oreochromis niloticus* (Kocher, personal communication). cDNA microarrays are available for *A. burtoni* (Renn et al. 2004) and the two Lake Victoria haplochromines (Kijimoto et al. 2005). In addition to these resources, low-coverage (0.1X) genomic sequences exist for five Malawian species (Loh et al. 2008). Importantly, four species are currently being prepared for whole genome sequencing (http://www.genome.gov/11007951). For the majority of cichlid genes, all interspecific comparisons that are available suggest little divergence of sequences within cichlids, especially in the protein-coding regions (Watanabe et al. 2004; Salzburger et al. 2008;

From the symposium "Genomics and Vertebrate Adaptive Radiation: A Celebration of the First Cichlid Genome" presented at the annual meeting of the Society for Integrative and Comparative Biology, January 3–7, 2009, at Boston, Massachusetts.
[1]E-mail: renns@reed.edu
Integrative and Comparative Biology, volume 49, number 6, pp. 644–659
doi:10.1093/icb/icp080
Advanced Access publication August 13, 2009

Loh et al. 2008). Among the five Lake Malawi species for which this information is available, diversity of genomic sequence (Watterson's $\theta_w = 0.26\%$) and Jukes-Cantor genetic distance (0.23–0.29%) are low (Loh et al. 2008). This ever-increasing wealth of sequence information, in combination with the abundant ecological and evolutionary literature for this group, makes cichlids ideal subjects for studies on the molecular basis of ecological and evolutionary processes.

The cDNA microarray constructed for the 'model cichlid', *A. burtoni*, has been used to identify in the brain co-regulated gene modules associated with complex social traits, such as social dominance (Renn et al. 2008; Fraser et al. in preparation; Larkins-Ford et al. in preparation). Overall, these results from one 'model cichlid' underscore the power of this system for explaining social and hormonal regulation of brain and behavior in molecular terms. However, to fully capitalize on the comparative power of the behavioral diversity of the cichlid system, it is important to pursue similar questions in other species.

The feasibility of 'heterologous hybridization' on a DNA microarray, as a means of obtaining gene-expression profiles from species other than the platform species, is well supported by both technical and biological reports (Renn et al. 2004; Buckley 2007; Kassahn et al. 2007, 2008; Cummings et al. 2008). Through heterologous hybridization, gene expression can be studied for a range of species for which little or no genomic information is available. Using the *A. burtoni* array, Renn et al. (2004) provided one of the first systematic reports that the utility of the cDNA array decreases somewhat with phylogenetic distance, but because of the low levels of genetic variation across cichlid species (Watanabe et al. 2004; Salzburger et al. 2008; Loh et al. 2008), the single cichlid array platform performs very well across a range of cichlid species (Renn et al. 2004).

In order to study the evolution of adaptive phenotypes, such as social behavior, it is imperative to move beyond the single-species analysis and take full advantage of species diversity by comparing expression profiles across species using the same array platform. Obviously, sequence divergence, causing substantial base-pair mismatches between the species of interest and the species from which the single microarray platform was derived, can lower the efficiency of hybridization. For a few species related to model genetic organisms, e.g. drosophilids (Clark et al. 2007), nematodes (Blaxter et al. 2004), salmonids (von Schalburg et al. 2008), and primates (Toleno et al. 2009), completed genome sequences, or large EST sets, allow for the computational assessment of sequence divergence that would affect microarray hybridization. However, this option is not (yet) available for the majority of ecologically or evolutionarily interesting model systems due to the lack of sequenced genomes (but see Rokas and Abbot, 2009). Alternative experimental methods must therefore be devised in order to identify microarray features that will, or will not, be useful in a specific experiment using interspecific heterologous hybridization. Array based comparative genomic DNA hybridizations (aCGH) to microarrays have been utilized to identify and 'mask' microarray features that are significantly affected by sequence divergence prior to experiments that employ heterologous expression profiling for a single within-species, heterologous hybridization experiment (Hammond et al. 2005; Kassahn et al. 2007; Cummings et al. 2008). Such a correction would be of even greater utility for the comparative analysis of transcriptomes across species, which hold great promise for modern biology. We first review this literature and then demonstrate a novel masking approach in the form of a proof of concept experiment.

Heterologous hybridizations are confounded by sequence divergence: a short review

Even though African cichlids are thought to be quite genetically homogeneous (Loh et al. 2008), it cannot be ruled out that certain genes show high levels of sequence variation, especially if they may have experienced adaptive evolution. Of course, the very same genes that show significant sequence variation across species (e.g. opsins: Spady et al. 2005) may also vary in expression (Carleton et al. 2008), thus potentially confounding transcriptome studies. In fact, it has been suggested that for some species there is a correlation between sequence divergence for a gene and divergence in expression regulation for that gene (*Drosophila*: Nuzhdin et al. 2004; *Xenopus*: Sartor et al. 2006; but see for yeast: Tirosh and Barkai, 2008). Therefore, in order to accurately identify interspecific variation in gene expression one must first identify array features that are influenced by sequence divergence.

While most ecologically and evolutionarily relevant questions regarding gene expression will benefit from a comparative approach and analysis in multiple species, the comparative studies conducted thus far have examined the gene-expression profile for a single non-model species and compared

it to that obtained for a rather closely related model-species using the same microarray platform. Many of these examples address questions of ecological relevance related to climatic change. Hypoxia response in the viviparous species *Xiphophorus* was compared to that in the genetic model, medaka, using a medaka 8K cDNA array (Boswell et al. 2009). Using a similar experimental design, transcriptional response to cadmium and zinc in the resistant *Thlaspi caerulescens* was compared to that of the more sensitive model species *Arabidospsis thaliana* using the *Arabidopsis* 60-mer oligo array (Hammond et al. 2005). Fewer studies have directly compared gene expression between multiple heterologous species (but see examples below). It is this last type of experiment that will provide the most information with regard to ecologically and evolutionary relevant questions by allowing direct comparison of divergent species. The most common method of comparative expression profiling is the competitive hybridization of two samples to a cDNA microarray. While next-generation sequencing technologies also offer the possibility of measuring transcript abundance directly via sequencing of un-normalized cDNAs from the mRNA sample of interest (e.g. Weber et al. 2007; Marioni et al. 2008; Shin et al 2008), these approaches are only viable for fully sequenced model organisms or for species that are phylogenetically close to an organism for which full genome sequence is available as a reference (e.g. Toth et al. 2008) and even then can be cost-prohibitive (but see Turner et al. 2009). Therefore, cDNA microarrays remain the method of choice for comparative expression studies.

There are three basic methods by which a single spotted cDNA glass microarray platform can be used to compare gene-expression levels among multiple species through heterologous hybridization strategies (reviewed in detail by Kassahn 2008). Each suffers from its own set of difficulties and drawbacks and each requires additional controls or data processing.

(1) The first, and most common, heterologous hybridization strategy is implemented by carrying out two (or more) intra-specific analyses in parallel in order to generate, for each species separately, a list of genes considered to be differentially expressed between two or more phenotypes of interest. These lists of genes can then be compared across species to identify shared patterns of gene regulation. Studies on temperature acclimation in teleosts provide an example of this strategy by using the microarrray constructed with cDNA sequences from the temperate goby *Gillichthys mirabilis* (Gracey 2008) to study species of fish as diverse as blue fin tuna (Castilho et al. 2009) and Antarctic notothenioids (Buckley and Somero 2009). Through meta-analysis of these and other studies of thermal stress conducted with other microarray platforms, e.g. carp (Gracey et al. 2004; Williams et al. 2008), catfish (Ju et al. 2002), and reef fish (Kassahn et al. 2007), a conceptual model of response to stress can include genomic responses (Kassahn et al. 2009). Given sufficient replication, each hybridization ratio within species can be used to determine statistically significant differences in gene regulation between species. This approach has been employed to compare brain gene expression between nurses and foragers among four species of bees using a single *Apis melifera* microarray platform (Sen Sarma et al. 2007) and also to identify differences among species for tissue-specific gene expression in leafy plants (Horvath et al. 2003). However, this parallel strategy provides no information on the relative level of gene expression between species, only the relative level within species.

(2) The second strategy again requires that two (or more) species are analyzed in parallel. Here, however, a reference sample pool, possibly from the platform species, is used in the parallel experiments such that the expression ratios between species of interest are inferred transitively. In an interesting twist on this strategy, the often-cited work of Abzhanov et al. (2006) used RNA from a basal (outgroup) species in a comparative study of the development of the beak across four species of Darwin's finches (one of which was the platform species). Similarly, a recent study using a rat 60-mer oligo array, revealed the regulation of 'biotransformation genes' in woodrat liver in response to creosote toxin by employing a reference design in which the reference sample pool comprised RNA samples from the 16 individuals (two species) under study (Magnanou et al. 2009). The main drawback of this approach is reduced statistical power (the greatest amount of information is gained for the reference sample, which is not of direct interest).

(3) Finally, in the third strategy, samples from the two species of interest are competitively hybridized directly against each other using a microarray platform constructed from cDNA

sequences of a third species. Again, the difficulty of untangling hybridization effects due to sequence divergence from those due to true differences in gene expression is a major obstacle. The relative phylogenetic distance between each of the heterologous species and the platform species will introduce an additional concern when using this third strategy. This strategy was employed in earlier work to address the differential liver function of two species of woodrat during detoxification. Skopec and colleagues (2007) appropriately employed a within-and-between-species design competitively hybridizing samples from each species of woodrat to 60 mer oligo array designed from rat genome sequence. Similarly, by maintaining equivalent phylogenetic distance between the platform species (*Salmo salar*) and the heterologous species of interest (two limnetic eoctypes of *Coregonus clupeaformis*), Derome and Bernatchez were able to use a heterologous reference design in order to address the hypothesis that parallel phenotypic adaptations of the two limnetic ecotypes involved parallel transcriptional changes at the same genes (Derome and Bernatchez 2006).

Clearly, many variations in design are possible, including, but not restricted to, these three basic strategies. For example, researchers have employed multi-species arrays in both technical (comparative primate: Gilad et al. 2005) and experimental (oocyte evolution: Vallee et al. 2006) studies. Furthermore, we have not discussed the use of single channel technologies such as the Affymetrix arrays that have been used in many comparative studies (Enard et al. 2002; Ji et al. 2004) nor subtractive hybridization strategies (e.g. Wang and Brown 2006). Regardless of the experimental details, it should be clear from this short survey that whenever multiple species are competitively hybridized to a single microarray platform it is of fundamental importance to separate the effects of sequence divergence from the effect of true differences in gene expression. The combined effects of variation in transcript abundance and variation in sequence can result in both false positives and false negatives. False positives will result when the hybridization ratio is driven by substantial variation in sequence between the species. False negatives will result when the transcript abundance and variation is obscured by sequence divergence between the two species. False negatives can also result simply from poor hybridization, using any of the above strategies. It is important to identify these false negatives when the ultimate goal includes comparison across species, or analysis according to representation of functional categories.

Gilad and colleagues (Gilad et al. 2005; Oshlack et al. 2007) have tested the effect of sequence divergence on the ability to accurately detect gene regulation in experiments that employ heterologous hybridization. Importantly, they demonstrated that normalization of the array data can introduce a systematic bias when the species are of different phylogenetic distance relative to the platform species. Various approaches have been used to address this issue such as systematic adjustment based upon observed similarity between species (Ranz et al. 2003), or the filtering of datasets according to array feature characteristics (Bar-Or et al. 2007). However, these techniques do not identify the feature-specific biases present in any interspecific comparison. Here we use two species from the phenotypically diverse Ectodini clade of Lake Tanganyikan cichlids for the first comparative, and directly interspecific, analysis of neural gene-expression profiles in relation to mating strategy.

Case study: comparative analysis of the evolution of mating systems

In contrast to the uniformly polygynous and maternal mating strategy employed by the haplochromine cichlids from Lakes Malawi and Victoria, many of the cichlid lineages in Lake Tanganyika display a wide range of mating strategies (ploygynous, polyandrous, and monogamous) and provide parental care in a maternal, bi-parental, or cooperative manner either via buccal incubation (mouth brooding) or substrate guarding (for review see Barlow 2000). The monophyletic clade of Ectodine cichlids, in particular, is well suited for a genomic exploration into the evolution of social systems and social behavior. About 36 species shared a common ancestor only about 1.5 million years ago and there have been several independent transitions from polygyny to monogamy (Koblmüller et al. 2005). In this group, habitat and social organization correlate with differences in structure of the brain and in mating system (Pollen et al. 2007), as well as in visual behavior (Dobberfuhl et al. 2005). In the present study we begin to determine the molecular basis of variation in social organization. As is shown in Fig. 1, we can, in principle, identify sets of genes associated with sex, pairbonding (in the monogamous species) or lack thereof (in the polygynous species), or parental care.

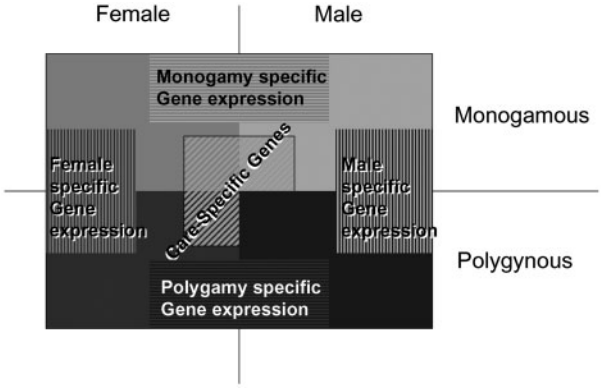

Fig. 1 Concept Map describing the expected shared and unique gene expression sets that underlie behavioral phenotypes. By exploring both sex and species, we can determine the molecular basis for different forms of social organization: genes related to pair bonding as well as those common to parenting among all three parental phenotypes (female monogamous, male polygynous, and male monogamous).

Methods

Gene expression: controlling species' behavior

We employed a loop design (Churchill 2002) to compare neural gene expression in reproductively active males and females from two sister species, *Xenotilapia flavipinnis* and *Enantiopus melanogenys*, that show different mating strategies. *X. flavipinnis* exhibit a monogamous mating strategy while *E. melanogenys* exhibit polygyny in a lek-like mating strategy. In order to confirm active reproductive stage, a group of *X. flavipinnis* males and females was observed in a common tank for 2 weeks. Males and females that interacted most frequently were separated from the group and qualitatively observed for 2 more weeks to verify species-specific behavior. Similarly, a group of *E. melanogenys* males and females was observed in a common tank for 4 weeks. Reproductive behaviors (e.g. sand-crater building, displays, and aggressive responses to other males) were observed for at least 2 weeks before dissection. Both sets of fish were housed in sandy habitats to model their shared wild habitat in Lake Tanganyika. The *E. melanogenys* were F1 laboratory-reared while the *X. flavipinnis* were wild-caught as juveniles in Nkamba Bay and reared in the laboratory.

Heterologous RNA hybridizations

Brains were removed within 9 min of the first disturbance of a tank. Dissected brains were placed in RNAlater solution (Ambion) within 3 min of death. All dissections were performed in the early afternoon between 2:30 and 4:00 pm to control for circadian effects on gene expression. Total RNA was extracted from brains using the TRIzol protocol (Invitrogen) following homogenization of brain tissue. RNA quality and concentrations were determined using the Bioanalyzer (Agilent).

For each sample, 2 μg of total RNA was reverse transcribed and labeled according to Renn et al. (2004), using oligodT (12–18) in a reverse transcription reaction with SuperScript II (Invitrogen) according to the manufacturer's protocol. This incorporates the amino-allyl dUTP (Sigma) to be dye-coupled with Cy3 or Cy5 CyDye Post-labeling Reactive Dye Pack (Amersham) following 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. Samples were competitively hybridized to the first generation *A. burtoni* microarray that contains 6000 brain cDNA features representing about 4000 unique genes (Renn et al. 2004) (GEO platform GPL928). A microarray loop design (Fig. 2) that incorporates dye swaps across phenotypes was used to control for any effects of dye while allowing for multiple comparisons using fewer microarrays (Churchill 2002).

Heterologous comparative genomic hybridization

Tissue samples from *X. flavipinnis* and *E. melanogenys* were previously stored in ethanol. Genomic DNA was extracted according to a standard ProteinaseK/Phenol protocol and quantified (Nanodrop 1000) for dilution to 0.25 μg/μl appropriate to size reduce by Hydroshear (Genome Solutions/Digilab) to an average of 1.0–1.5 kb, verified by gel electrophoresis. Genomic DNA samples from multiple individuals were combined and genomic DNA (3 μg) was fluorescently labeled with Alexa-Fluors conjugated dCTP by Klenow fragment polymerization (Invitrogen, Bio-Prime), the efficiency of which was quantified (Nanodrop 1000) such that competitive hybridizations were matched for concentration. Each heterologous species was labeled twice, once with each fluorophore, for a total of two direct competitive hybridizations. A dye-swap was included to account for dye bias and false positives due to individual variation in copy number (Redon et al. 2006). Hybridizations proceeded for ~16 h at 48°C in Ambion Hyb Buffer 1 (Ambion) blocked by Cot-1DNA (Invitrogen). The array used for these genomic

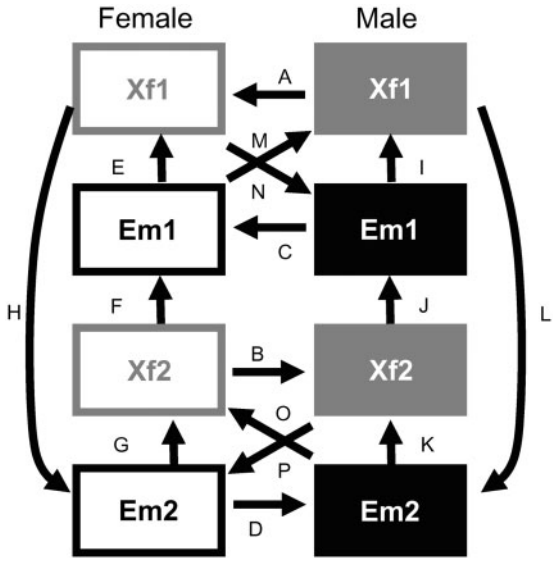

Fig. 2 Microarray comparisons within and across species. Em denotes *E. melanogenys* (black), Xf denotes, *X. flavipinnis* (gray), open boxes denote females, filled boxes denote males, 1 and 2 denote sample number. Each arrow represents one microarray with the base of the arrow denoting Cy3 and the head of the arrow denoting Cy5. The letters correspond to arrays as described in the text. Every individual was compared directly against two heterospecific animals of the same sex (i.e. also comparing mating systems) as well as against two individuals of the same sex, one of each species (also allowing to compare mating systems).

hybridizations (GEO platform GPL6416) contains the same brain cDNA library clones as were used in expression profiling.

Array scanning and LIMMA analysis

Hybridized arrays were scanned with an Axon 4000B scanner (Axon Instruments) using Genepix 5.0 software (Axon Instruments). Features of poor quality (fluorescence <2 SDs above background) were flagged and excluded. Raw data from Genepix were imported into R, and Linear Models for Microarray Data (LIMMA, Smyth 2005) was used to apply a background correction ('minimum') and within-array intensity normalization ('printtip loess'). All viable array features were used for normalization because the two species being hybridized are of similar phylogenetic distance relative to the platform species *A. burtoni*. When this is not the case, the normalization would be calculated based on the subset of genes known to be highly conserved, and then would be applied to all features on the array in order to avoid introduction of bias at this step (Gilad et al. 2005; Machado et al. in preparation). A linear model was fitted to the data using 'lmFit', and 'eBayes' provided an empirical Bayes shrinkage of probe-wise variation, borrowing information across genes (Smyth 2004). Unless otherwise noted, a significance threshold of $P<0.01$ without multiple testing correction was used, the low stringency of which allowed us to examine general patterns.

Although the experimental design relies upon a full loop of interconnected comparisons (Fig. 2), we chose to use results only from those comparisons for which the number of samples/arrays of direct interest was less than half of the number of total arrays. For these analyses with only two or four arrays, we do not differentiate between biological and technical replicates. Only when statistical analyses were applied to all 16 arrays in the full loop to identify overall sex-biased, species-biased, or pair-wise phenotypic comparisons were both biological and technical replication accommodated in the statistical model.

Meta-analysis

We used data from two previous studies performed with the *A. burtoni* cDNA array platform (*N. pulcher*: Aubin-Horth et al. 2007; *A. burtoni*: Renn et al. 2008) to compare sex-specific gene expression with the patterns found in *X. flavipinnis* and *E. melanogenys* (present study). The *A. burtoni* dataset consisted of comparisons of 36 arrays between the sexes or between male reproductive phenotypes (territorial and non-territorial), with biological replication. The *N. pulcher* dataset consisted of six microarrays, all of which directly compared adult breeding males and females, with biological replication. These datasets were reanalyzed according to sex, using the same normalization and testing procedures applied to the current *X. flavipinnis* and *E. melanogenys* datasets.

Genomic hybridization analysis and masking procedure

For the detection of differential hybridization of gDNA between *X. flavipinnis* and *E. melanogenys*, the same normalization and testing procedures were applied to the dye-swap pair of competitive genomic hybridizations. Array features for which the log hybridization ratio differed from 0 at significance of $P<0.01$ and then at $P<0.05$ were chosen for removal from the expression dataset to provide a genomic hybridization mask that could eliminate potential false positives and negatives that were due to greater sequence divergence or gene duplication in one species over another. One measure of the effect

of these masks was the comparison of Pearson correlation coefficients of the genomic and expression hybridization ratios (log2) of sets of significantly and non-significantly regulated genes before masking, after the first masking (with $P<0.01$), and after the second masking (with $P<0.05$). A final pair-wise expression analysis was performed on the second masked expression dataset for each combination of phenotypes (*X. flavipinnis* male, *X. flavipinnis* female, *E. melanogenys* male, and *E. melanogenys* female), using the loop design with biological replication.

Results

Sex-specific gene expression

For each species, two direct comparisons were available for the analysis of sex-specific gene expression. For this between-sex, within-species comparison, 4527 features survived for *X. flavipinnis* (Fig. 2: arrays A and C) and 4488 features survived filtering for *E. melanogenys* (Fig. 2: arrays B and D). These data allow a parallel intra-specific interrogation strategy, described above, as the first and most common heterologous hybridization approach. When the full complement of microarray comparisons was interrogated for sex-specific gene regulation, 4289 features survived in sufficient number of arrays to be analyzed for overall sex-specific regulation, regardless of species.

Overall, there were more male-enriched genes (80) than female-enriched genes (77) when the full complement of microarray comparisons was considered. This trend was reversed when the species were considered independently. The monogamous species, *X. flavipinnis*, showed only three male-enriched, but 46 female-enriched features ($P<0.01$). The polygynous species, *E. melanogenys*, showed a lesser number of sex-specific genes overall, but the same trend in that only six male-enriched features were identified but nine female-enriched features were identified ($P<0.01$). Figure 3 shows these results in the form of Venn diagrams. The majority of the sex-specific gene regulation is not shared between species, as demonstrated by the fact that there was almost no intersection for the lists of sex-specific regulation in each species when they were analyzed independently. This result is perhaps surprising, given the ability to detect a substantial amount of sex-specific regulation when the full complement of microarray comparisons was considered. We interpret this result to mean that while there is considerable sex-specific gene regulation, the most robust sex-specific gene regulation (identified even with low power of only two arrays) is species-specific.

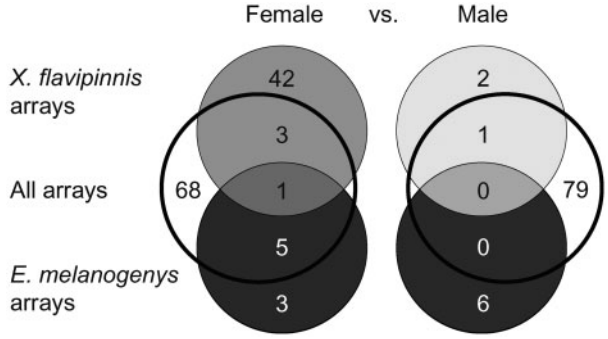

Fig. 3 Venn diagram for sex-specific gene expression in each species. The full loop analysis is included in the black circle. The set intersections show the number of genes that were identified as sex-enriched in more than one analysis.

In order to investigate the relative expression level and statistical significance for sex-associated expression in the absence of stringent thresholding, we plotted all sex-specific genes from one species for a very relaxed threshold of $P<0.05$ onto the volcano plot for the other species (Fig. 4). Interestingly there is little overall conservation of sex-specific gene expression between these two closely related species. A great number of genes that were male-enriched in one species showed a (non-significant) female-biased expression ratio in the other species and *vice versa*.

In order to carry the parallel intra-specific meta-analysis one step further and identify a core set of male-enriched and female-enriched genes, we interrogated the current results with those from two previous studies of sex-specific neural gene expression in cichlids. When we applied the same statistical treatment to those datasets as was applied to the current one, we identified 730 (of 3951) features that are enriched in either male or female *A. burtoni*, a lek-breeding species more similar to *E. melanogenys* (Fig. 5). We then plotted these sex-specific genes onto the expression-analysis results for the other available robust adult phenotype dataset for *N. pulcher*, a cooperatively brooding species more similar to *X. flavipinnis* (Aubin-Horth et al. 2007). As these two studies represent divergent mating strategies from distantly related species, we wanted to know whether we could identify a signature of species-invariant and lineage-invariant (i.e. conserved) sex-specific gene regulation. We found remarkably little conservation (less than 50%) for sex-specific gene regulation across species. Nonetheless, 177 male-biased features and 145 female-biased features were conserved across species (regardless of statistical significance in the *N. pulcher* analysis). We then plotted these conserved sex-specific

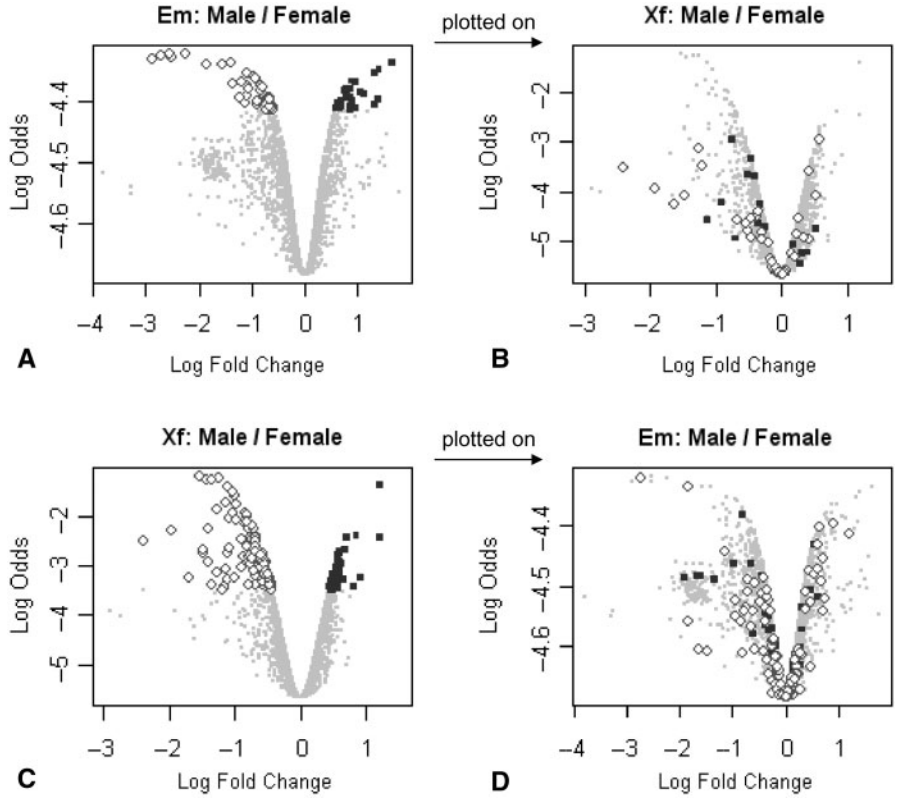

Fig. 4 Volcano plots of sex-specific expression within *E. melanogenys* (Em) and *X. flavipinnis* (Xf). Genes with significant ($P < 0.05$) male-specific (filled squares) and female-specific (open circles) expression are indicated within species (**A** and **C**) and then plotted on the analysis for the other species (**B** and **D**).

features onto the current polygynous (*E. melanogenys*) and monogamous (*X. flavipinnis*) gene-expression profiles as analyzed according to sex (Fig. 5). Again, the results show a remarkable lack of sex-specific conservation. In *E. melanogenys*, 51% of the sex-regulated genes were part of this conserved set (103 in male and 61 in female). In *X. flavipinnis*, 57% of the sex-regulated genes were conserved (110 in male and 73 in female). Across all four species, male-biased expression was conserved for only 57 array features (~50 expected by chance) and female-biased expression for 29 array features (~40 expected by chance) (see Supplementary Material Online Table 1). We consider these genes to represent the lineage-invariant and conserved core set of sex-regulated genes. There is no *a priori* reason to expect this core set to be greater or less than the number expected by chance.

An individual's neural gene expression pattern is influenced by many factors other than sex or overall genotype. Thus, only a portion of the overall pattern is expected to be directly related to the behavioral phenotype of interest (such as mating system). Our meta-analysis that compares the present data set with published intra-specific expression studies provides a powerful assessment of how similar transcript profiles are across species with similar social systems. For example, sex-specific expression profiles in the monogamous and biparental *X. flavipinnis* overlap substantively with those in the cooperatively brooding *N. brichardi* (24% of regulated genes show concordant sex-bias at a significance threshold of $P < 0.05$), yet show much less overlap when compared with the polygynous and maternal *A. burtoni* (9%). In comparison, sex-specific gene expression in the polygynous and maternal *E. melanogenys* is no more concordant with *A. burtoni* (7%) than with *N. brichardi* (10%), as one would expect if mating system alone is driving sex-specific gene expression. Instead, despite similarities in mating system, the males and females of these species exhibit dramatic differences in gene expression.

Mating-system-enriched gene expression

Our study also directly compared male *X. flavipinnis* (monogamous) with male *E. melanogenys* (polygynous) (Fig. 2 arrays I, J, K, L) and female *X. flavipinnis* (monogamous) with female *E. melanogenys* (ploygynous) (Fig. 2 arrays E, F, G, H) as an example of the third strategy for heterologous

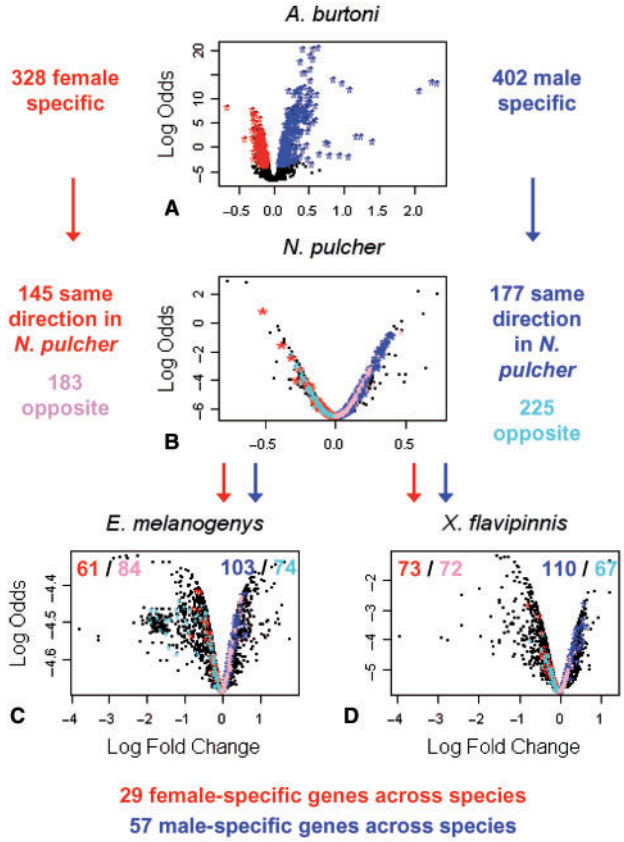

Fig. 5 Volcano plots of sex-specific expression as determined by a meta-analysis. (**A**) *A. burtoni*, sex-regulated genes (red: female, blue: male) ($P<0.01$) are plotted onto the (**B**) *N. pulcher* analysis to indicate those features that do not show conservation of sex-specific expression across species (pink and cyan) and those whose sex-specific expression is conserved (red and blue). The conserved set of sex-specific genes is then plotted onto current expression results for (**C**) *X. flavipinnis* and (**D**) *E. melanogenys*.

Table 1 Pair-wise comparison of differences in gene expression between each phenotype in this study ($P<0.01$)

	X. flav male	*X. flav* female	*E. mel* male	*E. mel* female
***X. flav* male**		61	417	164
***X. flav* female**	53		290	203
***E. mel* male**	411	392		124
***E. mel* female**	289	405	203	

Numbers indicate the number of genes with increased expression in the row-labeled phenotype compared with the column-labeled phenotype.

comparative expression profiling described above. Interestingly, we found a greater number of genes to be regulated by species (mating-system) than by sex. This is likely in part due to the increased statistical power moving from a single dye-swapped

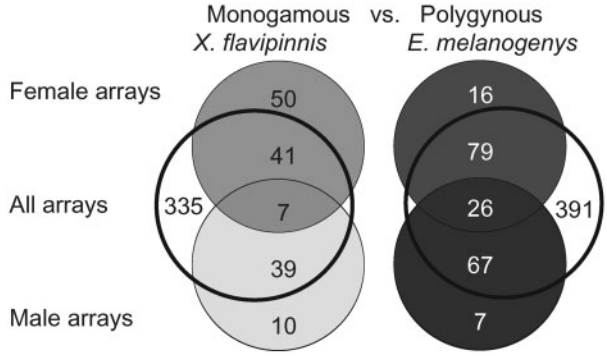

Fig. 6 Venn diagram for species (mating-system)-specific gene expression. For each sex, the monogamy-enriched set is indicated by lighter tones and the polygyny-enriched set by darker tones, while the full loop analysis is included in the black circle. The set intersections show the number of genes identified as species-enriched in more than one analysis.

pair of microarrays for the analysis of sex-specific gene expression (above) to a total of four microarrays for the analysis of mating style by comparing species (Clark and Townsend 2007). For this comparison between species (mating-system) and within each sex, 4624 features survived filtering for the analysis of males and 4605 features survived for the analysis of females. When the full complement of arrays was considered, 4389 features survived in sufficient number of arrays to be analyzed for overall species (mating-system)-specific regulation regardless of sex. Overall there were more *E. melanogenys*-enriched (polygyny) features (563) than *X. flavipinnis*-enriched (monogamy) features (421) when the full complement of microarrays was considered. This trend held for both sexes analyzed independently. While the analysis of females showed 118 *E. melanogenys* (polygyny)-enriched features and 104 *X. flavipinnis* (monogamy)-enriched features ($P<0.01$), the analysis of males showed a smaller number of species (mating-system)-specific genes with 99 *E. melanogenys* (polygyny)-enriched features and 57 *X. flavipinnis* (monogamy)-enriched features. The intersection of *X. flavipinnis* (monogamy)-enriched gene sets from the two sexes was seven, and the intersection of *E. melanogenys* (polygyny)-enriched gene sets was 26 ($P<0.01$). Interestingly, there was a greater intersection between the lists of genes specific to species (mating-system) derived from each sex (Fig. 6) than there had been between the lists of sex-specific genes derived from each species (Fig. 3).

In order to detect global patterns, we investigated the relative expression level and statistical significance for regulation between species by mapping all

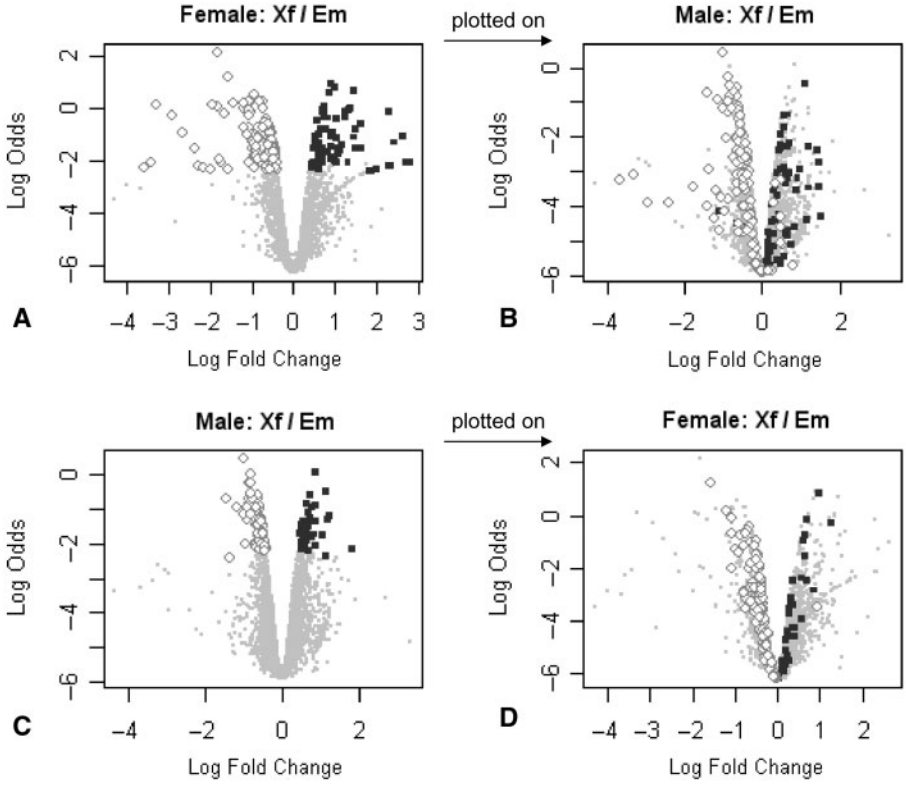

Fig. 7 Volcano plots of species (mating-system)-specific expression for each sex. Genes with significant ($P<0.05$) *X. flavipinnis*-specific (monogamous; filled squares) and *E. melanogenys*-specific (polygamous; open circles) expression are indicated within a sex (**A** and **C**) and then plotted on the analysis for the opposite sex (**B** and **D**).

species (mating-system)-specific genes from one sex onto the volcano plot for the other sex (Fig. 7). We find that the species-bias is conserved between the comparison of males and the comparison of females. The majority of the features that show a statistically significant species (mating-system) bias in the analysis among one sex do maintain the same directional bias in the analysis of arrays from the other sex.

It must be noted that in the above analyses, the results that are interpreted as enrichment of gene regulation according to mating system are confounded by any interspecific differences, including both biological differences not related to mating as well as DNA sequence. Because hybridization ratios may be driven by species differences in sequence similarity relative to *A. burtoni* causing a species bias in hybridization efficiency, we performed genomic hybridizations in order to address and quantify the impact of sequence bias.

Heterologous genomic hybridizations

We performed a single pair of dye-swap heterologous gDNA hybridizations that directly competed genomic DNA from the two focal species, *X. flavipinnis* and *E. melanogenys*. A total of 4624 features survived filtering. We applied the same statistical test to these hybridization ratios in order to identify features for which the genomic hybridization ratio was statistically different from the expected equal hybridization. Such a result would suggest the influence of sequence divergence. Unlike a simple competition between the platform species and one heterologous species, array features can be biased in either direction. Because we are using gDNA from two different heterologous species, the hybridization ratio depends on the similarity of sequences between the platform species and each of the heterologous species. When these genomic hybridization ratios are tested for statistical significance, at a rather stringent threshold ($P<0.01$), there was a total of 49 genes with significant genomic hybridization bias, 28 biased toward *X. flavipinnis* and 21 biased toward *E. melanogenys*; at a less stringent threshold ($P<0.05$), there were 273 features that showed significant genomic hybridization bias, 131 biased toward *X. flavipinnis* and 142 biased toward *E. melanogenys*.

While there was one gene that showed species-specific expression in each sex, yet in opposite

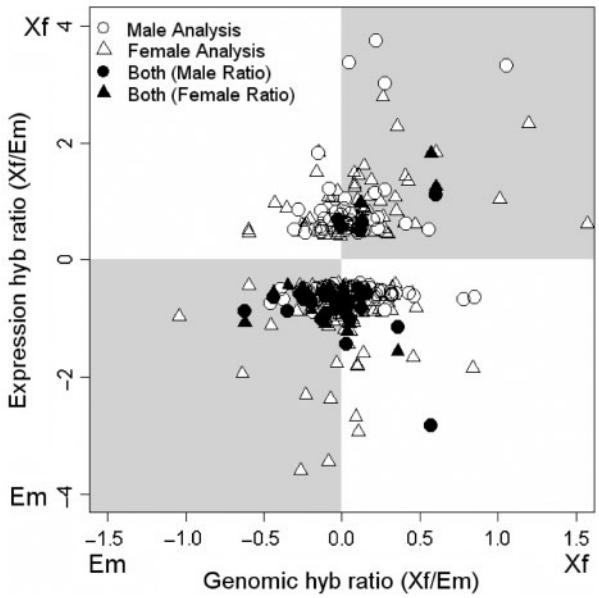

Fig. 8 Correspondence between hybridization ratio (log_2) from the expression arrays (*Y*-axis) compared to the genomic hybridization ratio (log_2) (*X*-axis) for all genes significantly regulated between species analyzed for each sex separately: males (circles), females (squares). Genes found regulated in both analyses are represented by filled symbols. Positive log ratios indicate *X. flavipinnis*-bias (Xf) and negative log ratios indicate *E. melanogenys*-bias (Em), such that gray quadrants contain the genes whose genomic hybridization ratios lie in the same direction as the expression hybridization ratios.

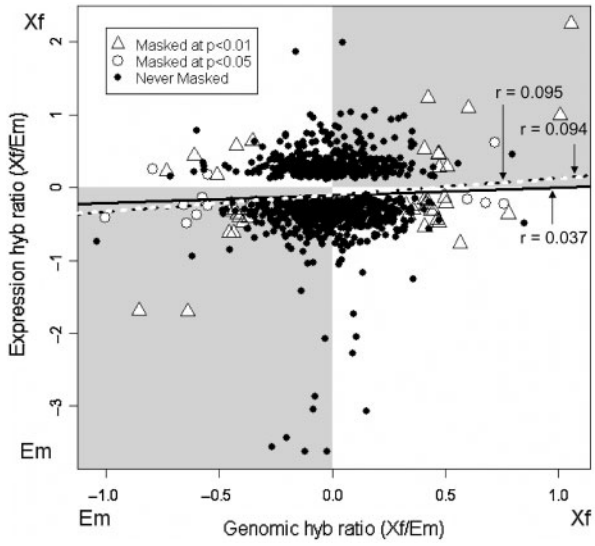

Fig. 9 Correspondence between genomic hybridization ratios and cDNA hybridization ratios for genes that showed statistically significant differences in expression between *X. flavipinnis* (Xf: positive ratios) and *E. melanogenys* (Em: negative ratios) ($P<0.01$; 945 features). Genes with significant differential gDNA hybridization to the microarray are 'masked' by a more liberal (white triangles $P<0.01$, 11 genes masked) and a more conservative (white circles $P<0.05$, 54 genes masked) genomic hybridization mask. The correlation between genomic hybridization ratios and expression hybridization ratios is shown with a regression line for before (white dotted line) and after masking at $P<0.01$ (black dotted line) and $P<0.05$ (black) with corresponding correlation coefficients.

directions, as it has greater expression in monogamous *X. flavipinnis* females than in polygynous *E. melanogenys* females, but has decreased expression in *X. flavipinnis* males compared to *E. melanogenys* males, there were 31 genes that were found to be species (mating-system)-specific in both the analysis of males and of females (indicated in black in Fig. 8). It was these features for which we were most concerned about a strong influence of sequence divergence. In order to visualize any possible bias and interrogate specifically those array features that figure prominently in our results, we analyzed the 304 array features that showed statistically significant regulation of gene expression between species when analyzed within males or within females ($P<0.01$) (Fig. 8). The fact that, overall, genes that were significantly regulated according to species (mating-system) were also more likely to show genomic hybridization bias in the direction of that species (56% of 325; exact binomial test $P=0.026$) (found in the gray area in Fig. 8) suggests that sequence divergence does influence expression profiling, and therefore must be taken into account in some way.

Building a hybridization mask

In order to determine the extent to which the results described above had been confounded by sequence divergence, we asked whether the expression ratios were correlated with the genomic hybridization ratios. A strong correlation would suggest that measurement of expression level was highly confounded by sequence variation. We found a slight ($r=0.033$) but nearly significant correlation ($P=0.066$) for the two hybridization ratios among the array features that were not found to be regulated between species (mating-systems) (3305 features) when all 16 microarrays were analyzed. We take this value to represent the baseline effect for the two genomes of interest. The same test for correlation, applied to the 945 array features that, according to this analysis of the full complement of microarrays, showed species (mating-system)-specific gene regulation, resulted in a correlation coefficient of 0.095 ($P=0.004$) (Fig. 9). This increased correlation between genomic hybridization ratio and expression ratio indicates a weak, yet statistically significant impact of sequence divergence. When we applied the first genomic mask to

the analysis of gene expression by eliminating all array features that showed significant genomic hybridization bias at a threshold of $P<0.01$ (49), this resulted in the removal of the 43 biased features (11 significantly regulated and 32 not regulated) that also survived initial quality controls for the expression analysis. There was little change in correlation coefficients for either the non-significantly regulated genes ($r=0.046$, $P=0.01$) or the remaining set of 918 regulated array features ($r=0.094$, $P=0.004$). When we applied the second, more stringent, mask by eliminating all array features that showed significant bias in genomic hybridization at a threshold of $P<0.05$, this resulted in the elimination from the expression analysis of 254 biased features (54 significantly regulated and 200 not regulated), as again not all of the 273 sequence-biased features survived the initial quality control for the expression analysis. Importantly, under this more stringent mask, there was a considerable reduction in the strength and significance of the correlations between genomic hybridization and expression ratios for the significantly regulated genes ($r=0.037$, $P=0.271$) (Fig. 9). This indicates a dramatic reduction in the number of potential false positives, i.e. genes whose expression hybridization ratios are more strongly influenced by genomic characteristics, such as level of sequence similarity, rather than representing true expression.

After applying the more stringent genomic hybridization mask we reassessed the gene expression profiles. Of the 31 genes that were identified as species (mating-system)-specific in both the analysis within males and the analysis within females, only three were eliminated by the genomic mask, suggesting that the observed species-bias in expression ratio represents parallel gene regulation in both males and females of the species possibly related to the mating-system. We then wanted to identify all genes that are differentially regulated between any two phenotypes. With the genomic mask applied in order to reduce false positives that are due to sequence divergence, and taking advantage of the full loop design, we conducted a pair-wise comparison among the different sexes and phenotypes (Table 1). We found that the polygynous males (*E. melanogenys*) showed the highest degree of phenotype-specific gene expression in all pair-wise comparisons. This divergence at the molecular level parallels the dramatically different behavioral phenotype for the polygynous male relative to the other three phenotypes studied.

Interestingly, several genes show expression regulation opposite to the genomic hybridization ratios (i.e. they fall in the white quadrant in Fig. 8) (see Supplementary Material Online Table 2). For this set of genes, there is a possibility that the sequence divergence is buffering the relative expression, such that expression differences are underestimated; those not found to be significantly regulated are possible false negatives, and those found to be regulated might be differentially expressed to an even greater extent than indicated by the data. If confirmed, such reversals in expression or inverse relationships between aCGH and mRNA expression level might provide interesting additional insights into the evolution of mating systems. Overall, the masking procedure identified 54 features that were potential false positives in the interspecies analysis. This represents 6% of the genes initially found to be regulated between species. This case study provides a proof of concept demonstrating how this novel approach allows the elimination of false positives and identification of those features that provide reliable data in the given species comparison.

Discussion

In this article, we first described the state of the genomic resources available for the powerful cichlid model of adaptive radiation. We then reviewed the challenges associated with heterologous hybridizations as an important tool for comparative functional genomics and provided several examples of attempts to address these challenges. Finally, we presented a method by which a detailed analysis of CGH to a cDNA microarray can inform transcriptome analysis across species.

With the focus of our proof of concept study on the use of CGH for validation and quality control, the sample sizes available for transcriptome analysis were necessarily small. While increased sampling will no doubt increase statistical power considerably, especially when comparing multiple independent transitions in the mating systems of Ectodines, we have already found some interesting patterns. One tantalizing insight from this part of our study is the suggestion that the difference between these two species representing monogamy and polygyny—in terms of neural gene expression—is not simply the behavior of the males (pairbonding and paternal in one species, solitary and displaying a lack of care in the other), which is in fact the defining feature at the organismic level. Rather, while *E. melanogenys* males do show the greatest number of significantly regulated genes according to the masked pair-wise comparison, it is also the females that show substantial differences in gene expression profiles, possibly related to the difference

in behavior necessitated by the presence or absence of a mate and fellow caregiver. Nevertheless, when we analyzed *E. melanogenys* expression profiles according to sex (Fig. 3), we found that all six of the genes overrepresented in the (polygynous) males were specific to this phenotype, whereas most (66%) of the genes overrepresented in the females of this species were also found in the overall comparison between males and females. In contrast, the situation appears to be reversed in the monogamous *X. flavipinnis*, in which 90% of the genes overrepresented in the females were specific to this phenotype (there were too few male-specific genes to allow comparison to the polygynous males). One tentative interpretation of these results is that under the derived monogamous conditions both female-specific and male-specific patterns of gene expression have diverged substantially from the ancestral polygynous expression patterns. Ongoing experiments that examine more than one of the independent transitions from polygynous to monogamous mating systems among the Ectodini cichlid clade, will clarify this intriguing possibility by removing the confounding factor of species, which is inherent to the present study.

Our finding that sex-specific gene expression was highly variable across species indicates that social organization, such as mating system, may play an important role in sculpting transcription profiles in the brain. However, our results suggest that there are core sets of genes whose expression is coordinated across species. Future studies comparing more species will provide us with a better understanding of how these gene sets relate to social phenotypes.

Overcoming technical challenges for the analysis of interspecies heterologous hybridization

While many studies filter expression data from heterologous hybridization based on intensity of hybridization during expression profiling itself (e.g. Rise et al. 2004), several technical studies have now shown that mere hybridization is insufficient for identifying features that are available to give accurate and robust results (Renn, et al. 2004). Similarly, a single hybridization using genomic DNA from the heterologous species may reveal the percentage of array features that are sufficiently conserved to give hybridization signals for potential use with experiments that employ intra-specific heterologous hybridization (e.g. Graham et al. 2007). However, a more rigorous measure of array utility includes a competitive genomic DNA hybridization between the heterologous species and the platform species in order to estimate the reliability of each feature (Kassahn, et al. 2007). Increased sequence divergence correlates with skewed hybridization ratios and therefore provides a better indicator of feature performance under expression profiling conditions.

For studies that employ interspecific heterologous hybridization, the competitive genomic DNA hybridization between the two different heterologous species onto the array platform is crucial in order to identify those features that will be substantially biased by sequence divergence. Without identifying the extent to which this effect will confound the expression analysis it is impossible to produce reliable results that do not suffer from a marked increase in false positives that result from, or have been augmented by, a bias in DNA sequence. A global correction factor (that is applied when an estimate is available for the two species of interest) does not account for variation in sequence divergence—and potential differences in hybridization dynamics—among individual genes/features. The physicochemical properties of the dynamics of nucleic acid hybridization constitute a challenging topic (Edwards-Ingram et al. 2004; Bar-Or et al. 2007), yet—as we have shown here—this does not prevent us from identifying features whose hybridization signals are affected by sequence divergence. In the present case of East African cichlids, the number of genes/features affected by sequence divergence is low. We describe an iterative process to determine the statistical threshold that provides an appropriate mask based upon genomic hybridization results that, when applied to an expression profiling experiment, is sufficient to reduce the number of false positives. In the current study, the correlation coefficient between genomic hybridization ratio and gene expression ratio, based upon genes not determined to be significantly regulated, is used as a baseline measure of hybridization bias influenced by sequence variation. We then increased the stringency of the genomic hybridization mask in order to reduce, to the baseline level, the correlation coefficient for those array features that are determined to be significantly regulated. The exact *P*-value threshold used either for determination of significant regulation or for the mask will depend upon the statistical power of each experiment (Clark and Townsend 2007) and also upon the inferences to be drawn from the results.

A subset of those microarray features that are eliminated by these genomic hybridization masks actually show expression bias (even significant expression bias) in a direction opposite to that observed in the genomic hybridization bias. These features likely represent genes for which the

difference in species-specific gene expression is profound, sufficiently so to counter (and even overcome) the effects of genomic hybridization. One would obviously attend to such individual results. Furthermore, these results highlight the fact that, in any experiment that uses heterologous hybridization, there are also false negatives due to genomic hybridization bias. Any feature that shows a strong genomic hybridization bias in one direction, yet a subthreshold expression bias in the opposite direction, may actually represent a gene that is differentially regulated between the species of interest. The existence of such false negatives is intriguing, as it suggests that genomic hybridization ratios could be used as covariates in a more complex statistical analysis of experiments that employ interspecific heterologous hybridization. While one could imagine a feature-by-feature correction factor that either subtracted, multiplied or otherwise adjusted the expression ratio based upon the genomic hybridization ratio, such a technique would require substantial rigorous development and validation in a species group for which genome sequence information is known and species-specific expression could be validated through alternate methods. Therefore, given our poor understanding of the relationship between sequence divergence and hybridization dynamics (Edwards-Ingram et al. 2004; Bar-Or et al. 2007), not to mention the very different nature of genomic DNA sequences versus mRNA sequences (e.g. presence of introns, total length, repetitive sequences), such a complex approach is currently not advisable in any non-traditional model system.

Conclusions

The recent advances in sequencing technology continue to lower the cost of obtaining large amounts of genomic, or coding-region specific sequence information (Pihlak et al. 2008; Shendure et al. 2008; Turner et al. 2009) which can then be used to construct species-specific microarray platforms (e.g. Glanville fritillary butterfly: Vera et al. 2008; axolotl: Cotter et al. 2008). While this may be a cost-effective option for some studies that involve repeated measures within a small number of species, the construction of a unique array would still be prohibitive for large comparative studies that address the level of gene expression across a greater number of species. We therefore posit that heterologous hybridizations to existing array platforms will continue to increase in prominence for transcriptome analysis across species, especially in important model systems, for studies in ecology, evolution, and behavior. The use of interspecific aCGH to mask analysis of gene expression provides an important tool for such studies of gene expression in that this novel technique provides increased accuracy for the interpretation of comparative analysis of gene expression. This approach should be applicable to a wide range of interesting questions related to the evolution and ecology of gene expression.

Supplementary material

Supplementary material is available at ICB online.

Funding

This research was supported by an NRSA post-doctoral fellowship, the Murdock Life Trust Foundation and National Institutes of Health grant NIGMS R15GM080727 (S.C.P.R.), and by National Institutes of Health grant NIGMS GM068763, the Bauer Center for Genomics Research, the Alfred P. Sloan Foundation and the Institute for Cellular & Molecular Biology (H.A.H.).

Acknowledgments

We are grateful to Josiah Altschuler and Melinda Snitow for animal care, Sarah Bahan for assistance with the behavioral experiments.

References

Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. 2006. The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. Nature 442:563–7.

Aubin-Horth N, Desjardins JK, Martei YM, Balshine S, Hofmann HA. 2007. Masculinized dominant females in a cooperatively breeding species. Mol Ecol 16:1349–58.

Bar-Or C, Czosnek H, Koltai H. 2007. Cross-species microarray hybridizations: A developing tool for studying species diversity. Trends in Genetics 23:200–7.

Barlow GW. 2000. Cichlid Fishes: Nature's Grand Experiment in Evolution. Harper Collins Publishers.

Blaxter M, Whitton C, Thompson M, Daub J, Guiliano D, Stirton M, Jieru Y, Aboobaker A, Parkinson J. 2004. Comparative nematode genomics. Proceedings of the Fourth International Congress of Nematology 2:557–71.

Boswell MG, Wells MC, Kirk LM, Ju Z, Zhang Z, Booth RE, Walter RB. 2009. Comparison of gene expression responses to hypoxia in viviparous (*Xiphophorus*) and oviparous (*Oryzias*) fishes using a medaka microarray. Com Biochem Physiol C-Toxicol Pharmacol 149:258–65.

Buckley BA. 2007. Comparative environmental genomics in non-model species: Using heterologous hybridization to DNA-based microarrays. J Exp Biol 210:1602–6.

Buckley BA, Somero GN. 2009. cDNA microarray analysis reveals the capacity of the cold-adapted Antarctic fish

Trematomus bernacchii to alter gene expression in response to heat stress. Polar Biology 32:403–15.

Carleton KL, Spady TC, Streelman JT, Kidd MR, McFarland WN, Loew ER. 2008. Visual sensitivities tuned by heterochronic shifts in opsin gene expression. BMC Biol 6:22.

Castilho PC, Buckley BA, Somero G, Block BA. 2009. Heterologous hybridization to a complementary DNA microarray reveals the effect of thermal acclimation in the endothermic bluefin tuna (*Thunnus orientalis*). Mol Ecol 18:2092–102.

Churchill GA. 2002. Fundamentals of experimental design for cDNA microarrays. Nat Genet 32:490–5.

Clark TA, Townsend JP. 2007. Quantifying variation in gene expression. Mol Ecol 16:2613–6.

Cotter JD, Storfer A, Page RB, Beachy CK, Voss SR. 2008. Transcriptional response of Mexican axolotls to *Ambystoma tigrinum* virus (ATV) infection. BMC Genomics 9.

Cummings ME, Larkins-Ford J, Reilly CRL, Wong RY, Ramsey M, Hofmann HA. 2008. Sexual and social stimuli elicit rapid and contrasting genomic responses. Proc R Soc Lond Ser B-Biol Sci 275:393–402.

Derome N, Bernatchez L. 2006. The transcriptomics of ecological convergence between 2 limnetic coregonine fishes (*Salmonidae*). Mol Biol Evol 23:2370–8.

Dobberfuhl AP, Ullmann JFP, Shumway CA. 2005. Visual acuity, environmental complexity, and social organization in African cichlid fishes. Behav Neurosci 119:1648–55.

Edwards-Ingram LC, Gent ME, Hoyle DC, Hayes A, Stateva LI, Oliver SG. 2004. Comparative genomic hybridization provides new insights into the molecular taxonomy of the *Saccharomyces sensu stricto* complex. Genome Res 14:1043–51.

Enard W, Khaitovich P, Klose J, Zollner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Paabo S. 2002. Intra- and interspecific variation in primate gene expression patterns. Science 296:340–43.

Gilad Y, Rifkin SA, Bertone P, Gerstein M, White KP. 2005. Multi-species microarrays reveal the effect of sequence divergence on gene expression profiles. Genome Res 15:674–80.

Gracey AY. 2008. The *Gillichthys mirabilis* Cooper array: a platform to investigate the molecular basis of phenotypic plasticity. J Fish Biol 72:2118–132.

Gracey AY, Fraser EJ, Li WZ, Fang YX, Taylor RR, Rogers J, Brass A, Cossins AR. 2004. Coping with cold: An integrative, multitissue analysis of the transcriptome of a poikilothermic vertebrate. Proc Natl Acad Sci USA 101:16970–5.

Graham NS, Broadley MR, Hammond JP, White PJ, May ST. 2007. Optimising the analysis of transcript data using high density oligonucleotide arrays and genomic DNA-based probe selection. BMC Genomics 8:344.

Hammond JP, Broadley MR, Craigon DJ, Higgins J, Emmerson ZF, Townsend HJ, White PJ, May ST. 2005. Using genomic DNA-based probe-selection to improve the sensitivity of high-density oligonucleotide arrays when applied to heterologous species. Plant Methods 1:10.

Hofmann HA. 2003. Functional genomics of neural and behavioral plasticity. J Neurobiol 54:272–82.

Horvath DP, Schaffer R, West M, Wisman E. 2003. *Arabidopsis* microarrays identify conserved and differentially expressed genes involved in shoot growth and development from distantly related plant species. Plant J 34:125–34.

Ji W, Zhou WL, Gregg K, Yu N, Davis S, Davis S. 2004. A method for cross-species gene expression analysis with high-density oligonucleotide arrays. Nucleic Acids Res 32:e93.

Ju Z, Dunham RA, Liu Z. 2002. Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation. Mol Genet Genomics 268:87–95.

Kassahn KS. 2008. Microarrays for comparative and ecological genomics: beyond single-species applications of array technologies. J Fish Biol 72:2407–34.

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. Mol Ecol 16:1749–63.

Kassahn KS, Crozier RH, Portner HO, Caley MJ. 2009. Animal performance and stress: responses and tolerance limits at different levels of biological organisation. Biol Rev 84:277–92.

Kijimoto T, Watanabe M, Fujimura K, Nakazawa M, Murakami Y, Kuratani S, Kohara Y, Gojobori T, Okada N. 2005. cimp1, a novel astacin family metalloproteinase gene from East African cichlids, is differentially expressed between species during growth. Mol Biol Evol 22:1649–60.

Koblmüller S, Duftner N, Katongo C, Phiri H, Sturmbauer C. 2005. Ancient divergence in bathypelagic Lake Tanganyika deepwater cichlids: Mitochondrial phylogeny of the tribe *Bathybatini*. J Mol Evol 60:297–314.

Kocher TD. 2004. Adaptive evolution and explosive speciation: The cichlid fish model. Nat Rev Genet 5:288–98.

Loh Y-HE, Katz LS, Mims MC, Kocher TD, Yi SV, Streelman JT. 2008. Comparative analysis reveals signatures of differentiation amid genomic polymorphism in Lake Malawi cichlids. Genome Biol 9:R13.

Magnanou D, Malenka JR, Dearing MD. 2009. Expression of biotransformation genes in woodrat (*Neotoma*) herbivores on novel and ancestral diets: Identification of candidate genes responsible for dietary shifts. Mol Ecol 18:2401–14.

Marioni JC, Mason CE, Mane SM, Stephens M, Gilad Y. 2008. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. Genome Res 18:1509–17.

Nuzhdin SV, Wayne ML, Harmon KL, McIntyre LM. 2004. Common pattern of evolution of gene expression level and protein sequence in *Drosophila*. Mol Biol Evol 21:1308–17.

Oshlack A, Chabot AE, Smyth GK, Gilad Y. 2007. Using DNA microarrays to study gene expression in closely related species. Bioinformatics 23:1235–42.

Pihlak A, Bauren G, Hersoug E, Lonnerberg P, Metsis A, Linnarsson S. 2008. Rapid genome sequencing with short universal tiling probes. Nat Biotech 26:676–84.

Pollen AA, Dobberfuhl AP, Scace J, Igulu MM, Renn SCP, Shumway CA, Hofmann HA. 2007. Environmnetal complexity and social organization sculpt the brain in lake Tanganyikan Cichlid Fish. Brain Behav Evol 70:21–39.

Ranz JM, Castillo-Davis CI, Meiklejohn CD, Hartl DL. 2003. Sex-dependent gene expression and evolution of the *Drosophila* transcriptome. Science 300:1742–5.

Renn SCP, Aubin-Horth N, Hofmann HA. 2004. Biologically meaningful expression profiling across species using heterologous hybridization to a cDNA microarray. BMC Genomics 5.

Renn SCP, Aubin-Horth N, Hofmann HA. 2008. Fish and chips: Functional genomics of social plasticity in an African cichlid fish. J Exp Biol 211:3041–56.

Rise ML, et al. 2004. Development and application of a salmonid EST database and cDNA microarray: Data mining and interspecific hybridization characteristics. Genome Res 14:478–90.

Rokas A, Abbot P. 2009. Harnessing genomics for evolutionary insights. TREE 24:192–200.

Salzburger W, Renn SCP, Steinke D, Braasch I, Hofmann HA, Meyer A. 2008. Annotation of expressed sequence tags for the east African cichlid fish *Astatotilapia burtoni* and evolutionary analyses of cichlid ORFs. BMC Genomics 9:1–14.

Sartor MA, Zorn AM, Schwanekamp JA, Halbleib D, Karyala S, Howell ML, Dean GE, Medvedovic M, Tomlinson CR. 2006. A new method to remove hybridization bias for interspecies comparison of global gene expression profiles uncovers an association between mRNA sequence divergence and differential gene expression in *Xenopus*. Nucleic Acids Res 34:185–200.

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.

Shendure JA, Porreca GJ, Church GM. Overview of DNA sequencing strategies. Curr Protoc Mol Biol Chapter 7:Unit 7.1.

Shin H, Hirst M, Bainbridge MN, Magrini V, Mardis E, Moerman DG, Marra MA, Baillie DL, Jones SJM. 2008. Transcriptome analysis for *Caenorhabditis elegans* based on novel expressed sequence tags. BMC Biol 6:30.

Skopec MM, Haley S, Dearing MD. 2007. Differential hepatic gene expression of a dietary specialist (*Neotoma stephensi*) and generalist (*Neotoma albigula*) in response to juniper (*Juniperus monosperma*) ingestion. Comp Bioch Phys D Genomics Proteomics 2:34–43.

Smyth GK. 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat App Genet Mol Biol 3:1–26.

Smyth GK. 2005. Limma: Linear models for microarray data. In: Gentleman R, Carey V, Dudoit S, Irizarry R, Huber W, editors. Bioinformatics and computational biology solutions using R and bioconductor. New York: Springer. p. 397–420.

Spady TC, Seehausen O, Loew ER, Jordan RC, Kocher TD, Carleton KL. 2005. Adaptive molecular evolution in the opsin genes of rapidly speciating cichlid species. Mol Biol Evol 22:1412–22.

Tirosh I, Barkai N. 2008. Evolution of gene sequence and gene expression are not correlated in yeast. Trends in Genetics 24:109–113.

Toth AL, et al. 2007. Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. Science 318:441–4.

Toleno DM, Renaud G, Wolfsberg TG, Islam M, Wildman DE, Siegmund KD, Hacia JG. 2009. Development and evaluation of new mask protocols for gene expression profiling in humans and chimpanzees. BMC Bioinformatics 10:77.

Vallee M, Robert C, Methot S, Palin MF, Sirard MA. 2006. Cross-species hybridizations on a multi-species cDNA microarray to identify evolutionarily conserved genes expressed in oocytes. BMC Genomics 7:113.

Vera JC, Wheat CW, Fescemyer HW, Frilander MJ, Crawford DL, Hanski I, Marden JH. 2008. Rapid transcriptome characterization for a nonmodel organism using 454 pyrosequencing. Mol Ecol 17:1636–47.

von Schalburg KR, Cooper GA, Leong J, Robb A, Lieph R, Rise ML, Davidson WS, Koop BF. 2008. Expansion of the genomics research on Atlantic salmon Salmo salar L. project (GRASP) microarray tools. J Fish Biol 72:2051–70.

Wang Z, Brown DD. 1991. A Gene-Expression Screen. Proceedings of the National Academy of Sciences of the United States of America 88:11505–9.

Watanabe M, Kobayashi N, Shin-I T, Horiike T, Tateno Y, Kohara Y, Okada N. 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 343:263–9.

Weber APM, Weber KL, Carr K, Wilkerson C, Ohlrogge JB. 2007. Sampling the *arabidopsis* transcriptome with massively parallel pyrosequencing. Plant Physiol 144:32–42.

Williams DR, et al. 2008. Genomic resources and microarrays for the common carp *Cyprinus carpio L.* J Fish Biol 72:2095–117.