

RESEARCH ARTICLE

Open Access



Neurogenomic divergence during speciation by reinforcement of mating behaviors in chorus frogs (*Pseudacris*)

Oscar E. Ospina^{1,2}, Alan R. Lemmon³, Mysia Dye¹, Christopher Zdyrski^{1,4}, Sean Holland¹, Daniel Stribling^{1,5}, Michelle L. Kortyna¹ and Emily Moriarty Lemmon^{1*} 

Abstract

Background: Species interactions can promote mating behavior divergence, particularly when these interactions are costly due to maladaptive hybridization. Selection against hybridization can indirectly cause evolution of reproductive isolation within species, a process termed cascade reinforcement. This process can drive incipient speciation by generating divergent selection pressures among populations that interact with different species assemblages. Theoretical and empirical studies indicate that divergent selection on gene expression networks has the potential to increase reproductive isolation among populations. After identifying candidate synaptic transmission genes derived from neurophysiological studies in anurans, we test for divergence of gene expression in a system undergoing cascade reinforcement, the Upland Chorus Frog (*Pseudacris feriarum*).

Results: Our analyses identified seven candidate synaptic transmission genes that have diverged between ancestral and reinforced populations of *P. feriarum*, including five that encode synaptic vesicle proteins. Our gene correlation network analyses revealed four genetic modules that have diverged between these populations, two possessing a significant concentration of neurotransmission enrichment terms: one for synaptic membrane components and the other for metabolism of the neurotransmitter nitric oxide. We also ascertained that a greater number of genes have diverged in expression by geography than by sex. Moreover, we found that more genes have diverged within females as compared to males between populations. Conversely, we observed no difference in the number of differentially-expressed genes within the ancestral compared to the reinforced population between the sexes.

* Correspondence: chorusfrog@bio.fsu.edu

¹Department of Biomedical Sciences, College of Veterinary Medicine, Iowa State University, 1800 Christensen Drive, 50011 Ames, IA, USA
Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Conclusions: This work is consistent with the idea that divergent selection on mating behaviors via cascade reinforcement contributed to evolution of gene expression in *P. feriarum*. Although our study design does not allow us to fully rule out the influence of environment and demography, the fact that more genes diverged in females than males points to a role for cascade reinforcement. Our discoveries of divergent candidate genes and gene networks related to neurotransmission support the idea that neural mechanisms of acoustic mating behaviors have diverged between populations, and agree with previous neurophysiological studies in frogs. Increasing support for this hypothesis, however, will require additional experiments under common garden conditions. Our work points to the importance of future replicated and tissue-specific studies to elucidate the relative contribution of gene expression divergence to the evolution of reproductive isolation during incipient speciation.

Keywords: Cascade reinforcement, Behavioral evolution, Mate choice, Transcriptomics, Neurotransmission, Synaptic transmission, Weighted correlation networks, Gene modules

Background

Species interactions can lead to the divergence of reproductive behaviors, thus reducing the probability of costly hybridization, a process termed reinforcement [1–4]. Furthermore, as interacting taxa diverge, populations within a hybrid zone (sympatry) may incidentally diverge from populations in other hybrid zones or outside of the hybrid zone (allopatry), leading to reproductive isolation within species [3, 5–7]. Theoretical models indicate that the latter process, cascade reinforcement [8], can act as a species multiplier, increasing the rate of speciation within a clade [9–11]. Specifically, as additional taxa arise, novel and spatially-varying species assemblages are formed leading to different selection pressures on reproductive behaviors. The strength of the selection depends upon the combination of closely-related species present in each geographic area [12]. These varying selection pressures may thus contribute to the rapid proliferation of new species [13]. This process has been documented to drive divergence across a broad spectrum of taxa, including fish [14], frogs [15–17], *Drosophila* [18–23], walking-stick insects [24], beetles [25], and flowering plants [26, 27].

Systems experiencing cascade reinforcement provide a unique opportunity to study the genetic changes that occur during the first stages of speciation [28]. Although little is known about the genetic targets of selection underlying in this mechanism, some information is available from studies of the more general process of reinforcement between species [26, 29–31]. For example, previous work has identified signatures of selection, such as selective sweeps, at documented [32] and candidate [33, 34] reinforcement loci, under the assumption that loci experiencing strong selection in sympatry should show decreased genetic variation [28].

As predicted by theory [35], the evolution of reproductive isolation during reinforcement can result from structural changes at the genomic level [36, 37]. Diverging species and populations are also likely to undergo differentiation of gene regulation pathways [38, 39].

Because gene expression is regulated through complex interactions of sequences and proteins within the cell, these interacting regulatory elements are thought to be co-adapted [40]. When these co-adapted complexes are disrupted, such as in hybrids, the regulatory targets may be mis-expressed, resulting in incompatibilities between species or divergent populations [39, 41]. These incompatibilities are particularly likely to contribute to reproductive isolation when selection targets different components of a gene network [42], which can occur in replicate sympatric populations during cascade reinforcement.

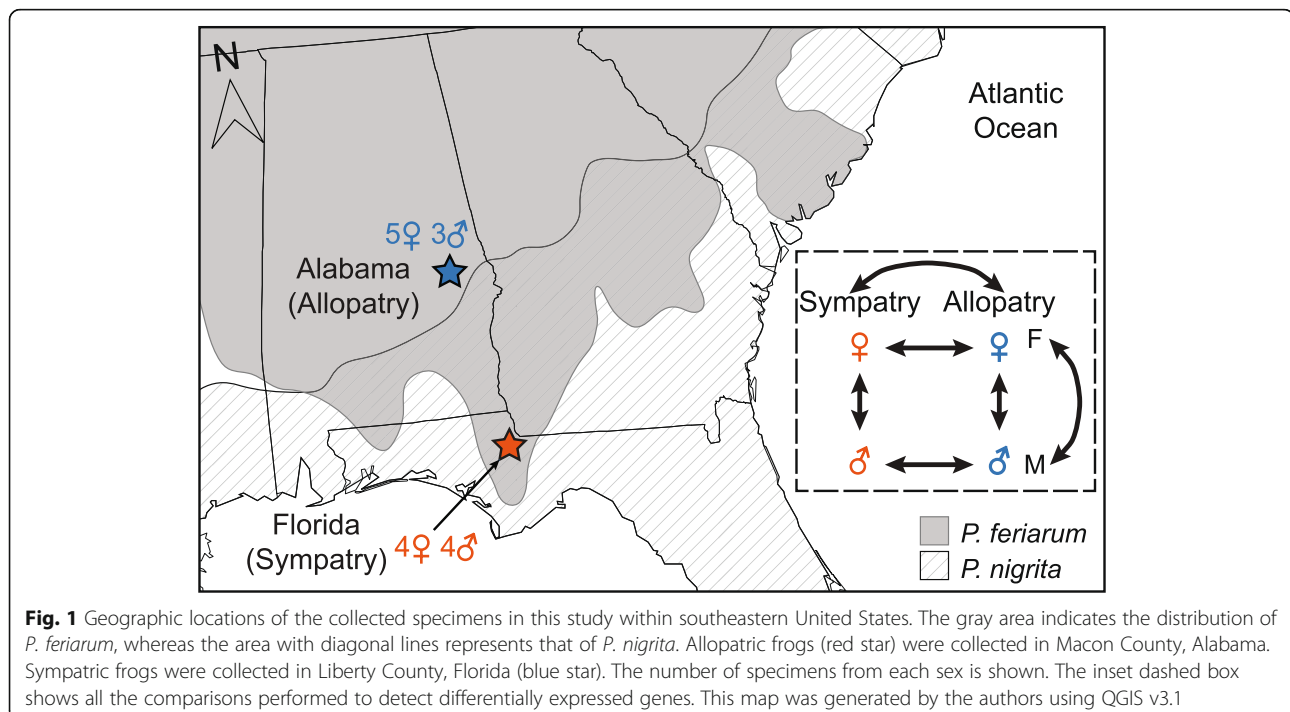
In the well-studied European house mouse system (*Mus musculus* spp.), Loire et al. [34] identified candidate mate recognition genes by studying gene expression divergence during reinforcement. Specifically, they found differential gene expression of candidate vomeronasal receptor genes between populations of the same subspecies within and outside of a reinforced contact zone. Loire et al. [34] expected to find divergence of these candidate genes because behavioral studies indicated that mice use pheromone-based information in urine for discriminating between subspecies [43–47]. Their findings suggest that divergence of olfactory gene expression is also contributing to the evolution of assortative mating among populations within one of the subspecies [47]. In the only study of gene expression in a frog system undergoing reinforcement, the spadefoot toads (*Spea* sp.), Seidl et al. [41] determined that divergence of gene expression profiles accompanies divergence of reproductive behaviors in sympatry. In spadefoots, the interacting species have evolved increased species discrimination within their contact zone, due to the negative consequences of hybridization [48, 49] (but see [50]). Seidl et al. [41] found that hybrids derived from interspecific crosses of parents originating in sympatry differed significantly in gene expression from their counterparts generated from allopatric parents. They concluded that ongoing hybridization in sympatry is driving the divergence of gene expression between geographic

regions. These empirical studies support theoretical predictions that reinforcing selection can contribute to divergence of gene expression across populations, pointing to the need for studies on additional non-model systems.

In the North American chorus frogs (genus *Pseudacris*), reinforcement in contact zones has caused one species (*P. feriarum*) to diverge in male acoustic signals and female mating preferences from another species (*P. nigrita*) [51]. As a consequence, increased reproductive isolation has evolved between the two species in sympatry [51, 52]. The lineages giving rise to *Pseudacris feriarum* and *P. nigrita* diverged ~ 8 million years ago (ma) based on estimates from mitochondrial markers, and the two taxa have come into secondary contact more recently [53, 54]. Phylogeographic evidence indicates that *P. feriarum* expanded into the range of *P. nigrita* multiple times, by following the floodplains of river systems that bisect the Coastal Plain of the southeastern United States (Fig. 1 [55]). Thus, allopatric *P. feriarum* represent the ancestral state for the species, whereas sympatric *P. feriarum* represent the derived, reinforced state. Empirical work indicates that populations within *P. feriarum* are diverging via cascade reinforcement: sympatric *P. feriarum* females have evolved strong preferences for sympatric over allopatric male signals, and male calls have diverged substantially between regions [51]. Estimates of the timing of secondary contact with *P. nigrita* are uncertain; mitochondrial-based estimates range from 0 to 4 ma [56]. Within *P. feriarum*, genetic divergence between allopatric and sympatric (with *P. nigrita*) regions

including the populations in this study are low ($F_{ST} = 0.04$), and less than between adjacent allopatric regions of similar geographic distance ($F_{ST} = 0.12$ [55]). The significant differentiation of both male and female reproductive behaviors between ancestral and reinforced populations suggests that the sexual decision-making genetic pathways have diverged in the brains of sympatric frogs [51]. Furthermore, because the cost of hybridization is higher for females than males, female *P. feriarum* may have diverged to a greater degree than males, due to strong natural and sexual selection on female mate preferences [52].

Reinforcement selection may have driven divergence in expression of genes involved with mating behaviors, particularly within the auditory midbrain. Neuroethological studies in another chorus frog species (*P. regilla*), have revealed that critical elements of male acoustic signals are decoded in the inferior colliculus of the auditory midbrain [57, 58]. Electrophysiological analyses have identified two classes of pulse-rate selective neurons in this brain region: interval-counting neurons (ICNs) and long-interval neurons (LINs) [59–63]. ICNs show band-pass selectivity for pulse rate and respond only after a threshold number of pulses are presented with optimal interpulse onset intervals [59]. Importantly, these two acoustic characteristics of mating signals diverged between allopatric and sympatric populations of *P. feriarum* [51]. The tuning of ICNs for pulse rate and interval counting involves interplay between excitation via AMPA-type and NMDA-type ionotropic glutamate



receptors and inhibition via ionotropic GABA_A receptors [61, 64]. Our preliminary electrophysiological data suggest that allopatric and sympatric *P. feriarum* differ with respect to the tuning of ICNs in the auditory midbrain (E. Lemmon and G. Rose, unpublished observations). These differences may be mediated by divergence in genes and/or their expression that govern the presence and function of inhibitory and excitatory synapses. For example, up-regulation of one or more genes that encode GABAergic receptors in ICNs may increase the relative amplitude of inhibition and as a consequence increase the number of pulses required for eliciting responses [64, 65] (G. Rose, unpublished observations). Thus, changes in expression of ionotropic receptor genes and other genes related to synaptic transmission might underlie the observed differences in mating behaviors between ancestral and reinforced populations.

In this study, we aim to better understand the changes that occur in the earliest stages of speciation via cascade reinforcement in *P. feriarum*. Examining populations of the same species that are experiencing different selection pressures on their reproductive behaviors allows us to focus more directly on early genetic changes as speciation proceeds [28]. To this aim, we generate a comprehensive reference transcriptome for *P. feriarum* using a variety of tissues, then use a hypothesis-testing approach to narrow down a large set of candidate genes involved in synaptic transmission in the brain to a smaller set that have diverged between ancestral and reinforced populations. We predict that one or more genes involved in synaptic transmission will be differentially expressed between ancestral and reinforced populations in one or both sexes.

In order to understand the general patterns of divergence across the entire transcriptome, we also study expression changes (between sexes and between populations) for the entire set of genes expressed in the brain. We make several predictions based on our current understanding of the neurophysiology and evolutionary history of these frogs. First, we expect that gene expression patterns have diverged to a greater degree between ancestral and reinforced populations than between the sexes, due to the effects of natural and sexual selection via reinforcement [52]. Second, we predict that females from ancestral and reinforced populations will show the greatest divergence, since females experience a greater cost to hybridization and thus stronger selection than males. If the first prediction is upheld, the pattern could alternatively be explained by differences in environmental selection (i.e., abiotic factors, such as climate, or other biotic factors, such predator communities), or by historical demographic differences between populations. However, if the second prediction is also upheld, these two alternative hypotheses would be less likely because

the sexes should be affected similarly by environmental factors or demographic shifts. Third, we predict that genes involved in synaptic transmission will be overrepresented in one or more gene networks that have diverged between ancestral and reinforced populations, since neural processing by auditory neurons of the midbrain are thought to mediate mate choice in this system.

Results

Reference transcriptome assembly

We obtained a comprehensive reference transcriptome. Before filtering and processing, we obtained between 28 million and 56 million read pairs from each tissue (brain, eye, testis, liver, heart, lung, skin, and leg muscle) from two individuals. In all cases, 98% of the reads had high average quality with Phred score > 28 and average length of 100 bp (Additional file 1: Table S1). Following trimming and rRNA filtering our transcriptome data set contained between 27 million and 55 million read pairs per tissue type with 99% of the sequences showing Phred score > 28. After assembly, we obtained 598,748 contigs representing 387,298 unique transcripts with average, median and N50 length of 639,325, and 1019 bp respectively. This set of transcripts was considered as our assembled reference transcriptome for all downstream analyses. A different set of sequences (*Results—Brain Expression Transcriptomes*) was mapped to the reference transcriptome to study gene expression. The BUSCO (Benchmarking Universal Single-Copy Orthologs) search for ortholog genes in the Tetrapod ortholog database showed a coverage of 95.2% for the assembled transcriptome, representing 3761 complete single-copy orthologs. The mean and median length of the contigs matching subjects in BUSCO was 2669 and 2272 bp, respectively, meaning that contigs larger than the whole transcriptome median contig length (~ 360 bp) contributed to ortholog detection. Length coverage analyses to assess protein length (based on blast searches; *Methods—Reference Transcriptome Assembly*) showed that 10,993 of the UniProt/SwissProt and 13,779 proteins of the *Xenopus tropicalis* databases matched more than 90% of their aminoacidic lengths to transcripts in our assembled transcriptome (Additional file 1: Table S2).

The annotation (performed in Trinotate; *Methods—Reference Transcriptome Assembly*) based on the UniProt/SwissProt database yielded identification for 99,774 (26%) of the transcripts. Of those, 18,358 transcripts were associated with *Xenopus tropicalis*. For the UniProt/SwissProt annotations, our reference transcripts matched with 21,841 Gene Ontologies (GOs). Annotation using the Pfam database for protein families yielded 48,469 annotations (12.5% of transcripts), of which 2283 were associated with GO terms. Overall, 11,512 different GO terms were associated with biological processes,

3995 with molecular functions, and 1751 with cellular components (Additional file 2: Fig. S1).

Brain transcriptomes

We quantified gene expression in 17 individual *P. feriarum* from four groups: sympatric females ($n = 5$), sympatric males ($n = 4$), allopatric females ($n = 5$), and allopatric males ($n = 3$). After processing and filtering, the number of paired reads from each sample ranged from 14.9 to 97.6 million sequences (Additional file 1: Table S3). Of the 17 samples sequenced, all but one produced data of sufficiently high quality to be included in downstream analyses (Additional file 3: Fig. S2); one sympatric female sample produced substantially fewer post-filter read pairs (14.9 million) than the remainder of the samples (mean = 55 million, min = 34 million). Data for this individual were not carried forward in the downstream analyses presented below. On average, 64% of the reads from each sample mapped to our reference transcriptome (61–69%). Studies mapping expression reads to a more comprehensive, pre-existing transcriptome have achieved similar mapping percentages (e.g., [66]). None of the 16 samples appeared as outliers in the multidimensional scaling (MDS) analysis of the Trimmed Mean of M-values (TMM) log₂ counts per million (CPM) values (Additional file 4: Fig. S3).

Identification of candidate synaptic transmission genes

From published studies [61, 64, 65, 67, 68], we identified 39 target protein families associated with synaptic transmission (Table 1; *Methods—Identification of Candidate Gene Transcripts Involved in Synaptic Transmission*). We obtained matches to 97% (38 of 39) of these target protein families in the expression transcriptome: 1104 (~2%) of reference transcripts contained one or more keywords. The one exception is, Pp60src, a tyrosine kinase that phosphorylates synaptophysin and synaptogyrin. The keywords matching this target gene were not found in the annotation terms of any of the transcripts in the reference transcriptome. Of the matching reference transcripts, 554 were present in the expression data set (i.e., passed the quality/coverage filters). These transcripts, which we hereafter refer to as the candidate synaptic transmission genes, included 54 ionotropic receptors, 259 synaptic vesicle proteins, 126 proteins that associate with synaptic vesicles, 99 synaptic plasma membrane proteins, and 16 proteins that reversibly associate with plasma membrane proteins [67].

Differential expression of candidate synaptic transmission genes

Analysis of the candidate synaptic transmission genes revealed seven genes with significant expression differences by geography between populations, two in males

only and five when both sexes were combined (Additional file 1: Table S4). The log difference ratios of these genes are highlighted in Fig. 2. Five of the differentially-expressed (DE) genes are synaptic vesicle proteins: Secretory carrier-associated membrane protein 3 (SCAMP3, False-Discovery Rate [FDR] = 0.5×10^{-3}), Synaptotagmin-2-B (ESYT2-b, FDR = 0.038), GABA transporter 2 (Slc6a13, FDR = 0.038), Sodium and chloride-dependent betaine/GABA transporter (Slc6a12, FDR = 0.038), and Zinc transporter ZIP10 (Slc39a10, FDR = 0.038). The two remaining differentially-expressed genes include a protein that associates with synaptic vesicles, kinesin-associated protein 3 (KIF3A, FDR = 4.87×10^{-6}), and a synaptic plasma membrane protein, cadherin-15 (CDH15, FDR = 0.038). Note that sampling 554 genes (candidate synaptic transmission genes) at random from the entire set of genes results in 8.4 significant genes on average, suggesting that the synaptic transmission gene candidate set does not contain a greater number of differentially expressed genes than would be expected [$P(\text{number significant}) > 6 = 0.634$], given the size of the gene set.

Comparing expression divergence among the four groups

Expression trees showing the relative divergence among the four groups (Additional file 5: Fig. S4; Additional file 1: Table S5) did not identify a group in which expression divergence was greatest (all FDR were greater than 0.08). Moreover, when comparing the ratio of branch lengths (representing the relative amount of evolution of two groups) between the candidate genes and non-candidate genes, we found no difference (all FDR were greater than 0.08), suggesting that the relative rates of evolution for the four groups were not enhanced in the candidate gene set.

Transcriptome-wide patterns of divergence

Our transcriptome-wide analysis, based on 48,254 genes, (*Methods—Evaluating Transcriptome-wide Patterns of Divergence*), revealed a substantial number of differentially-expressed genes (Fig. 3), with the greatest difference in expression seen between ancestral and reinforced populations (Fig. 4). Among the four Geography x Sex comparisons, we detected 298 differentially-expressed genes based on an FDR threshold of 0.05 (Additional file 1: Table. S6). The highest number of differentially-expressed genes detected were between sympatric and allopatric females ($n = 196$), however, a substantial number was also identified between sympatric and allopatric males ($n = 78$), 19 of which were shared between these two comparisons. A multidimensional scaling (MDS) plot based on pairwise log₂-fold changes of all DE genes separated the samples primarily

Table 1 Categories of predicted differentially-expressed genes between allopatric and sympatric populations. Categories and genes involved in synaptic transmission for the anuran auditory pathway are summarized from [67, 68]. Gene and protein names indicate significant differentially-expressed transcripts between allopatric and sympatric populations for each gene/gene family after FDR correction (0.05)

Categories of Predicted Proteins	Diff. Expr. Gene Name	Diff. Expr. Protein Name
1. Ionotropic receptors		
Glutamate receptor ionotropic, NMDA	–	–
Glutamate receptor ionotropic, AMPA	–	–
Gamma-aminobutyric acid receptor (GABA)	–	–
2. Synaptic vesicle proteins		
Neurotransmitter transporters	Slc6a12; BGT1	Sodium- and chloride-dependent betaine transporter /GABA transporter
Neurotransmitter transporters	Slc6a13; GAT2	Sodium- and chloride-dependent GABA transporter 2
Cysteine string protein (CSP)	–	–
Cytochrome b561	–	–
Rab and Ra1 proteins	–	–
Rabphilin-3A	–	–
Secretory carrier membrane proteins (SCAMPs)	SCAMP3	Secretory carrier-associated membrane protein 3
Synaptic vesicle protein (SV2)	–	–
Synapsins	–	–
Synaptobrevins	–	–
Synaptogyrin	–	–
Synaptophysin	–	–
Synaptotagmin	ESYT2-b	Extended synaptotagmin-2-B
Transport proteins (channels) for chloride and zinc ^a	SLC39A10	Zinc transporter ZIP10
Vacuolar proton pump	–	–
3. Proteins that associate with synaptic vesicles		
Amphiphysin	–	–
Endophilin	–	–
AP2 and clathrin	–	–
Ca ²⁺ , calmodulin-dependent protein kinases I and II (CaMKI and CaMKII)	–	–
Dynamin-1	–	–
Dynein	–	–
Kinesins	KIFA3	Kinesin-associated protein 3
Guanine nucleotide exchange factor MSS4	–	–
Pp60src (tyrosine kinase that phosphorylates synaptophysin and synaptogyrin)	–	–
Complexin/synaphin	–	–
4. Synaptic plasma membrane proteins		
Munc13s	–	–
Neurexins	–	–
SNAP-25	–	–
Syntaxins	–	–
Voltage-gated Ca ²⁺ channels	–	–
RIM	–	–
Neurologin	–	–

Table 1 Categories of predicted differentially-expressed genes between allopatric and sympatric populations. Categories and genes involved in synaptic transmission for the anuran auditory pathway are summarized from [67, 68]. Gene and protein names indicate significant differentially-expressed transcripts between allopatric and sympatric populations for each gene/gene family after FDR correction (0.05) (*Continued*)

Categories of Predicted Proteins	Diff. Expr. Gene Name	Diff. Expr. Protein Name
Cadherin	CDH15; M-cadherin	Cadherin-15
5. Proteins that reversibly associate with plasma membrane proteins		
Munc18s/syntaxin binding protein	–	–
N-ethylmaleimide-sensitive factor (NSF)	–	–
$\alpha/\beta/\gamma$ SNAPs	–	–

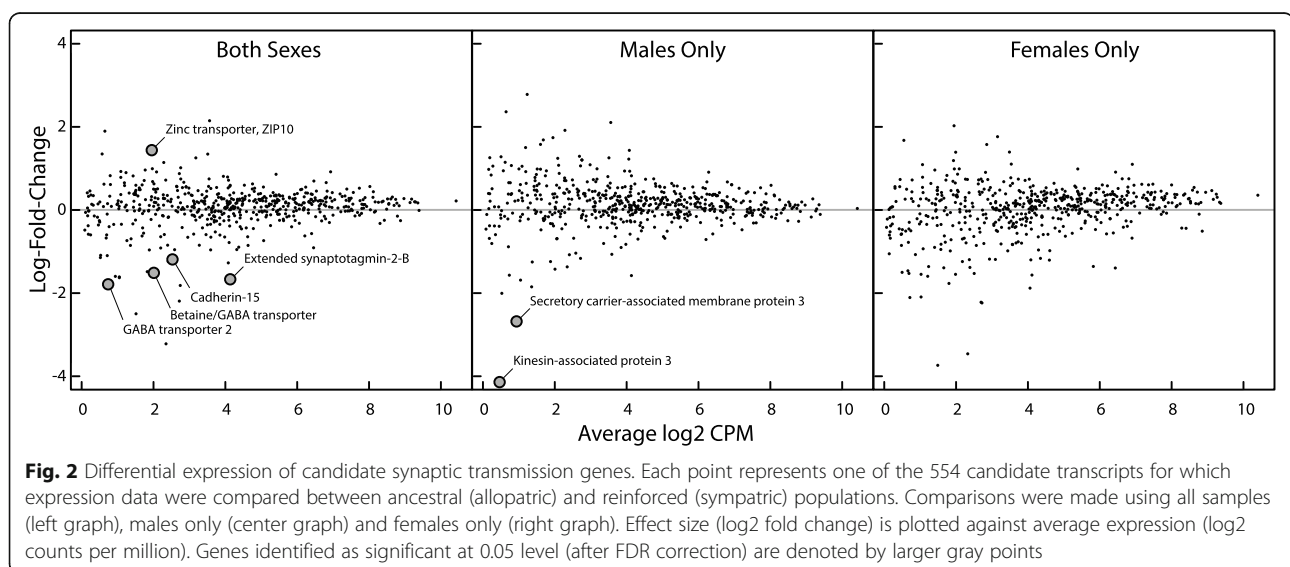
^aMembers of the ZnT zinc transporter family are from Category 2, but those from the ZIP zinc transporter family, which occur in the plasma membrane instead, are from Category 4 [94]s

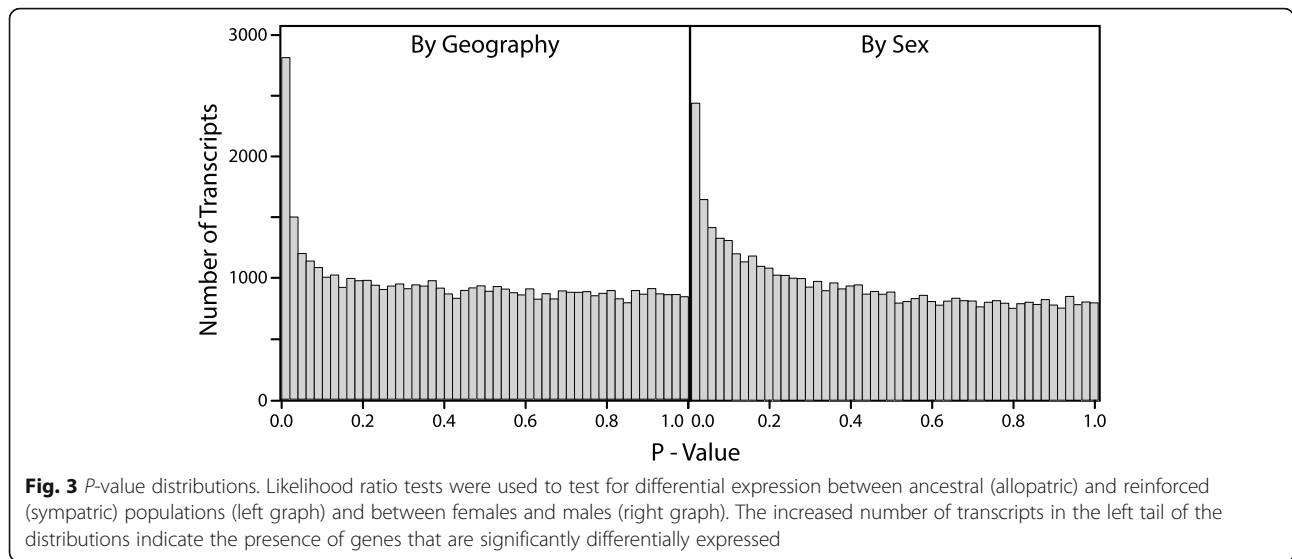
by geography (sympatric versus allopatric populations) and to a lesser degree by sex (Fig. 5).

The majority of differentially-expressed genes were observed between populations. The results from the LRTs (Likelihood ratio tests) showed 517 differentially-expressed genes between all allopatric frogs and all sympatric frogs. Note that with the FDR correction and an alpha threshold of 0.05, there is only a 5% chance of one or more genes showing significant differential expression. Of the genes differentially expressed by geography, 166 were overexpressed in sympatry and 351 were underexpressed in sympatry (see Additional file 1: Table S6 for effect sizes). The hierarchical clustering analysis supported the separation between the two groups with an approximately unbiased p -value (AU, a multiscaled version of bootstrap [69]), $AU > 86\%$ (Additional file 6: Fig. S5).

Between allopatric and sympatric females, 196 genes were differentially expressed. A clustering analysis based on those genes provided high support ($AU > 89\%$) for the separation of allopatric females from sympatric

females (Fig. 6A). Between allopatric and sympatric males, 78 genes were differentially expressed that separated these two groups ($AU = 100\%$, Fig. 6B). As expected, differentially-expressed genes between females and males allowed statistically significant clustering of transcriptomic profiles based on sex (Additional file 7: Fig. S6). The contrast between brains of all female and all male frogs yielded 129 differentially-expressed genes. Of those, 81 were underexpressed in females and 42 were overexpressed in females. The hierarchical clustering analysis supported a cluster including most females except two from sympatry ($AU = 88\%$; Additional file 7: Fig. S6). The same two females were included in a second cluster with the male samples ($AU = 73\%$). We found 24 differentially-expressed genes between females and males in allopatry. Overall, these differences in gene expression did not form hierarchical clusters as well supported as those observed in the comparison between allopatric and sympatric frogs. Allopatric males and allopatric females





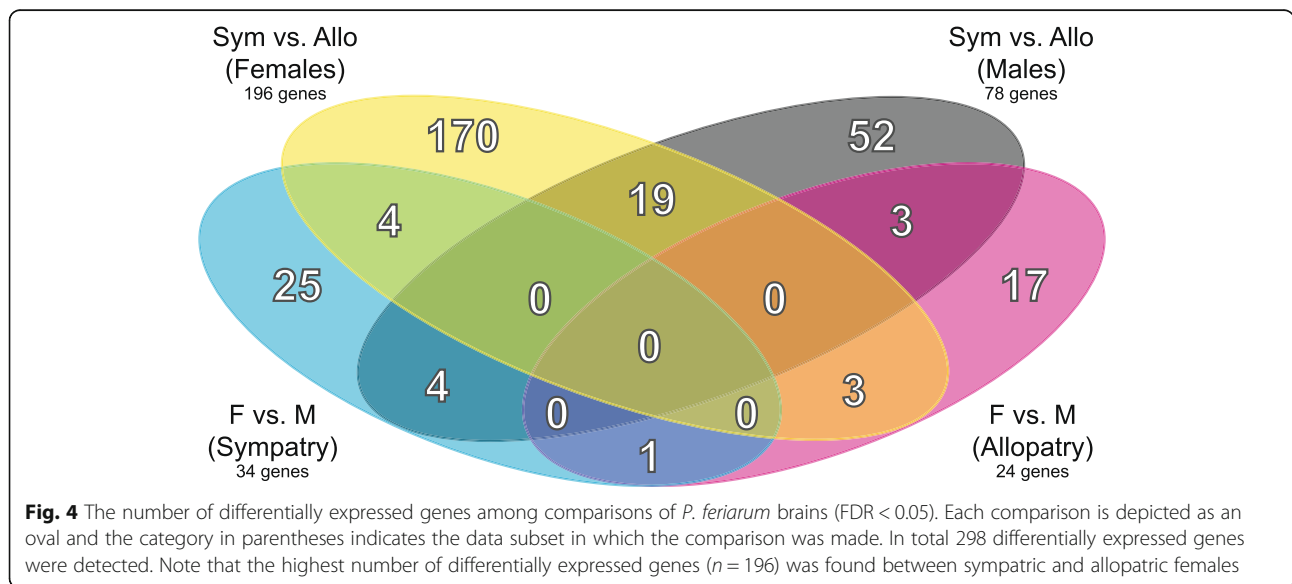
formed two clusters supported by AU = 99% (Additional file 8: Fig. S7).

More genes were differentially expressed by geography than by sex, as indicated by our randomization tests (ratio = 4.4737, $p < 0.0001$). The randomization tests also demonstrated that females had more genes differentially expressed by geography than males (ratio = 2.51, $p < 0.0001$). We did not, however, find a significant difference between sympatry and allopatry in the number of genes differentially expressed by sex (ratio = 1.4167, $p = 0.1217$).

Divergence of brain gene co-expression networks

The analysis of co-expression network yielded 25 modules (31 before merging) of genes with correlated

expression levels and high connectivity among them (Additional file 9: Fig. S8; Additional file 1: Table S7-S8). The co-expression networks were constructed from 4826 genes in the top 10% with respect to the amount of among-sample variation in normalized CPM. The network filtered at correlation coefficients (r) lower than $r = 0.05$, contained 2871 transcripts with $r \geq 0.1$, and 513 transcripts showing connections with $r \geq 0.3$. The average degree of connectivity (number of connections) in this co-expression network was 325.6, with average clustering coefficient (C) and average neighborhood connectivity (NC) of $C = 0.7$ and $NC = 637.1$, respectively, indicating overall high connection among all genes and their neighbors. A total of 372 differentially-expressed



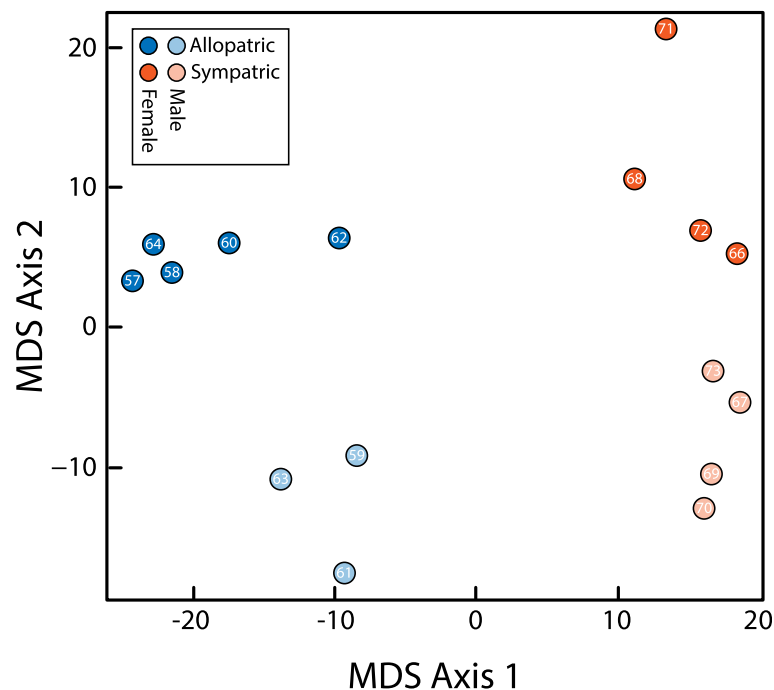


Fig. 5 Multidimensional Scaling Plot of pairwise expression differences, based on normalized log₂ counts per million for the differentially expressed genes in the transcriptome-wide gene set. Note the discrete break between the two populations (allopatric and sympatric). Also note the higher level of dispersion for sympatric samples than allopatric samples. Numbers inside points correspond to the last two digits of the sample IDs in Additional file 1: Table S10

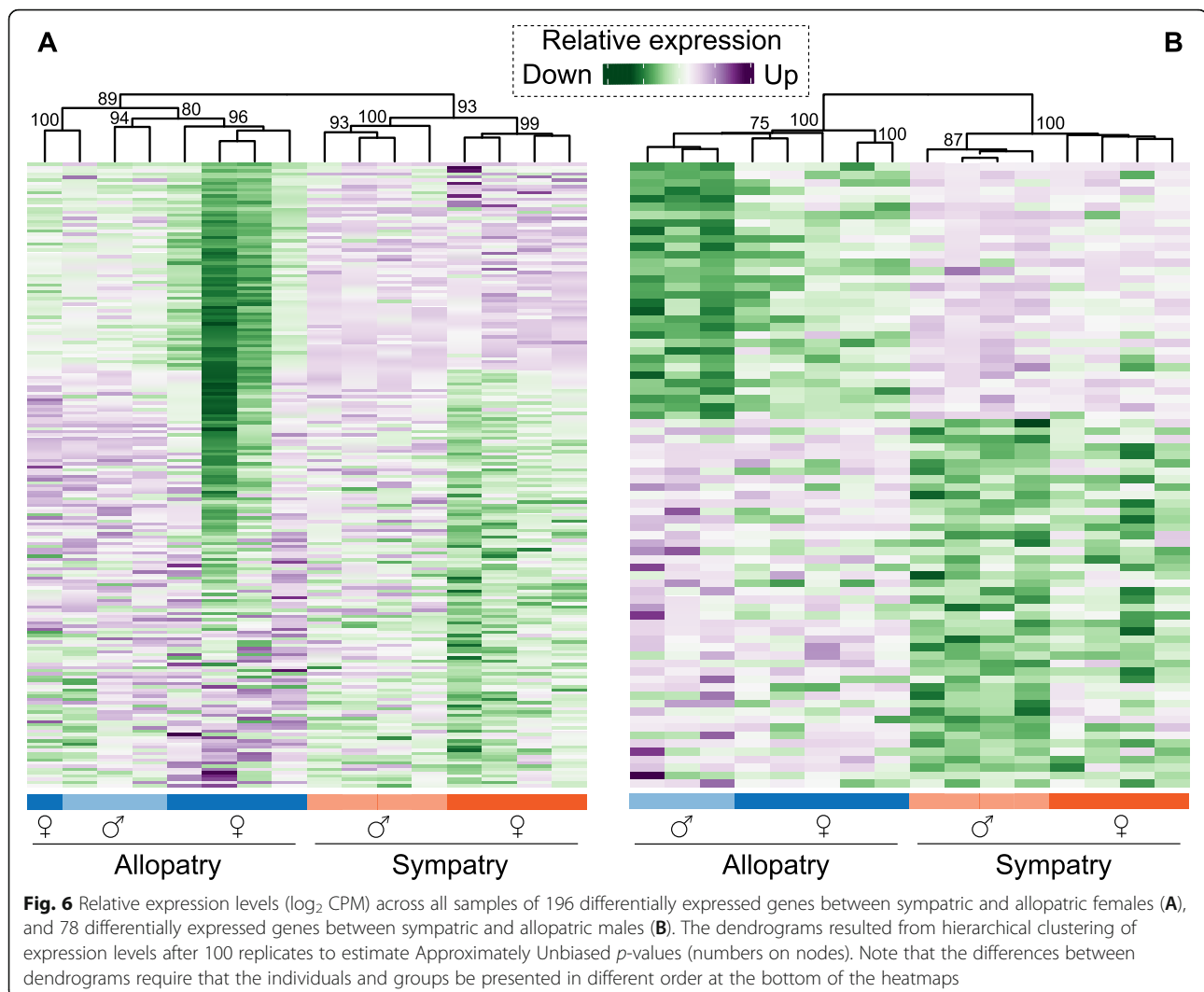
genes from any of the previously mentioned comparisons were included in this network analysis.

Correlation tests yielded four modules with significant association between overall gene expression and geography (Fig. 7; Fig. S8; Additional file 1: Table S8-S9). Those modules were named darkred, purple, red, and black following WGCNA's module notation. One of those modules, named darkred, included 71 genes that showed expression levels associated with geography ($r = 0.65$, $FDR = 0.0260$). This module showed average connection degree of 9.0, $C = 0.6$, and $NC = 13.7$. The darkred module contained 14 differentially-expressed genes. A second significant module, termed the purple module, included 156 genes that exhibited expression levels associated with geography ($r = 0.93$, $FDR < 0.0001$). The purple module showed an average degree of connectivity of 15.4 and $C = 0.6$, and $NC = 28.9$ and contained 63 differentially-expressed genes. A third module, named red, contained 196 genes showing association of expression levels with geography ($r = -0.8$, $FDR = 0.0019$). The average degree of genes within this module was 41.6, $C = 0.8$, and $NC = 81.3$. The red module contained 39 differentially-expressed genes. The fourth module, which was called the black module, contained 190 genes with gene

expression patterns correlated with geography ($r = 0.6$, $p = 0.0260$). The genes in this module had an average degree of 39.6, $C = 0.76$, and $NC = 68.2$. The module included 14 differentially-expressed genes.

Only the blue module had gene expression associated with sex (Fig. 7; Additional file 9: Fig. S8; Additional file 1: Table S8-S9). The blue module included 575 genes with expression levels correlated with a frog's sex ($r = 0.64$, $p < 0.0260$). The genes in this module showed an average degree of 279.2, $C = 0.9$, and $NC = 381.0$. Of the 575 genes in the blue module, 44 were differentially expressed.

The 36 neurotransmission-related enriched GO pathways identified (Additional file 1: Table S9) were not randomly distributed across the modules but are instead significantly concentrated in the darkred module (15 of 28; $FDR < 0.0001$), the purple module (9 of 27, $FDR = 0.0088$), and the white module (11 of 40; $FDR = 0.0125$; note the white module was not significantly correlated with geography or sex; Fig. S8; Table S8-S9). Moreover, the modules associated with geography contained significantly more synaptic pathways than modules not associated with geography (randomization test, ratio = 3.40, $p < 0.0001$).



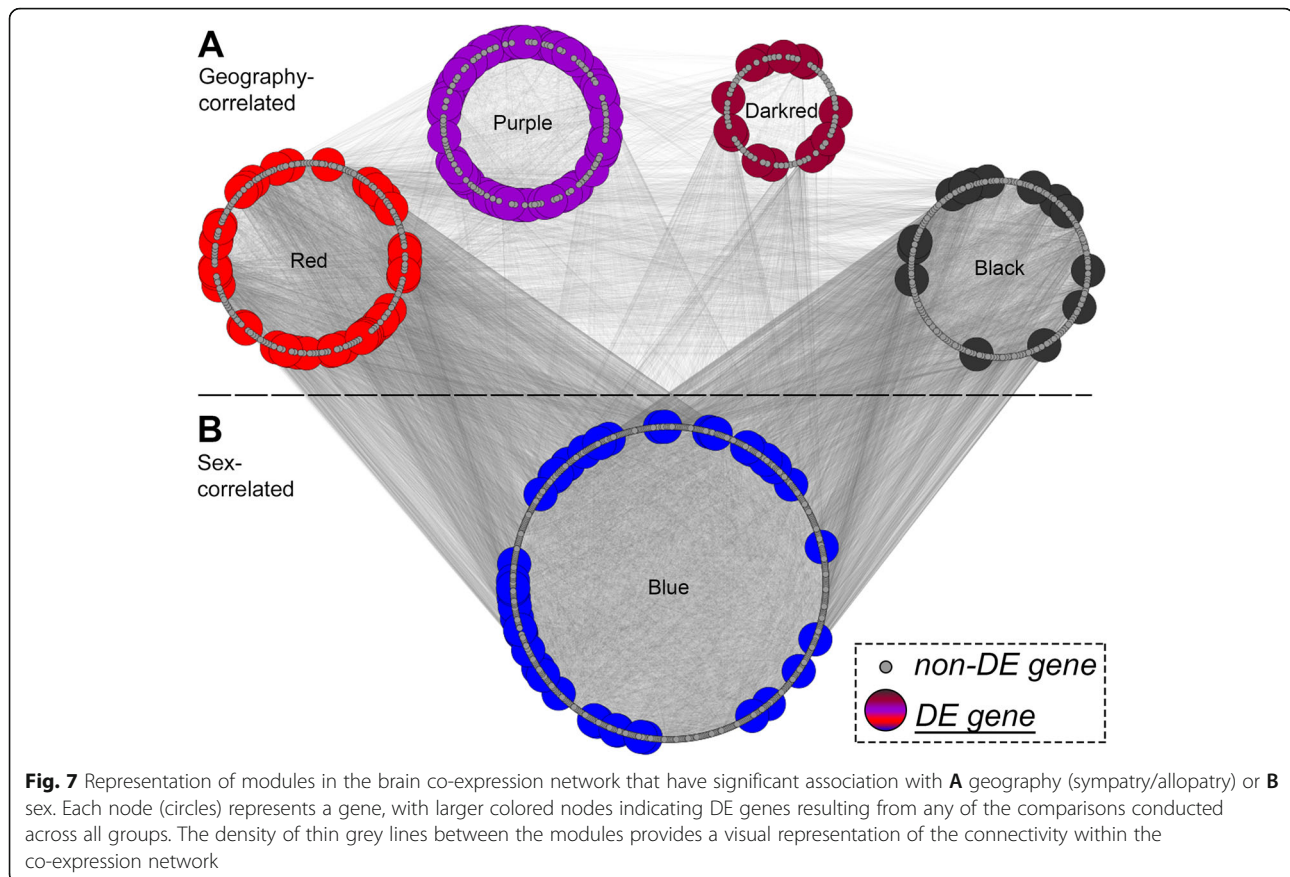
Discussion

Our study has revealed multiple candidate synaptic genes and four gene modules that have diverged between ancestral (allopatric) and reinforced (sympatric) populations of *P. feriarum*. Two of these modules are significantly enriched for synaptic transmission functions. We found a greater number of genes have diverged in expression between ancestral and reinforced populations than between the sexes. Furthermore, we observed a significantly greater number of genes that are differentially expressed by population in females compared to males. In contrast, we found no difference in the number of genes differentially expressed by sex within the ancestral compared to the reinforced population. Although this study represents just the first glimpse into brain gene expression divergence during cascade reinforcement in this system, our results are consistent with neurophysiological and behavioral studies suggesting that changes in synaptic transmission

may lead to divergence of mate recognition mechanisms during speciation.

Transcriptome-wide divergence of ancestral and reinforced populations

Our data are consistent with the idea that natural and sexual selection driving reinforcement in this system have contributed to differential gene expression between ancestral and reinforced populations. Support for this interpretation comes from the patterns of greater gene expression divergence by geography than by sex and of greater divergence by geography in females than males. Together, these patterns suggest that environmental selection and demographic processes may not be the only factors contributing to divergence by geography, but rather, stronger selection against hybridization in females than males may also underlie these differences [51, 52].



Some of the observed gene expression divergence by geography may be due to factors other than reinforcement, such as environmental selection or demographic history of the populations. For example, in terms of environmental differences, sympatric *P. feriarum* in Florida primarily breed in cypress-gum swamps along river floodplains, whereas allopatric conspecifics in Alabama utilize wetlands in more general mixed temperate forest habitats. These differences suggest that sympatric frogs may be more restricted with regard to their habitat requirements. A previous study of environmental selection on acoustic signals found, however, that sensory drive [70–72] is not driving divergence of these behaviors [73]. This study, which included the two populations in the current study, suggests that habitat selection has not contributed significantly to divergence of acoustic signals between populations. Moreover, overall environmental conditions of the two locations are similar (e.g., Level III Ecoregion [74]).

The two populations examined in our study may also differ somewhat in their demographic histories. Repeated contact zones have formed across *P. feriarum* and *P. nigrita* distributions [55]. In these contact zones, *P. feriarum* has expanded into the range of the latter species multiple independent times by following different river

systems of the southeastern U.S. (e.g., Escambia, Apalachicola, Altamaha, Edisto/Santee, and James/Anna Rivers). In each contact zone, migration estimates point to ongoing unidirectional gene flow, from the ancestral allopatric region into the reinforced sympatric areas [55]. Estimates of effective population sizes (N_E) are equivalent to or larger in replicate sympatric compared to allopatric populations across all five river systems [55], suggesting that that drift is not predicted to be stronger in peripheral sympatric populations. Consistent with the migration data, estimates of genetic divergence between allopatric and sympatric regions including the populations in this study are low ($F_{ST} = 0.04$), and less than between adjacent allopatric regions of similar geographic distance ($F_{ST} = 0.12$ [55]).

Divergence of brain gene co-expression networks

Of the four genetic modules that diverged between the ancestral and reinforced populations, two were enriched for and showed significant concentration of terms related to neurotransmission (darkred $FDR < 0.0001$ and purple modules $FDR = 0.0088$) compared to all other network modules. Specifically, darkred module terms included integral and intrinsic components of the pre-synaptic and postsynaptic membrane, postsynaptic

density, and extrinsic component of plasma membrane, among others (Additional file 1: Table S9). These terms encompass the products of the candidate genes that were found to be differentially expressed by geography in this study (see below). Purple module terms involved metabolism of the neurotransmitter nitric oxide (NO). The NO pathway modulates the release of multiple neurotransmitters [75, 76], and has been shown to be involved in mating behaviors [77, 78]. Mate choice studies in *Drosophila* have shown that in neurons of the ellipsoid body of the brain, NO acts as a second messenger after stimulation by GABA via GABA-A receptors to contribute to activation of NMDA receptors, resulting in the alteration of mate decisions by females [79, 80]. This connection between nitric oxide, synaptic transmission via ionotropic receptors, and mate choice is consistent with computational models of how the temporal information present in acoustic signals is processed in the auditory midbrain [81].

Of the other two divergent modules that did not show concentration of neurotransmission terms, one primarily included terms for enzymatic activity (red module), and another for transcription factor binding and citric acid cycle (black; Additional file 1: Table S9). Overall, a variety of synaptic pathways were concentrated in the significant modules ($p < 0.0001$), suggesting that neural pathways are diverging in gene expression between ancestral and reinforced populations. Note that the divergent modules only contained a marginally significant number of synaptic pathways (per module) compared to the non-significant modules ($p = 0.0561$).

Although the other 21 modules did not show significant divergence geographically, one of these was also enriched for neurotransmission-related terms, including inner ear receptor cell development and differentiation, sensory organ development, neurogenesis, and mechanoreceptor differentiation (white). The only sex-associated divergent module was enriched for terms such as transmembrane transporter activity of solutes, hormone activity, and neuroactive ligand-receptor interaction (blue; Additional file 1: Table S9).

Divergence of candidate synaptic transmission genes

We predicted that gene expression has diverged between ancestral and reinforced populations across five possible categories of proteins involved in synaptic transmission (Table 1). We identified seven candidate genes across three categories, all of which were also present in our transcriptome-wide set of differentially-expressed genes (Additional file 1: Table S4). For the ionotropic receptors (Category 1), we found no evidence of differential gene expression. Because neurophysiological studies of mid-brain neurons [57–60, 63–65] indicated that these receptors mediate excitation and inhibition of auditory

neurons, we hypothesized that down- or up-regulation of one or more genes that code for these receptors may alter how acoustic information is processed. We did not find any evidence for expression changes in ionotropic receptor genes, however, at the whole-brain level.

With respect to synaptic vesicle proteins (Category 2), we found five genes that were differentially expressed between the ancestral and reinforced populations. Two of these genes were GABA transporters (Sodium- and chloride-dependent GABA transporter [GAT2; Slc6a13] and Sodium- and chloride-dependent betaine transporter/GABA transporter [BGT1; Slc6a12]). The primary function of GABA transporters in the brain is to recycle GABA neurotransmitter from the extracellular space of the synaptic cleft and through the cell membrane into the presynaptic neuron [82]. Although expressed in brain tissue, GAT2 and BGT1, are less common in GABAergic synapses than other GABA transporters (e.g., GAT1 and GAT3). GAT2 and BGT1 are not thought to play a major role in GABA reuptake in neurons compared to GAT1 and GAT3 due to lower expression levels, lower binding affinities to GABA, and localization to other areas of the brain [83–86].

In regard to other synaptic vesicle proteins (Category 2), the other three differentially-expressed genes included Secretory carrier-associated membrane protein 3 (SCAMP3), Extended-synaptotagmin-2-B (ESYT2-b), and Zinc transporter ZIP10 (Slc39A1). The SCAMP3 gene encodes a secretory carrier membrane protein that functions as a carrier to the cell surface in post-golgi recycling pathways [87]. One function of SCAMP proteins is to bind to neurotransmitter transporters (Slc6 genes, see above) and regulate their expression at the cell surface [88, 89]. Although little is known about the role of SCAMPs in behavior, a study in *Drosophila* reported multiple abnormal behaviors in SCAMP-null mutants, underscoring the importance of this gene in neurotransmission [90].

The extended-synaptotagmins (ESYTs) are a class of endoplasmic reticulum (ER)-resident proteins that are related to but predate the synaptotagmin genes evolutionarily [91]. These proteins are thought to coordinate membrane tethering and lipid exchange between the endoplasmic reticulum and plasma membrane. Studies in *Drosophila melanogaster* of the single ESYT ortholog in this species found that neurotransmission is reduced in ESYT mutants, indicating that this gene is required to facilitate presynaptic release of neurotransmitter [92]. Furthermore, synaptic growth is promoted by over-expression of this gene [92].

Zinc transporters of the zinc importer (ZIP or Slc39A) family control the influx of zinc from outside the cell into the cytoplasm and from vesicles to increase

cytosolic zinc concentrations [93–95]. Unlike the second family of zinc transporters (ZnT or Slc30), which can be components of synaptic vesicles, ZIP family proteins can be located in the plasma membrane of neurons [95]. Zinc plays a diversity of roles in the brain, influencing neurotransmission and sensory processing and is of particular interest to the current study because its involvement with inhibition of NMDA and GABA-A receptors in the postsynaptic membrane [96–98]. Genome-wide association studies have found a strong association between ZIP zinc transporters and several psychiatric disorders [99, 100], indicating the key role of these transporters for healthy functioning of the central nervous system [97].

In terms of proteins that associate with synaptic vesicles (Category 3), Kinesin-associated protein 3 (KIFAP3 or KAP3) was differentially expressed between ancestral and reinforced populations. Kinesins are anterograde (cell body to the synaptic junctions) motor proteins that traffic vesicles by walking them along microtubules [101, 102]. The KIFAP3 gene encodes a non-motor accessory protein that forms a complex (KIF3 complex) with two kinesin 2 family proteins to generate a functional motor. This complex provides essential transport of vesicles during axonal elongation, conventional transport, and flagellar transport [103]. The KIFAP3 protein is the element within the motor complex that binds to the vesicle cargo [103, 104]. Defects in the KIF3 complex can lead to neurodegenerative diseases that are caused by problems with long-distance vesicle trafficking [105, 106].

With regard to synaptic plasma membrane proteins (Category 4), Cadherin-15 (CDH15) was divergently expressed between populations. Cadherins are calcium-dependent cell adhesion proteins occurring on both pre- and postsynaptic membranes that align the active zone to the postsynaptic membrane, thus enabling efficient neurotransmission [107, 108]. These proteins regulate dendritic and synaptic architecture and influence function and plasticity in neurons [109–111]. Cadherins have been associated with various neural disorders, for example, deletion of CDH15 results in mental retardation [112, 113]. These studies underscore the vital roles of cadherins in synapse stabilization and gene regulation during neurotransmission [109]. For the final category, proteins that reversibly associate with plasma membrane proteins (Category 5), we did not recover any differentially-expressed genes (Table 1).

Genetic basis of acoustic mating behaviors in other taxa

The genetic architecture of mating behaviors is understudied in anuran amphibians, with examples being largely limited to visual signals [114, 115]. Information regarding the genetic basis of acoustic signaling and

processing in frogs is even more scarce [116]. Nonetheless, insight can be gained from insects that exhibit temporal variation in elements of acoustic mating signals in a similar way to many frogs. This work has shown that temporal variation has a polygenic genetic architecture across diverse insect groups [117–127]. Studies in *Drosophila melanogaster* found that candidate genes controlling pulse rate of acoustic signals include ion channel genes, transcription factors, and regulators of transcription and translation [128–131]. Quantitative trait mapping studies on Hawaiian crickets (genus *Laupala*) have shown that candidate genes underlying temporal information in acoustic signals include genes encoding a cyclic nucleotide-gated ion channel, a calcium release-activated calcium channel, a signal peptide peptidase-like protein gene, a putative synaptic vesicle related protein [127], three of which fall within the bounds of synaptic transmission gene categories targeted in this study (Table 1). In frogs, [116], suggested that membrane currents of fast pulse rate neurons in the parabranchial nucleus of the brain are likely to mediate divergence of acoustic signals among *Xenopus* species and that these differences probably involve differential expression of ion channels. Though little is known about the genetic basis for species-specific acoustic differences, their ongoing work will examine gene expression differences within specific vocal and auditory neurons across this genus [116].

Caveats

One of the limitations of our study was the lack of population replication—we studied a pair of populations representing a single shift of *P. feriarum* from allopatry into sympatry along the Apalachicola River floodplain. This experimental design limits our ability to infer general, repeatable patterns of gene expression divergence. Since we have shown previously that this species independently formed contact zones with *P. nigrita* multiple times [55], our future work will investigate whether gene expression has diverged in a similar manner across contact zones or if populations diversifying through cascade reinforcement are undergoing non-parallel genetic divergence [5–7, 10]. A second limitation is that individuals were collected directly from the field under peak mating conditions, rather than raised in a common garden setting in the laboratory. Raising animals from different population treatments under the same conditions in the laboratory would remove confounding environmental variation across geographic sites. In future work, we will remove this source of variation by rearing experimental animals in the same setting. A third limitation of this study was our focus on whole brain gene expression, since relevant shifts in expression may be most pronounced in specific areas of the brain. Although this

coarse-scale approach provided the first insight into expression divergence in this species, in future work we will utilize the information gained here to predict patterns of divergence in specific brain regions where changes in neural gene expression are most likely to have occurred [60, 116]. Comparing these patterns of divergence to those seen in non-neural and non-reproductive tissues could allow us to identify expected levels of divergence due to geography and demographic factors. We expect that this fine-scale targeted approach will enable us to uncover additional genes underlying the divergence of mating behaviors in this system.

Future directions

Systems undergoing cascade reinforcement provide an exciting opportunity to study how speciation can be accelerated by spatially-varying selection pressures on reproductive traits generated through species interactions [3, 5–7]. The diversification of populations that can result suggests that different genes may be the targets of selection in different populations. Indeed, studies of parallel phenotypic evolution under similar selection pressures have often uncovered as much or more evidence of non-parallel than parallel divergence at the genomic sequence and gene expression levels [132–135] (but see [126]). In this way, the conditions of divergent selection generated by cascade reinforcement may drive the evolution of reproductive isolation via mechanisms such as gene expression divergence [38] or evolution of structural variants [36, 37] across conspecific populations. Moreover, at the broader level of gene network evolution, different elements of the same gene networks or even different networks may diverge in replicate populations, thereby accelerating the evolution reproductive isolation [39, 42].

In the current study, we identified seven candidate genes involved in synaptic transmission that are differentially-expressed geographically. Five of these genes encode synaptic vesicle proteins. Moreover, we found two gene networks with a concentration of neurotransmission terms that have diverged between ancestral and reinforced populations. Our future work will focus on testing the repeatability of evolution with respect to gene and gene network expression and genome structure in naturally-replicated populations undergoing cascade reinforcement. Moreover, by through examining specific brain regions, we will gain insight into how the genetic architecture underlying mate choice diverges during speciation.

Conclusions

As Upland chorus frog (*Pseudacris feriarum*) populations experience strong selection via cascade reinforcement, the genetic pathways underlying neural

processing of mating behaviors are predicted to diverge. Using Illumina RNA-seq and a new reference transcriptome for this species, we show evidence that multiple candidate synaptic transmission genes have diverged in expression between ancestral and reinforced populations. Moreover, gene networks with a significant concentration of neurotransmission enrichment terms have also diverged between these populations. Overall, we found greater divergence of gene expression between ancestral and reinforced populations, particularly in females, than between the sexes. Although some elements of our study design limit the generality of our conclusions, our results are consistent with the idea that divergent selection on reproductive behaviors promotes the evolution of expression in neural processing genes. Our findings lay the groundwork for future studies in non-model organisms that integrate neurogenetics, behavior, and evolution to understand the origin of new species.

Methods

Reference transcriptome sequencing

Eight tissues (brain, eye, testis, liver, heart, lung, skin, and leg muscle) from two adult allopatric male *P. feriarum* (catalog numbers ECM7962 and ECM7965, Nash County, NC, USA, N 35.95999°, W 78.07131°) were utilized to generate a reference transcriptome. The frogs in this study were wild-caught under the appropriate state scientific collecting permits for Alabama and Florida and procedures followed the Florida State University's Animal Care and Use Committee Protocol #1747. Frogs were dissected within 5 min of euthanizing via decapitation, and tissues were immediately flash-frozen in liquid nitrogen after dissection to minimize RNA degradation. Nucleic acids were extracted using a phenol-chloroform method briefly described here: Each tissue sample was homogenized in 250 μ L Trizol with a pestle and was kept on dry ice throughout the procedure. Then 50 μ L of phenol-chloroform (pH 5.5) were added followed by centrifugation (12,000 $g \times 15$ min, 4 °C) for phase separation. Nucleic acids were then precipitated with isopropyl alcohol and sodium acetate, followed by centrifugation (12,000 $g \times 10$ min, 4 °C). The pellet was washed in ethanol and remaining liquid was allowed to evaporate at room temperature. Extracted nucleic acids were then suspended in RNase-free water, flash frozen in liquid nitrogen, and stored at –80 °C. Quality analysis of the extraction was made by quantitation via Nanodrop (Thermo Fisher Scientific), Qubit (Thermo Fisher Scientific), and Bioanalyzer 2100 (Agilent Technologies). Samples with RNA Integrity Number (RIN) lower than 7.5 were discarded from further processing.

From the extracted RNA, sequencing libraries were constructed at the Georgia Genomics and Bioinformatics Core (Athens, GA, USA) using the KAPA Stranded

mRNA-Seq Kit. Poly-T magnetic beads were used to capture extracted mRNA. Heat fragmentation at 94 °C for 6 min was carried out to obtain fragments 200–300 bp in length. Synthesis of the second cDNA strand included the incorporation of dUTP for identification and posterior selective amplification of 3' cDNA. After amplification, A-tails and Illumina Truseq LT adaptors were added to the 3' cDNAs, which were subsequently quantitated by Qubit and Fragment Analyzer (Agilent Technologies). Sequencing on an Illumina HiSeq 2500 platform (with 150 bp paired-end reads) was carried out at Florida State University's College of Medicine Translational Science Laboratory.

Reference transcriptome assembly

The processing, assembly, and annotation of the reads followed a general protocol published elsewhere [136], which is briefly described here. Raw sequence files were inspected in FastQC v0.11 and MultiQC v1.7 [137, 138] to verify their quality. All reads originating from the same tissue were pooled together and trimmed with Trim Galore! v0.6 [139] using a stringency of 3 bp and a 5' hard trim of 9 bp to remove low quality bases. Reads were not de-duplicated. Trimmed reads were passed to SortMeRNA v2.1 [140] in order to filter out rRNA sequences based on similarity with the SILVA v111 and Rfam v11.0 databases [141, 142]. Reads were inspected again with FastQC and MultiQC to ensure quality of the data [143, 144]. Normalization of the data set was performed with the `insilico_read_normalization.pl` script from Trinity v2.8 [145] to obtain a maximum coverage of 50X. Normalized reads from each tissue were pooled and transcripts were assembled with Trinity v2.8 using a k-mer size of 25. Assembled transcripts were then validated with BUSCO v3 [146], run on transcriptome mode against the Tetrapoda Ortholog Database v8 [147] and with cutoff e-value = 1×10^{-6} . Other BUSCO parameters were left as default. In addition, blastx searches (cutoff E-value = 1×10^{-20}) were performed using the UniProt/SwissProt v2019_04 and the *Xenopus tropicalis* protein database (downloaded May 2019) to obtain estimates of protein length coverage.

Trinotate v3.1 [148] was used to generate gene annotations of the assembled transcripts. Annotations were collected from the UniProt/SwissProt and Pfam databases as provided by Trinotate v3.1 (updated to January 2015). The entire annotation process was automated by using the `autoTrinotate.pl` script included in the Trinotate distribution. The script used detection of long open reading frames ("LongOrfs") to identify candidate coding regions in TransDecoder v5.5 [145]. The reference transcripts and the predicted TransDecoder peptides were matched against the UniProt/SwissProt database using blastx and blastp searches (E-value = 1×10^{-5}). In addition, protein

family assignments were made by hmmscan v3.2 by querying the Pfam database. The results of the blast and hmmscan searches were then loaded into an SQLite database included in Trinotate v3.1. The resulting annotations were summarized with trinotateR [149] as implemented in R v3.6 [150]. A flowchart of the reference transcriptome assembly and annotation process is shown in the Supplemental material (Additional file 10: Fig. S9).

Brain gene expression sequencing and processing

To detect differences in brain gene expression associated with reproductive character displacement, whole brains were dissected from 20 *P. feriarum*, ten of which were obtained from a population sympatric with *P. nigrita* (Liberty County, FL, USA, N 30.10145°, W 85.09951°), and ten of which were obtained from an allopatric population (Macon County, AL, USA, N 32.44314°, W 85.65523°; Fig. 1, Additional file 1: Table S10). We refer to the allopatric site as the ancestral population and the sympatric site as reinforced population throughout this paper because our phylogeographic and population genomic analyses indicate that the historical pattern of range expansion and current direction of gene flow is from allopatric into sympatric regions [55].

Our experimental design corresponds to a single shift by *P. feriarum* from allopatry to sympatry with *P. nigrita* without replication along the Apalachicola River drainage, corresponding to one of the expansions identified by Banker et al. [55]. Although more populations could have been sampled within the immediate geographic area to increase replication, sampling independent shifts in other river drainages would not have been an appropriate approach to replication for the questions we address in this study, due to the disparate phenotypic outcomes observed [51] and divergent gene expression profiles expected [42] in other contact zones. Frogs sampled from each population represented the two sexes in a 50:50 ratio. All frogs were captured in amplexus in the field while mating and laying eggs. Pairs were separated and immediately euthanized via decapitation as described above (*Methods—Reference transcriptome sequencing*). Females were not acclimated to laboratory conditions because we aimed to examine brain gene expression while frogs were at the breeding peak. Following Florida State University's Animal Care and Use Committee protocol #1747, brains were removed in the field within 1 min of decapitation, transferred to cryotubes, flash-frozen in liquid nitrogen, and then stored at -80 °C. RNA extraction, quality analysis, and library preparation were performed as described above. Three of the samples (two allopatric males and one sympatric male) produced poor-quality libraries and were not carried forward in downstream processing analyses. Sequencing of the 17 remaining libraries was carried out

on an Illumina NovaSeq 6000 platform, with a target sequencing effort of 200 million reads per sample.

After raw reads were quality checked with FastQC/MultiQC, a Bayesian approach [151] was used to remove of adapters and merge of read pairs. Low quality bases at the start of the reads were trimmed with hts_CutTrim v1.0 [152] by setting a hard trim of 10 bp on the 5' end. After PCR duplicates were removed with hts_SuperDeduper v1.0 [152], a second quality check was performed using FastQC/MultiQC. RNA reads were mapped to our *P. feriarum* reference transcriptome by using HISAT v2.1 [137], then BAM files were produced in SAMtools v1.9 [138]. Raw read counts per reference transcript were obtained directly from the BAM files using a custom script (see *Availability of data and materials*). One sample containing fewer than 20 million post-filtered read pairs was removed from downstream analyses (see Results).

Raw read counts per gene/transcript were collated and used as input for edgeR [139]. After CPM were calculated, transcripts with average CPM < 1 across all samples were discarded from the data set [153]. Calculation of CPM was made using a prior.count = 2 in order to avoid undefined logarithms when calculating log₂ CPM. To account for large differences in gene counts among libraries, normalization factors for each sample were obtained by applying a Trimmed Mean of M-values (TMM) [154]. The resulting values were visualized using multidimensional scaling in order to check for outlier samples (plotMDS function used in R). As no outlier samples were obvious (Additional file 4: Fig. S3), the downstream analyses were conducted using all 16 of the samples.

Tagwise (gene-wise) dispersions were estimated and outlier effects were reduced with the estimateDisp function in edgeR (using the robust = T option). Tagwise (gene-wise) dispersions were estimated and outlier effects were reduced with the estimateDisp function in edgeR (using the robust = T option). After calculating the Biological Coefficient of Variation (BCV, Additional file 11: Fig. S10), negative binomial generalized linear models (GLM) were fitted for each gene. A flow-chart of the RNA-Seq read processing into raw counts and differential expression analyses is available in the Supplemental material (Additional file 10: Fig. S9).

Identification of candidate gene transcripts involved in synaptic transmission

We predicted that some genes involved in synaptic transmission during the processing of acoustic information should be differentially expressed between reinforced and non-reinforced populations. We compiled a list of these genes based on Sudhof (Table 9–1 in [67]) and Luo (Table 3–1 in [68]). This list included synaptic vesicle proteins, proteins that associate with synaptic vesicles, synaptic plasma membrane proteins, and

proteins that reversibly associate with plasma membrane proteins during synaptic transmission (Table 1). Because ionotropic receptors (AMPA-type and NMDA-type glutamate and GABA_A-type receptors) have been shown experimentally to mediate processing of temporal information in anuran acoustic signals [61, 64, 65], we also included genes for these receptors (Table 1). We identified candidate transcripts in our reference transcriptome by searching for appropriate keywords (Additional file 12) in the transcriptome annotation (see *Availability of data and materials*). Some broad search terms were restricted further if many irrelevant matches were recovered in preliminary searches. These restrictions followed functional descriptions by Mignogna and D'Adamo ([155]; Rab proteins), Hirokawa et al. ([103]; kinesins), and Pfister et al. ([156]; dyneins) and are defined in Additional file 12.

Testing candidate synaptic transmission transcripts for differential expression

We tested for significant differential expression between three pairs of samples: sympatric vs. allopatric females, sympatric vs. allopatric males, and all sympatric vs. all allopatric frogs. Using the GLM model coefficients corresponding to these three comparisons, we tested whether or not each candidate gene was differentially expressed between each comparison. Likelihood ratio tests (LRT) were carried out using the glmLRT function from edgeR. Specifically, the LRT compared a GLM including the average log-counts of two given groups as coefficients to a null GLM in which both average log-counts are equal [157]. False-discovery rate corrections were applied to tests from each comparison [158]. Genes with $q < 0.05$ in one or more comparison were regarded as differentially expressed.

Testing for unequal expression divergence among the four groups

We evaluated the magnitude of gene expression evolution in each of the four groups and compared the magnitudes among the groups to determine if more evolution had taken place in any of the groups. Following [159], we performed the following steps: 1) we standardized the expression levels for each gene to a mean of zero and a standard deviation of one, 2) we separated the expression data into two sets, one for candidate and one for non-candidate genes, 3) we computed the average expression level for each group/gene/set combination (producing eight values for each gene), 4) we computed a distance matrix (using the *dist* function in R) for each gene and for each set representing the pairwise divergence between four groups, and 5) we estimated from each distance matrix an expression divergence tree using neighbor joining (*nj* method in R).

The resulting tip branch lengths were interpreted as the relative contribution of each group to the overall expression divergence. In Additional file 5: Fig. S4, we represent for each group the relative contribution as the \log_2 ratio of the branch lengths for two groups. To test for significance, we repeated steps 3–5 for each gene, computed the average branch length difference across the genes within the set (the test statistic), and performed randomization test in which the assignment of candidate and non-candidate was shuffled among the genes before recomputing the test-statistic. The null distribution was comprised of 10,000 replicates. In order to test whether each difference was higher in the candidate than non-candidate genes, we compared the difference (candidate minus non-candidate) to a null distribution generated using the approach above. Note that zero length branches for individual genes required that differences were used instead of ratios. The p -values from these tests (8 tests corresponding to the group \times set combinations, and 4 tests to compare the two sets) were used to compute FDR to correct for 12 total tests.

Evaluating transcriptome-wide patterns of divergence

In order to elucidate general expression patterns in the entire transcriptome data set, we expanded our analysis to all transcripts for the following 6 comparisons: (1) all sympatric vs. allopatric frogs, (2) all female vs. male frogs, (3) sympatric vs. allopatric females, (4) sympatric vs. allopatric males, (5) male vs. female allopatric frogs, and (6) male vs. female sympatric frogs. For the last four comparisons, a group-based analysis was performed in which each sample was assigned to a group by geography and sex. As above, to ascertain whether or not each gene was differentially expressed between each comparison, likelihood ratio tests (LRT) were carried out using the `glmLRT` function from `edgeR`. Specifically, the LRT compared a GLM including the average log-counts of two given groups as coefficients to a null GLM in which both average log-counts are equal [157]. False-discovery rate corrections were applied to each comparison and genes with $FDR < 0.05$ [158] were regarded as differentially expressed. We used randomization tests with 10,000 null replicates to determine: (1) whether more genes are differentially expressed by geography than by sex, (2) whether more genes are differentially expressed by geography in females than in males, and (3) whether more genes are differentially expressed by sex in sympatry than in allopatry. In each case, we computed the test statistic as the ratio of the number of differentially expressed genes (e.g. `nDiffExpressedGeo/nDiffExpressedSex`). To compute each sample from the null distribution, we first randomly shuffled numbers between the first and second category for each gene separately (e.g., shuffled values horizontally in a table with

one column for each category and one row per gene), then recomputed the test statistic using the permuted Table. P -values were computed as the proportion of null replicate values that were greater than or equal to the test statistic.

Based on the resulting differentially expressed genes from the four comparisons mentioned above, a multi-dimensional scaling plot (MDS) was created to verify concordance of group membership with expression profiles using the `plotMDS` function. Heatmaps were also generated using normalized CPM of differentially expressed genes using the `Heatmap` function from the `ComplexHeatmap` and `RColorBrewer` packages [160, 161]. To obtain statistical support (approximately unbiased, AU p -values) and improve clustering of samples in the heatmaps, a hierarchical clustering function was applied using the `pvclust` and `dendextend` packages [69, 151]. Detection of differentially-expressed genes associated with differences between sympatric and allopatric frogs was also performed by pooling all sympatric samples in a single group, and all allopatric samples in another group (comparison 1). Similarly, tests for differences between females and males was carried out (comparison 2). All analyses were performed in R and graph support was provided by the `gplots` package [162].

Weighted correlation network of brain genes

A weighted correlation network analysis was performed to detect groups of genes (modules) that showed patterns of co-expression across samples, geography (allopatry or sympatry), and sex (male or female). Co-expression networks were constructed with the R package `WGCNA` [163] as follows: TMM-normalized raw counts were obtained from `edgeR`, and in order to reduce noise from low-variation transcripts as suggested elsewhere [164], only the top 10% of genes with the highest variation in CPM across samples were retained. The resulting data set was checked for missing entries and low-variation genes using `WGCNA`'s function `goodSampleGenes`. In `WGCNA`, networks are assumed to follow a scale-free topology fashion, a property of biological networks in which only a few genes are highly connected, resembling a power function [165]. A scale-free topology is obtained by elevating gene correlations coefficients to a power (β) that enhances the differences between weak and strong correlations. The function `pickSoftThreshold` was used to iteratively ascertain which β approximated a scale-free network. The procedure fits a curve between the number of connections and the frequency of those numbers of connections in the network. A threshold was set to obtain a correlation coefficient $R^2 > 0.9$ for the aforementioned curve, resulting in $\beta = 6$ (Additional file 13: Fig. S11). The transformed correlations, representing the adjacency matrix, were

then converted to a Topological Overlap Matrix (TOM) which describes the overall degree of connectedness of each gene to other genes.

Module identification—

Genes were clustered based on expression level similarity using TOM and a hierarchical clustering method implemented in the `hclust` function of R. Gene models were defined by cutting the resulting dendrogram at a height of 0.99 (R function `cuttreeDynamic` with module size ≥ 40 genes and `deepsplit = 3`), Modules were further merge by grouping highly correlated modules based on overall module expression, with modules being merged when eigengene dissimilarities were less than 0.3. The correlation between the module expression levels and the frogs' geography or sex were also determined using the eigengenes. The resulting p -values were corrected for multiple tests (25 modules \times two groupings = 50 tests) using the q value function in R. A threshold of $\alpha = 0.05$ was used to determine significance.

Module analysis—

Each of the resulting modules was visualized in Cytoscape v3.7.2 [166] after filtering of low gene-gene correlations ($r < 0.05$). Pathway enrichment analyses were performed using gene annotations in the modules from Uniprot using the R package `g:Profiler2` [167]. The enrichment analysis selected gene pathways enriched at an $FDR < 0.1$ by obtaining annotations from the Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG), and TRANSFAC databases [168, 169]. The background gene set was all the annotated genes in our reference transcriptome. The enrichment analyses were made based on the unfiltered modules (i.e., low correlations were not filtered out from modules).

Randomization tests pertaining to modules—

In order to determine if synaptic pathways were concentrated in each of the 25 modules and also overall, we conducted randomization tests. To this aim, we performed a search for each enriched pathway (Additional file 1: Table S9) in the EMBL-EBI QuickGo database (Binns et al. 2009). If the terms “synap-,” (for synaptic transmission) or “neuro-,” or “neura-” (for neurotransmission) was found in the Ancestor Chart or Child Terms list or in the name of the enriched pathway itself, we counted it as a synaptic pathway. To test for enrichment of synaptic pathways in each separate module we used (as the test statistic) the number of synaptic pathways for that module. To test if the differentially expressed modules were enriched for synaptic pathways overall, we used (as the test statistic) the total number of synaptic pathways in the differentially expressed modules.

To simulate the null distribution to evaluate these two test statistics, we shuffled the assignment of pathway to module (holding the total number of pathways per module constant) and recomputed the test statistics (10,000 null samples were simulated). Finally, to test whether the number of synaptic pathways per differentially expressed module was higher than expected, we computed as the test statistic the average number of synaptic pathways per differentially expressed module. This test statistic was evaluated by shuffling the designation of significant/non-significant across the modules (while keeping the number of synaptic pathways pathways constant). Again 10,000 null samples were simulated. The resulting p -values for the 25 individual modules were corrected for multiple tests using the q value function in R. A threshold of $FDR < 0.05$ was used to determine significance in each test.

We also conducted an additional randomization test to determine whether modules associated with geography contained significantly more synaptic pathways than modules not associated with geography. We followed the same procedure described in the previous test, but with a single test statistic being calculated: $\text{meanNSynapticGeo}/\text{meanNSynapticNonGeo}$. Again, $\alpha = 0.05$ was used to determine significance.

Abbreviations

ADP: Adenosine diphosphate; AMPA: α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid; AU: Approximately unbiased p -value; BCV: Biological coefficient of variation; BUSCO: Benchmarking Universal Single-Copy Orthologs; cADPR: Cyclic ADP ribose; CPM: Counts per million; DE: Differentially expressed gene; FDR: False-discovery rate; F_{57} : Fixation index; GABA: Gamma aminobutyric acid; GAT: GABA transporter; GLM: Generalized linear model; GO: Gene ontology; IEG: Immediate early gene; KEGG: Kyoto Encyclopedia of Genes and Genomes; LRT: Likelihood ratio tests; LTD: Long-term depression; LTP: Long-term potentiation; MDS: Multi-dimensional scaling; NADH: Nicotinamide adenine dinucleotide; NC: Neighborhood connectivity; N_e : Effective population size; NMDAR: N-methyl-D-aspartate receptor; NO: Nitric oxide; ORF: Open reading frame; RIN: RNA integrity number; SGN: Spiral ganglion neuron; SLC: Solute carrier; SNARE: SNAP receptor; SPG: Synaptic plasticity gene; TE: Transposable element; TMM: Trimmed mean of M -values; TOM: Topological overlap matrix; WGCNA: Weighted correlation network analysis

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-021-07995-3>.

Additional file 1: Table S1. Number of read pairs, average sequence length, and percentage of sequences with Phred Score. **Table S2.** Coverage of assembled transcripts against proteins in the Uniprot/Swissprot and *Xenopus tropicalis* databases. **Table S3.** Sample information and statistics of brain RNA-Seq reads used in the differential expression analysis. Please see the main text Methods for detail on read processing and quality control. **Table S4.** Differential Expression of Candidate Synaptic Transmission Genes between reinforced (sympatric) and ancestral (allopatric) populations. The 554 candidate transcripts for which expression data were available are shown. Significance of likelihood ratio tests conducted were corrected for multiple tests within each of the three sets of samples: 1) both sexes combined (left section), 2) females only (center section), and 3) males only (right section). Q-values resulting

from the correction were compared to $\alpha = 0.05$ for significance (highlighted). Note that log fold change (logFC) is expressed so that positive values mean that sympatric populations are overexpressed relative to allopatry, whereas negative values mean that sympatric populations are underexpressed relative to allopatry. **Table S5.** Results of randomization tests comparing divergence evolution. The test statistic for the first eight tests is the difference between the branch lengths for the 2 compared groups, averaged across the genes in the set. *Q*-values were obtained by correcting the *p*-values for twelve tests. **Table S6.** Differential Expression of all Transcripts between reinforced (sympatric) and ancestral (allopatric) populations (left three sections), as well as between males and females (right three sections). Significance of likelihood ratio tests conducted were corrected for multiple tests within each of the six sets of samples. FDR values resulting from the correction were compared to $\alpha = 0.05$ for significance (FDR shown for significant tests only). Note that log fold change (logFC) is expressed so that positive values mean that sympatric populations are overexpressed relative to allopatry, whereas negative values mean that sympatric populations are underexpressed relative to allopatry. Note LogFC is the effect size, with positive values corresponding to overexpression in sympatry. **Table S7.** Identity, gene ontology (GO) pathways, and co-expression network module membership of differentially expressed genes. **Table S8.** Module properties and hub genes. The hub genes were selected via the chooseTopHubInEachModule function in WGCNA. The secondary annotated hub gene was defined as the gene, other than the hub gene, with the highest degree (number of connections) that included annotation. Neighborhood connectivity of a node refers to the average connectivity of all its neighbors. The clustering coefficient of a node is the ratio of connections of its neighbors and the maximum number of possible connections between neighbors. The number of differentially expressed genes between any comparison is shown for each module, as well as the number of possible social/mate preference, synaptic plasticity (SPG), and immediate early genes (IEG).

Table S9. Trait-Module correlation analysis and enriched pathways (based on the *Xenopus tropicalis* and *Homo sapiens* Gene Ontology databases) at FDR < 0.1 via g:Profiler2. Significance of correlation with Geography and Sex were corrected simultaneously (50 tests) to produce *q*-values. Pathway terms containing keywords ("synap-" in red and "neuro-" and "neura-" in green in Enriched pathways column) were analyzed to determine whether they were concentrated in each module (see Methods). Significance values, based on randomization tests (see Methods), for these 25 tests were corrected for multiple tests to produce *q*-values that were compared to 0.05 to determine significance. **Table S10.** Information of the samples used in the differential gene expression analyses. The RIN (RNA Integrity Number) are provided as measured in a Agilent Bioanalyzer. Ratios are shown as measured from Qubit (Thermo Fisher).

Additional file 2: Figure S1. The number of transcripts from the reference transcriptome mapping to each Gene Ontology category after gene annotation with Trinotate.

Additional file 3: Figure S2. Number of reads post filter for the 17 samples for which data were collected. The bottom-most sample was excluded from downstream analyses because it produced substantially fewer reads than the other 16 samples. Each bar is labeled with the corresponding Sample IDs from Additional file 1: Table S10.

Additional file 4: Figure S3. Multidimensional scaling (MDS) plot of pairwise expression differences, based on normalized \log_2 counts per million for genes in the transcriptome-wide gene set. This plot was used to verify that no outliers existed. Numbers inside points correspond to the last two digits of the sample IDs in Additional file 1: Table S10.

Additional file 5: Figure S4. A comparison of expression evolution. Tip branch lengths on the two neighbor joining trees at the top represent the amount of expression evolution in each of the four groups compared: sympatric males (SM), allopatric males (AM), sympatric females (SF), and allopatric females (AF). The bar graph below indicates the ratios of branch lengths used to compare the relative expression evolution between two groups. Error bars show the 95% bootstrap confidence interval for each comparison. Additional file 1: Table S5 presents results of randomization tests used to determine if each ratio was significantly different from zero and if ratios for candidate loci were significantly

greater than corresponding ratios from non-candidate loci. None of the tests were significant at the 0.05 level after correcting for multiple tests.

Additional file 6: Figure S5. Relative expression levels (\log_2 CPM) of 517 differentially expressed genes between sympatric and allopatric frogs. The dendrograms resulted from hierarchical clustering of expression levels after 100 replicates to estimate Approximately Unbiased *p*-values (numbers on nodes). Gene names are Uniprot identifiers and can be found in the reference transcriptome annotation file (see *Availability of data and materials*).

Additional file 7: Figure S6. Relative expression levels (\log_2 CPM) of 129 differentially expressed genes between female and male frogs. The dendrograms resulted from hierarchical clustering of expression levels after 100 replicates to estimate Approximately Unbiased *p*-values (numbers on nodes). Gene names are Uniprot identifiers and can be found in the reference transcriptome annotation file (see *Availability of data and materials*).

Additional file 8: Figure S7. Relative expression levels (\log_2 CPM) across all samples of 24 differentially expressed genes between females and males in allopatry (A), and 34 differentially expressed genes between females and males in sympatry (B). The dendrograms resulted from hierarchical clustering of expression levels after 100 replicates to estimate Approximately Unbiased *p*-values (numbers on nodes). Gene names are Uniprot identifiers and can be found in the reference transcriptome annotation file (see *Availability of data and materials*).

Additional file 9: Figure S8. Correlations of geography (left column; allopatry/sympatry) or sex (right column; male/female) with each of the module's eigengenes (rows). The color of the cells indicates positive (orange) or negative (blue) correlations. The numbers in the cell are Pearson's coefficients with associated FDR in parentheses, which were computed by correcting *p*-values for the 50 tests. Significantly correlated traits are shown in bold and italics.

Additional file 10: Figure S9. Flowchart describing computational steps to process, annotate, assemble RNA-Seq reads for the reference transcriptome (green boxes), and steps to process and count reads of brain RNA-Seq for analysis of differential gene expression analysis of candidate transcripts (orange boxes) and all transcripts (blue boxes).

Additional file 11: Figure S10. Biological Coefficient of Variation (BCV) for each transcript in the data set. Each dot represents the average variation in transcript counts among samples (tagwise dispersion). The baseline reference used to adjust tagwise variation is indicated by dark gray line, showing the common dispersion. The model used to adjust the tagwise variation yielded a trend variation (light gray line).

Additional file 12. List of keywords used in the detection of synaptic transmission genes within the reference transcriptome. The list was created by querying the existing literature (see Methods-Identification of candidate genes transcripts involved in synaptic transmission).

Additional file 13: Figure S11. The effect of raising node correlations in the WGCNA analysis to a power β (soft threshold) on the fit (R^2) to the assumption of a scale-free topology network. In a scale-free topology network, some nodes are highly connected (hub genes). The red line indicates a fit of $R^2 = 0.9$ to a scale-free topology, indicating that gene-correlations should be elevated to a $\beta = 6$.

Acknowledgements

We thank Craig Guyer (Auburn University) for offering support during fieldwork and specimen collection. We are grateful to Kim Hughes, Darin Rokyta, and Joseph Travis from Florida State University for their advice in developing and designing this study, and to Gary Rose (University of Utah) for providing comments that improved the quality of this manuscript.

Authors' contributions

EML, MD and AL designed the study. EML, AL, and MD collected the frog specimens. MLK and MD performed the molecular laboratory techniques. OEO and AL analyzed the brain differential expression data and constructed weighted correlation networks. OEO, EML, AL, and CZ interpreted the results. OEO, AL, SH, and DS collaborated in the construction of the reference transcriptome. OEO, EML, and AL wrote the manuscript. All authors read and approved the final manuscript.

Funding

This project was funded the National Science Foundation, Division of Environmental Biology (DEB) through grant #1120516 to EML. This funding agency did not contribute to the design of the study or in the collection, analysis, and interpretation of data, or in the writing and editing of this manuscript.

Availability of data and materials

Raw reads used to draw conclusions in this article are stored in the Sequence Read Archive (NCBI-SRA BioProject PRJNA723357). The reference transcriptome, assembled contigs, and gene annotations are stored in the Zenodo repository under Digital Object Identifier (DOI): <https://doi.org/10.5281/zenodo.4709988> (<https://zenodo.org/record/4709988#.YR7a3pNKhaw>). Scripts and input files, including the matrix with read raw counts per sample are stored at Github (https://github.com/oospina/brain_gene_expr_P_feriarum). Databases were obtained from the following sources: SILVA v111 (https://www.arb-silva.de/no_cache/download/archive/release_111); Rfam v11.0 (<http://ftp.ebi.ac.uk/pub/databases/Rfam/11.0>); Tetrapoda Ortholog Database (OrthoDB) v8 (<https://www.orthodb.org/v8/index.html>); Uniprot (*Xenopus tropicalis*, <https://www.uniprot.org/proteomes/UP000008143>); EMBL-EBI QuickGo (www.ebi.ac.uk/QuickGO/).

Declarations

Ethics approval and consent to participate

Procedures were reviewed and approved by the Animal Care and Use Committee at Florida State University (protocol #1747).

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Biomedical Sciences, College of Veterinary Medicine, Iowa State University, 1800 Christensen Drive, 50011 Ames, IA, USA. ²Present address: Department of Biostatistics and Bioinformatics, Moffitt Cancer Center, 13131 USF Magnolia Drive, Tampa, FL 33612, USA. ³Department of Scientific Computing, Florida State University, 400 Dirac Science Library, Tallahassee, FL 32306, USA. ⁴Present address: Genetics and Genomics Program, Iowa State University, 2437 Pammel Drive, Ames, IA 50011, USA. ⁵Present address: Department of Molecular Genetics and Microbiology, Genetics Institute, University of Florida, Gainesville, FL 32610, USA.

Received: 2 December 2020 Accepted: 10 September 2021

Published online: 02 October 2021

References

- Dobzhansky T. Speciation as a stage in evolutionary divergence. *Am Nat.* 1940;74(753):312–21. <https://doi.org/10.1086/280899>.
- Blair WF. Mating call and stage of speciation in the *Microhyla olivacea-M. carolinensis* complex. *Evolution (N Y).* 1955;9:469–80.
- Howard DJ. Reinforcement: origin, dynamics, and fate of an evolutionary hypothesis. In: Hybrid zones and the evolutionary process; 1993. p. 46–69.
- Servedio MR, Noor MAF. The role of reinforcement in speciation: theory and data. *Annu Rev Ecol Syst.* 2003;34(1):339–64. <https://doi.org/10.1146/annurev.ecolsys.34.011802.132412>.
- Ortiz-Barrientos D, Grealy A, Nosil P. The genetics and ecology of reinforcement: implications for the evolution of prezygotic isolation in sympatry and beyond. *Ann N Y Acad Sci.* 2009;1168(1):156–82. <https://doi.org/10.1111/j.1749-6632.2009.04919.x>.
- Pfennig KS, Pfennig DW. Character displacement: ecological and reproductive responses to a common evolutionary problem. *Q Rev Biol.* 2009;84(3):253–76. <https://doi.org/10.1086/605079>.
- Hoskin CJ, Higgie M. Speciation via species interactions: the divergence of mating traits within species. *Ecol Lett.* 2010;13:409–20.
- Fuller RC. Reconciling concepts, theory, and empirical patterns surrounding cascade reinforcement. *Curr Zool.* 2016;62(2):131–4. <https://doi.org/10.1093/cz/zow011>.
- Pfennig KS, Ryan MJ. Reproductive character displacement generates reproductive isolation among conspecific populations: an artificial neural network study. *Proc R Soc B Biol Sci.* 2006;273(1592):1361–8. <https://doi.org/10.1098/rspb.2005.3446>.
- Yukilevich R, Aoki F. Is cascade reinforcement likely when sympatric and allopatric populations exchange migrants? *Curr Zool.* 2016;62(2):155–67. <https://doi.org/10.1093/cz/zow007>.
- McPeck MA, Gavrillets S. The evolution of female mating preferences: differentiation from species with promiscuous males can promote speciation. *Evolution (N Y).* 2006;60:1967–80.
- Thompson JN. The geographic mosaic of coevolution: University of Chicago Press; 2005. <https://doi.org/10.7208/chicago/9780226118697.001.0001>.
- Calabrese GM, Pfennig KS. Reinforcement and the proliferation of species. *J Hered.* 2020;111(1):138–46. <https://doi.org/10.1093/jhered/esz073>.
- Kozak GM, Roland G, Rankhorn C, Falater A, Berdan EL, Fuller RC. Behavioral isolation due to cascade reinforcement in *Lucania* killifish. *Am Nat.* 2015; 185(4):491–506. <https://doi.org/10.1086/680023>.
- Littlejohn M, Loftus-Hills J. Premating isolation in the *Hyla ewingi* complex (Anura: Hylidae). *Evolution (N Y).* 1968;22:659–63.
- Hoskin CJ, Higgie M, McDonald KR, Moritz C. Reinforcement drives rapid allopatric speciation. *Nature.* 2005;437(7063):1353–6. <https://doi.org/10.1038/nature04004>.
- Pfennig KS, Rice AM. Reinforcement generates reproductive isolation between neighbouring conspecific populations of spadefoot toads. *Proc R Soc B Biol Sci.* 2014;281:20140949.
- Zouros E, D'Entremont CJ. Sexual isolation among populations of *Drosophila mojavensis*: Response to pressure from a related species. *Evolution (N Y).* 1980;34:421–30.
- Higgie M, Blows MW. The evolution of reproductive character displacement conflicts with how sexual selection operates within a species. *Evolution (N Y).* 2008;62(5):1192–203. <https://doi.org/10.1111/j.1558-5646.2008.00357.x>.
- Bewick ER, Dyer KA. Reinforcement shapes clines in female mate discrimination in *Drosophila subquinaria*. *Evolution (N Y).* 2014;68:3082–94.
- Humphreys DP, Rundle HD, Dyer KA. Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? *Curr Zool.* 2016;62:183–91.
- Dyer KA, White BE, Sztepanacz JL, Bewick ER, Rundle HD. Reproductive character displacement of epicuticular compounds and their contribution to mate choice in *Drosophila subquinaria* and *Drosophila recens*. *Evolution (N Y).* 2014;68:1163–75.
- Rundle HD, Dyer KA. Reproductive character displacement of female mate preferences for male cuticular hydrocarbons in *Drosophila subquinaria*. *Evolution (N Y).* 2015;69:2625–37.
- Nosil P, Crespi BJ, Sandoval CP. Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proc R Soc B Biol Sci.* 2003;270(1527):1911–8. <https://doi.org/10.1098/rspb.2003.2457>.
- Porretta D, Urbanelli S. Evolution of premating reproductive isolation among conspecific populations of the sea rock-pool beetle *Ochthebius urbanelliae* driven by reinforcing natural selection. *Evolution (N Y).* 2012; 66(4):1284–95. <https://doi.org/10.1111/j.1558-5646.2011.01535.x>.
- Hopkins R, Rausher MD. Identification of two genes causing reinforcement in the Texas wildflower *Phlox drummondii*. *Nature.* 2011;469:411–4.
- Hopkins R, Rausher MD. The cost of reinforcement: selection on flower color in allopatric populations of *Phlox drummondii*. *Am Nat.* 2014;183(5):693–710. <https://doi.org/10.1086/675495>.
- Garner AG, Goulet BE, Farnitano MC, Molina-Henao YF, Hopkins R. Genomic signatures of reinforcement. *Genes (Basel).* 2018;9:191.
- Ortiz-Barrientos D, Reiland J, Hey J, Noor MAF. Recombination and the divergence of hybridizing species. *Genetica.* 2002;116(2/3):167–78. <https://doi.org/10.1023/A:1021296829109>.
- Ortiz-Barrientos D, Counterman BA, Noor MAF. The genetics of speciation by reinforcement. *PLoS Biol.* 2004;2:e416.
- Sæther SA, Sætre GP, Borge T, Wiley C, Svedin N, Andersson G, et al. Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science (80-).* 2007;318:95–7.
- Hopkins R, Levin DA, Rausher MD. Molecular signatures of selection on reproductive character displacement of flower color in *Phlox drummondii*. *Evolution (N Y).* 2012;66(2):469–85. <https://doi.org/10.1111/j.1558-5646.2011.01452.x>.
- Smadja CM, Loire E, Caminade P, Thoma M, Latour Y, Roux C, et al. Seeking signatures of reinforcement at the genetic level: a hitchhiking mapping and candidate gene approach in the house mouse. *Mol Ecol.* 2015;24(16):4222–37. <https://doi.org/10.1111/mec.13301>.

34. Loire E, Tusso S, Caminade P, Severac D, Boursot P, Ganem G, et al. Do changes in gene expression contribute to sexual isolation and reinforcement in the house mouse? *Mol Ecol*. 2017;26:5189–202.
35. Kirkpatrick M, Ravigné V. Speciation by natural and sexual selection: Models and experiments. *Am Nat*. 2002;159(3 Suppl):S22–35.
36. North HL, Caminade P, Severac D, Belkhir K, Smadja CM. The role of copy-number variation in the reinforcement of sexual isolation between the two European subspecies of the house mouse. *Philos Trans R Soc Lond Ser B Biol Sci*. 2020;375(1806):20190540. <https://doi.org/10.1098/rstb.2019.0540>.
37. Berdan EL, Fuller RC, Kozak GM. Genomic landscape of reproductive isolation in *Lucania killifish*: the role of sex loci and salinity. *J Evol Biol*. 2021; 34:157–74.
38. Mack KL, Nachman MW. Gene regulation and speciation. *Trends Genet*. 2017;33:68–80.
39. Yang CH, Scarpino SV. Reproductive barriers as a byproduct of gene network evolution. *bioRxiv*. 2020:1–27. <https://doi.org/10.1101/2020.06.12.147322>.
40. Dover GA, Flavell RB. Molecular coevolution: DNA divergence and the maintenance of function. *Cell*. 1984;38(3):622–3. [https://doi.org/10.1016/0092-8674\(84\)90255-1](https://doi.org/10.1016/0092-8674(84)90255-1).
41. Seidl F, Levis NA, Jones CD, Monroy-Eklund A, Ehrenreich IM, Pfennig KS. Variation in hybrid gene expression: implications for the evolution of genetic incompatibilities in interbreeding species. *Mol Ecol*. 2019;28(20):4667–79. <https://doi.org/10.1111/mec.15246>.
42. Johnson NA, Porter AH. Rapid speciation via parallel, directional selection on regulatory genetic pathways. *J Theor Biol*. 2000;205:527–42.
43. Smadja C, Ganem G. Subspecies recognition in the house mouse: a study of two populations from the border of a hybrid zone. *Behav Ecol*. 2002;13(3):312–20. <https://doi.org/10.1093/beheco/13.3.312>.
44. Ganem G, Ginane C, Ostrowski MF, Orth A. Assessment of mate preference in the house mouse with reference to investigations on assortative mating. *Biol J Linn Soc*. 2005;84(3):461–71. <https://doi.org/10.1111/j.1095-8312.2005.00447.x>.
45. Ganem G, Litel C, Lenormand T. Variation in mate preference across a house mouse hybrid zone. *Heredity (Edinb)*. 2008;100(6):594–601. <https://doi.org/10.1038/hdy.2008.20>.
46. Latour Y, Perriat-Sanguinet M, Caminade P, Boursot P, Smadja CM, Ganem G. Sexual selection against natural hybrids may contribute to reinforcement in a house mouse hybrid zone. *Proc R Soc B Biol Sci*. 2013;281(1776):20132733. <https://doi.org/10.1098/rspb.2013.2733>.
47. Smadja C, Ganem G. Divergence of odorant signals within and between the two European subspecies of the house mouse. *Behav Ecol*. 2008;19(1):223–30. <https://doi.org/10.1093/beheco/arm127>.
48. Pfennig KS. Female spadefoot toads compromise on mate quality to ensure conspecific matings. *Behav Ecol*. 2000;11(2):220–7. <https://doi.org/10.1093/beheco/11.2.220>.
49. Pfennig KS, Simovich MA. Differential selection to avoid hybridization in two toad species. *Evolution (N Y)*. 2002;56:1840–8.
50. Pfennig KS. Facultative mate choice drives adaptive hybridization. *Science (80-)*. 2007;318:965–7.
51. Lemmon EM. Diversification of conspecific signals in sympatry: geographic overlap drives multidimensional reproductive character displacement in frogs. *Evolution (N Y)*. 2009;63(5):1155–70. <https://doi.org/10.1111/j.1558-5646.2009.00650.x>.
52. Lemmon EM, Lemmon AR. Reinforcement in chorus frogs: lifetime fitness estimates including intrinsic natural selection and sexual selection against hybrids. *Evolution (N Y)*. 2010;64:1748–61.
53. Lemmon ECM. Patterns and processes of speciation in North American chorus frogs (*Pseudacris*). Austin: University of Texas; 2007.
54. Lemmon EM, Lemmon AR, Collins JT, Lee-Yaw JA, Cannatella DC. Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs (*Pseudacris*). *Mol Phylogenet Evol*. 2007;44:1068–82.
55. Banker SE, Lemmon AR, Hassinger AB, Dye M, Holland SD, Kortyna ML, et al. Hierarchical hybrid enrichment: multitermed genomic data collection across evolutionary scales, with application to chorus frogs (*Pseudacris*). *Syst Biol*. 2020;69(4):756–73. <https://doi.org/10.1093/sysbio/syz074>.
56. Lemmon AR, Lemmon EM. A likelihood framework for estimating phylogeographic history on a continuous landscape. *Syst Biol*. 2008;57(4):544–61. <https://doi.org/10.1080/10635150802304761>.
57. Rose G, Capranica RR. Temporal selectivity in the central auditory system of the leopard frog. *Science*. 1983;219(4588):1087–9. <https://doi.org/10.1126/science.6600522>.
58. Rose GJ, Capranica RR. Sensitivity to amplitude modulated sounds in the anuran auditory nervous system. *J Neurophysiol*. 1985;53(2):446–65. <https://doi.org/10.1152/jn.1985.53.2.446>.
59. Edwards CJ, Alder TB, Rose GJ. Auditory midbrain neurons that count. *Nat Neurosci*. 2002;5:934–6.
60. Rose GJ. Time computations in anuran auditory systems. *Front Physiol*. 2014; 5:1–7. <https://doi.org/10.3389/fphys.2014.00206>.
61. Rose GJ. The numerical abilities of anurans and their neural correlates: insights from neuroethological studies of acoustic communication. *Philos Trans R Soc B Biol Sci*. 2017;373(1740):20160512. <https://doi.org/10.1098/rstb.2016.0512>.
62. Rose GJ, Hanson JL, Leary CJ, Graham JA, Alluri RK, Vasquez-Opazo GA. Species-specificity of temporal processing in the auditory midbrain of gray treefrogs: interval-counting neurons. *J Comp Physiol A Neuroethol Sensory Neural Behav Physiol*. 2015;201:485–503.
63. Hanson JL, Rose GJ, Leary CJ, Graham JA, Alluri RK, Vasquez-Opazo GA. Species specificity of temporal processing in the auditory midbrain of gray treefrogs: long-interval neurons. *J Comp Physiol A*. 2016;202:67–79.
64. Alluri RK, Rose GJ, Hanson JL, Leary CJ, Vasquez-Opazo GA, Graham JA, et al. Phasic, suprathreshold excitation and sustained inhibition underlie neuronal selectivity for short-duration sounds. *Proc Natl Acad Sci U S A*. 2016;113(13):E1927–35. <https://doi.org/10.1073/pnas.1520971113>.
65. Alluri RK, Rose GJ, Leary CJ, Palaparthi A, Hanson JL, Vasquez-Opazo GA, et al. How auditory selectivity for sound timing arises: the diverse roles of GABAergic inhibition in shaping the excitation to interval-selective midbrain neurons. *Prog Neurobiol*. 2021;199:101962. <https://doi.org/10.1016/j.pneurobio.2020.101962>.
66. Schunter C, Vollmer SV, Macpherson E, Pascual M. Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics*. 2014;15:167.
67. Südhof T. Composition of synaptic vesicles. In: Siegel GJ, Agranoff BW, Albers RW, et al., editors. *Basic neurochemistry: Molecular, cellular and medical aspects*. 6th ed. Philadelphia: Lippincott-Raven; 1999. p. 7–9.
68. Luo L. Principles of neurobiology, Garland Science; 2015. <https://doi.org/10.1201/9781317553830>.
69. Suzuki R, Shimodaira H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*. 2006;22(12):1540–2. <https://doi.org/10.1093/bioinformatics/btl117>.
70. Endler JA. Signals, signal conditions, and the direction of evolution. *Am Nat*. 1992;139:S125–53. <https://doi.org/10.1086/285308>.
71. Boughman JW. How sensory drive can promote speciation. *Trends Ecol Evol*. 2002;17(12):571–7. [https://doi.org/10.1016/S0169-5347\(02\)02595-8](https://doi.org/10.1016/S0169-5347(02)02595-8).
72. Boughman JW. Divergent sexual selection enhances reproductive isolation in sticklebacks. *Nature*. 2001;411(6840):944–8. <https://doi.org/10.1038/35082064>.
73. Malone JH, Ribado J, Lemmon EM. Sensory drive does not explain reproductive character displacement of male acoustic signals in the upland chorus frog (*Pseudacris feriarum*). *Evolution (N Y)*. 2014;68(5):1306–19. <https://doi.org/10.1111/evo.12366>.
74. Omerik JM, Griffith GE. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. *Environ Manag*. 2014;54(6):1249–66. <https://doi.org/10.1007/s00267-014-0364-1>.
75. Kuriyama K, Ohkuma S. Role of nitric oxide in central synaptic transmission: effects on neurotransmitter release. *Jpn J Pharmacol*. 1995;69:1–8.
76. Vincent SR. Nitric oxide neurons and neurotransmission. *Prog Neurobiol*. 2010;90(2):246–55. <https://doi.org/10.1016/j.pneurobio.2009.10.007>.
77. Agustín-Pavón C, Martínez-Ricós J, Martínez-García F, Lanuza E. Role of nitric oxide in pheromone-mediated intraspecific communication in mice. *Physiol Behav*. 2009;98:608–13.
78. Hellier V, Brock O, Candlish M, Desrozières E, Aoki M, Mayer C, et al. Female sexual behavior in mice is controlled by kisspeptin neurons. *Nat Commun*. 2018;9(1):400. <https://doi.org/10.1038/s41467-017-02797-2>.
79. Dauwalder B. Mate choice: should I mate or should I go? *Curr Biol*. 2020; 30(3):R118–20. <https://doi.org/10.1016/j.cub.2019.12.001>.
80. Ishimoto H, Kamikouchi A. A feedforward circuit regulates action selection of pre-mating courtship behavior in female *Drosophila*. *Curr Biol*. 2020;30:396–407.e4.
81. Naud R, Houtman D, Rose GJ, Longtin A. Counting on dis-inhibition: a circuit motif for interval counting and selectivity in the anuran auditory system. *J Neurophysiol*. 2015;114:2804–15.
82. Scimemi A. Structure, function, and plasticity of GABA transporters. *Front Cell Neurosci*. 2014;8:1–14.

83. Takanaga H, Ohtsuki S, Hosoya KI, Terasaki T. GAT2/BGT-1 as a system responsible for the transport of γ -aminobutyric acid at the mouse blood-brain barrier. *J Cereb Blood Flow Metab.* 2001;21:1232–9.
84. Zhou Y, Danbolt NC. GABA and glutamate transporters in brain. *Front Endocrinol (Lausanne).* 2013;4:1–14.
85. Kempson SA, Zhou Y, Danbolt NC. The betaine/GABA transporter and betaine: roles in brain, kidney, and liver. *Front Physiol.* 2014;5:1–16.
86. Kicking S, Hellsberg E, Frølund B, Schousboe A, Ecker GF, Wellendorp P. Structural and molecular aspects of betaine-GABA transporter 1 (BGT1) and its relation to brain function. *Neuropharmacology.* 2019;161:107644.
87. O'Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44:D733–45.
88. Müller HK, Wiborg O, Haase J. Subcellular redistribution of the serotonin transporter by secretory carrier membrane protein 2. *J Biol Chem.* 2006;281(39):28901–9. <https://doi.org/10.1074/jbc.M602848200>.
89. Fjorback AW, Müller HK, Haase J, Raarup MK, Wiborg O. Modulation of the dopamine transporter by interaction with secretory carrier membrane protein 2. *Biochem Biophys Res Commun.* 2011;406(2):165–70. <https://doi.org/10.1016/j.bbrc.2011.01.069>.
90. Zheng JLC, Tham CT, Keatings K, Fan S, Liou AYC, Numata Y, et al. Secretory carrier membrane protein (SCAMP) deficiency influences behavior of adult flies. *Front Cell Dev Biol.* 2014;2:1–9. <https://doi.org/10.3389/fcell.2014.00064>.
91. Saheki Y, De Camilli P. The extended-Synaptotagmins. *Biochim Biophys Acta - Mol Cell Res.* 1864;2017:1490–3.
92. Kikuma K, Li X, Kim D, Sutter D, Dickman DK. Extended synaptotagmin localizes to presynaptic ER and promotes neurotransmission and synaptic growth in *Drosophila*. *Genetics.* 2017;207(3):993–1006. <https://doi.org/10.1534/genetics.117.300261>.
93. Eide DJ. Zinc transporters and the cellular trafficking of zinc. *Biochim Biophys Acta - Mol Cell Res.* 1763;2006:711–22.
94. Bosomworth HJ, Thornton JK, Coneyworth LJ, Ford D, Valentine RA. Efflux function, tissue-specific expression and intracellular trafficking of the Zn transporter ZnT10 indicate roles in adult Zn homeostasis. *Metallomics.* 2012;4(8):771–9. <https://doi.org/10.1039/c2mt20088k>.
95. Hara T, Aki TT, Takagishi T, Fukue K, Kambe T, Fukada T. Physiological roles of zinc transporters: molecular and genetic importance in zinc homeostasis. *J Physiol Sci.* 2017;67(2):283–301. <https://doi.org/10.1007/s12576-017-0521-4>.
96. Tóth K. Zinc in neurotransmission. *Annu Rev Nutr.* 2011;31(1):139–53. <https://doi.org/10.1146/annurev-nutr-072610-145218>.
97. Krall RF, Tzounopoulos T, Aizenman E. The function and regulation of zinc in the brain. *Neuroscience.* 2021;457:235–58. <https://doi.org/10.1016/j.neuroscience.2021.01.010>.
98. Krall RF, Moutal A, Phillip MB, Asraf H, Johnso JW, Khanna R, et al. Synaptic zinc inhibition of NMDA receptors depends on the association of GluN2A with the zinc transporter ZnT1. *Sci Adv.* 2020;6:eabb1515.
99. Baum AE, Hamshere M, Green E, Cichon S, Rietschel M, Noethen MM, et al. Meta-analysis of two genome-wide association studies of bipolar disorder reveals important points of agreement. *Mol Psychiatry.* 2008;13:466–7.
100. Fullard JF, Charney AW, Voloudakis G, Uzilov AV, Haroutunian V, Roussos P. Assessment of somatic single-nucleotide variation in brain tissue of cases with schizophrenia. *Transl Psychiatry.* 2019;9:21.
101. Wang W, Cao L, Wang C, Gigant B, Knossow M. Kinesin, 30 years later: recent insights from structural studies. *Protein Sci.* 2015;24:1047–56.
102. Siddiqui N, Straube A. Intracellular cargo transport by kinesin. *Biochem.* 2017;82:803–15.
103. Hirokawa N, Noda Y, Tanaka Y, Niwa S. Kinesin superfamily motor proteins and intracellular transport. *Nat Rev Mol Cell Biol.* 2009;10(10):682–96. <https://doi.org/10.1038/nrm2774>.
104. Verhey KJ, Kaul N, Soppina V. Kinesin assembly and movement in cells. *Annu Rev Biophys.* 2011;40(1):267–88. <https://doi.org/10.1146/annurev-biophys-042910-155310>.
105. Mandelkow E, Mandelkow E-M. Kinesin motors and disease. *Trends Cell Biol.* 2002;12:585–91.
106. Alsabban AH, Morikawa M, Tanaka Y, Takei Y, Hirokawa N. Kinesin Kif3b mutation reduces NMDAR subunit NR 2A trafficking and causes schizophrenia-like phenotypes in mice. *EMBO J.* 2020;39:1–19.
107. Benson DL, Huntley GW. Building and remodeling synapses. *Hippocampus.* 2012;22(5):954–68. <https://doi.org/10.1002/hipo.20872>.
108. Friedman LG, Benson DL, Huntley GW. Cadherin-based transsynaptic networks in establishing and modifying neural connectivity. In: *Physiology & behavior*; 2015. p. 415–65.
109. Basu R, Taylor MR, Williams ME. The classic cadherins in synaptic specificity. *Cell Adhes Migr.* 2015;9:193–201.
110. Seong E, Yuan L, Arikath J. Cadherins and catenins in dendrite and synapse morphogenesis. *Cell Adhes Migr.* 2015;9:202–13.
111. Arikath J, Reichardt LF. Cadherins and catenins at synapses: roles in synaptogenesis and synaptic plasticity. *Trends Neurosci.* 2008;31(9):487–94. <https://doi.org/10.1016/j.tins.2008.07.001>.
112. Bhalla K, Luo Y, Buchan T, Beachem MA, Guzauskas GF, Ladd S, et al. Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. *Am J Hum Genet.* 2008;83(6):703–13. <https://doi.org/10.1016/j.ajhg.2008.10.020>.
113. Hirano S, Takeichi M. Cadherins in brain morphogenesis and wiring. *Physiol Rev.* 2012;92(2):597–634. <https://doi.org/10.1152/physrev.00014.2011>.
114. Rodríguez A, Mundy NI, Ibáñez R, Pröhl H. Being red, blue and green: the genetic basis of coloration differences in the strawberry poison frog (*Oophaga pumilio*). *BMC Genomics.* 2020;21(1):1–16. <https://doi.org/10.1186/s12864-020-6719-5>.
115. Stuckert AMM, Chouteau M, McClure M, LaPolice TM, Linderoth T, Nielsen R, et al. The genomics of mimicry: gene expression throughout development provides insights into convergent and divergent phenotypes in a Müllerian mimicry system. *Mol Ecol.* 2021;30(16):4039–61. <https://doi.org/10.1111/mec.16024>.
116. Kelley DB, Ballagh IH, Barkan CL, Bendsky A, Elliott TM, Evans BJ, et al. Generation, coordination, and evolution of neural circuits for vocal communication. *J Neurosci.* 2020;40(1):22–36. <https://doi.org/10.1523/JNEUROSCI.0736-19.2019>.
117. Shaw KL. Polygenic inheritance of a behavioral phenotype: interspecific genetics of song in the hawaiian cricket genus *Laupala*. *Evolution (N Y).* 1996;50:256–66.
118. Williams MA, Blouin AG, Noor MAF. Courtship songs of *Drosophila pseudoobscura* and *D. persimilis*. II. Genetics of species differences. *Heredity (Edinb).* 2001;86:68–77.
119. Xu M, Shaw KL. Extensive linkage and genetic coupling of song and preference loci underlying rapid speciation in *Laupala* crickets. *J Hered.* 2021;112(2):204–13. <https://doi.org/10.1093/jhered/esab001>.
120. Henry CS, Wells MLM, Holsinger KE. The inheritance of mating songs in two cryptic, sibling lacewing species (Neuroptera: Chrysopidae: *Chrysoperla*). *Genetica.* 2002;116(2/3):269–89. <https://doi.org/10.1023/A:1021240611362>.
121. Gleason JM, Ritchie MG. Do quantitative trait loci (QTL) for a courtship song difference between *Drosophila simulans* and *D. sechellia* coincide with candidate genes and intraspecific QTL? *Genetics.* 2004;166(3):1303–11. <https://doi.org/10.1534/genetics.166.3.1303>.
122. Saldamando CI, Miyaguchi S, Tatsuha H, Kishino H, Bridle JR, Butlin RK. Inheritance of song and stridulatory peg number divergence between *Chorthippus brunneus* and *C. jacobsi*, two naturally hybridizing grasshopper species (Orthoptera: Acrididae). *J Evol Biol.* 2005;18:703–12.
123. Shaw KL, Parsons YM, Lesnick SC. QTL analysis of a rapidly evolving speciation phenotype in the hawaiian cricket *Laupala*. *Mol Ecol.* 2007;16:2879–92.
124. Ellison CK, Wiley C, Shaw KL. The genetics of speciation: genes of small effect underlie sexual isolation in the hawaiian cricket *Laupala*. *J Evol Biol.* 2011;24(5):1110–9. <https://doi.org/10.1111/j.1420-9101.2011.02244.x>.
125. Limousin D, Streiff R, Courtois B, Dupuy V, Alem S, Greenfield MD. Genetic architecture of sexual selection: QTL mapping of male song and female receiver traits in an acoustic moth. *PLoS One.* 2012;7(9):e44554. <https://doi.org/10.1371/journal.pone.0044554>.
126. Blankers T, Oh KP, Shaw KL. Parallel genomic architecture underlies repeated sexual signal divergence in hawaiian *Laupala* crickets. *Proc R Soc B Biol Sci.* 2019;286(1912):20191479. <https://doi.org/10.1098/rspb.2019.1479>.
127. Xu M, Shaw KL. The genetics of mating song evolution underlying rapid speciation: linking quantitative variation to candidate genes for behavioral isolation. *Genetics.* 2019;211(3):1089–104. <https://doi.org/10.1534/genetics.118.301706>.
128. Gleason JM. Mutations and natural genetic variation in the courtship song of *Drosophila*. *Behav Genet.* 2005;35(3):265–77. <https://doi.org/10.1007/s10151-005-3219-y>.
129. Turner TL, Miller PM, Cochrane VA. Combining genome-wide methods to investigate the genetic complexity of courtship song variation in *Drosophila melanogaster*. *Mol Biol Evol.* 2013;30:2113–20.
130. Fedotov SA, Bragina JV, Besedina NG, Danilenkova LV, Kamysheva EA, Panova AA, et al. The effect of neurospecific knockdown of candidate genes

- for locomotor behavior and sound production in *Drosophila melanogaster*. Fly (Austin). 2014;8(3):176–87. <https://doi.org/10.4161/19336934.2014.983389>.
131. Fedotov SA, Bragina JV, Besedina NG, Danilenkova LV, Kamysheva EA, Kamyshev NG. Gene CG15630 (fipi) is involved in regulation of the interpulse interval in *Drosophila* courtship song. J Neurogenet. 2018;32(1):15–26. <https://doi.org/10.1080/01677063.2017.1405000>.
 132. Bolnick DI, Barrett RDH, Oke KB, Rennison DJ, Stuart YE. (Non) Parallel Evolution. Annu Rev Ecol Evol Syst. 2018;49:303–30.
 133. Rivas MJ, Saura M, Pérez-Figueroa A, Panova M, Johansson T, André C, et al. Population genomics of parallel evolution in gene expression and gene sequence during ecological adaptation. Sci Rep. 2018;8(1):1–12. <https://doi.org/10.1038/s41598-018-33897-8>.
 134. Fischer EK, Roland AB, Moskowitz NA, Vidoudez C, Ranaivoroza N, Tapia EE, et al. Mechanisms of convergent egg provisioning in poison frogs. Curr Biol. 2019;29:4145–4151.e3.
 135. Jacobs A, Carruthers M, Yurchenko A, Gordeeva NV, Alekseyev SS, Hooker O, et al. Parallelism in ecomorphology and gene expression despite variable evolutionary and genomic backgrounds in a Holarctic fish. PLoS Genet. 2020;16(4).
 136. Griffith M, Walker JR, Spies NC, Ainscough BJ, Griffith OL. Informatics for RNA sequencing: a web resource for analysis on the cloud. PLoS Comput Biol. 2015;11(8):1–20. <https://doi.org/10.1371/journal.pcbi.1004393>.
 137. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat Methods. 2015;12:357–60.
 138. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics. 2009;25:2078–9.
 139. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
 140. Kopylova E, Noé L, Touzet H. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics. 2012;28:3211–7.
 141. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yara P, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41(D1):590–6. <https://doi.org/10.1093/nar/gks1219>.
 142. Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA families database. Nucleic Acids Res. 2009;37(Suppl. 1):136–40. <https://doi.org/10.1093/nar/gkn766>.
 143. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
 144. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarizing analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32(19):3047–8. <https://doi.org/10.1093/bioinformatics/btw354>.
 145. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. De novo transcript sequence reconstruction from RNA-seq using the trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
 146. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–2. <https://doi.org/10.1093/bioinformatics/btv351>.
 147. Kriventseva EV, Tegenfeldt F, Petty TJ, Waterhouse RM, Simão FA, Pozdnyakov IA, et al. OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software. Nucleic Acids Res. 2015;43:D250–6.
 148. Bryant DM, Johnson K, DiTommaso T, Tickle T, Couger MB, Payzin-Dogru D, et al. A tissue-mapped axolotl de novo transcriptome enables identification of limb regeneration factors. Cell Rep. 2017;18:762–76.
 149. Stubben C. trinator. 2016. <https://github.com/cstubben/trinatorR>.
 150. R Core Team. R: A language and environment for statistical computing (Version 3.5) [Software]. 2018. <https://www.r-project.org/>.
 151. Galili T. Dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics. 2015;31(22):3718–20. <https://doi.org/10.1093/bioinformatics/btv428>.
 152. HTStream. <https://github.com/s4hts/HTStream>.
 153. Li X, Brock GN, Rouchka EC, Cooper NGF, Wu D, O'Toole TE, et al. A comparison of per sample global scaling and per gene normalization methods for differential expression analysis of RNA-seq data. PLoS One. 2017;12:1–22.
 154. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010;11:1–9.
 155. Mignogna ML, D'Adamo P. Critical importance of RAB proteins for synaptic function. Small GTPases. 2018;9(1-2):145–57. <https://doi.org/10.1080/21541248.2016.1277001>.
 156. Kevin Pfister K, Fisher EMC, Gibbons IR, Hays TS, Holzbaur ELF, Richard McIntosh J, et al. Cytoplasmic dynein nomenclature. J Cell Biol. 2005;171(3):411–3. <https://doi.org/10.1083/jcb.200508078>.
 157. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012;40(10):4288–97. <https://doi.org/10.1093/nar/gks042>.
 158. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
 159. Somel M, Liu X, Tang L, Yan Z, Hu H, Guo S, et al. MicroRNA-driven developmental remodeling in the brain distinguishes humans from other primates. PLoS Biol. 2011;9(12):e1001214. <https://doi.org/10.1371/journal.pbio.1001214>.
 160. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics. 2016;32(18):2847–9. <https://doi.org/10.1093/bioinformatics/btw313>.
 161. Neuwirth E. RColorBrewer. 2014. <https://cran.r-project.org/web/packages/RColorBrewer/RColorBrewer.pdf>.
 162. Warnes GR, Bolker B, Bonebakker L, Gentleman R, Huber W, Liaw A, et al. gplots: Various R programming tools for plotting data, R Package version, vol 2; 2009. p. 1.
 163. Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008;9(1):1–13. <https://doi.org/10.1186/1471-2105-9-559>.
 164. Zhang J, Nie Q, Si C, Wang C, Chen Y, Sun W, et al. Weighted gene co-expression network analysis for RNA-sequencing data of the varicose veins transcriptome. Front Physiol. 2019;10:1–10. <https://doi.org/10.3389/fphys.2019.00278>.
 165. Zhang B, Horvath S. A general framework for weighted gene co-expression network analysis. Stat Appl Genet Mol Biol. 2005;4(1). <https://doi.org/10.2202/1544-6115.1128>.
 166. Shannon P, Markiel A, Ozier O, Baliga NS, Wang J, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13(11):2498–504. <https://doi.org/10.1101/gr.1239303>.
 167. Raudvere U, Kolberg L, Kuzmin I, Arak T, Adler P, Peterson H, et al. G:profiler: a web server for functional enrichment analysis and conversions of gene lists (2019 update). Nucleic Acids Res. 2019;47(W1):W191–8. <https://doi.org/10.1093/nar/gky369>.
 168. Kanehisa M, Sato Y, Furumichi M, Morishima K, Tanabe M. New approach for understanding genome variations in KEGG. Nucleic Acids Res. 2019;47(D1):D590–5. <https://doi.org/10.1093/nar/gky962>.
 169. The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25:25–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

