Received: 11 March 2020 | Revised: 21 May 2020 | Accepted: 1 June 2020

DOI: 10.1111/mec.15512

ORIGINAL ARTICLE

MOLECULAR ECOLOGY WILEY

Ecological divergence in sympatry causes gene misexpression in hybrids

Joseph A. McGirr[1] | Christopher H. Martin[1,2]

[1]Department of Biology, University of North Carolina, Chapel Hill, NC

[2]Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, CA

Correspondence
Christopher H. Martin, Department of Integrative Biology and Museum of Vertebrate Zoology, University of California, Berkeley, CA 94720, USA.
Email: chmartin@berkeley.edu

Funding information
National Institute of Dental and Craniofacial Research, Grant/Award Number: 7 R01 DE 027052; NSF CAREER, Grant/Award Number: 1938571

Abstract

Ecological speciation occurs when reproductive isolation evolves as a byproduct of adaptive divergence between populations. Selection favouring gene regulatory divergence between species could result in transgressive levels of gene expression in F1 hybrids that may lower hybrid fitness. We combined 58 resequenced genomes with 124 transcriptomes to identify patterns of hybrid gene misexpression that may be driven by adaptive regulatory divergence within a young radiation of *Cyprinodon* pupfishes, which consists of a dietary generalist and two trophic specialists—a molluscivore and a scale-eater. We found more differential gene expression between closely related sympatric specialists than between allopatric generalist populations separated by 1,000 km. Intriguingly, 9.6% of genes that were differentially expressed between sympatric species were also misexpressed in F1 hybrids. A subset of these genes were in highly differentiated genomic regions and enriched for functions important for trophic specialization, including head, muscle and brain development. These regions also included genes that showed evidence of hard selective sweeps and were significantly associated with oral jaw length—the most rapidly diversifying skeletal trait in this radiation. Our results indicate that divergent ecological selection in sympatry can contribute to hybrid gene misexpression which may act as a reproductive barrier between nascent species.

KEYWORDS

allele-specific expression, Dobzhansky–Muller incompatibility, ecological speciation, gene misexpression, gene misregulation, RNAseq

1 | INTRODUCTION

Adaptive radiations showcase dramatic instances of biological diversification resulting from ecological speciation, which occurs when reproductive isolation evolves as a byproduct of adaptive divergence between populations (Nosil, 2012; Schluter, 2000). Ecological speciation predicts that populations adapting to different niches will accumulate genetic differences due to divergent ecological selection, indirectly resulting in reduced gene flow. Gene regulation is a major target of selection during adaptive divergence, with many known cases of divergent gene regulation underlying ecological traits (Abzhanov, Protas, Grant, Grant, & Tab in, 2011; Jones et al., 2012; Manceau, Domingues, Mallarino, & Hoekstra, 2011; Parry et al., 2005; Thompson et al., 2018). However, it is still unknown whether selection on gene regulation can also contribute to reproductive isolation during ecological speciation (Mack & Nachman, 2017; Pavey, Collin, Nosil, & Rogers, 2010).

Hybridization between divergent populations can break up coadapted genetic variation, resulting in (Bateson) Dobzhansky–Muller incompatibilities (DMIs) if divergent alleles from parental populations are incompatible in hybrids and cause reduced fitness (Coyne & Orr, 2004; Orr, 1996). DMIs between divergent

regulatory alleles can contribute to patterns of hybrid gene misexpression: transgressive expression levels that are significantly higher or lower in F1 hybrids than either parental population. Because gene expression is largely constrained by stabilizing selection, gene misexpression is expected to disrupt highly coordinated developmental processes and reduce hybrid fitness (Bedford & Hartl, 2009; Signor & Nuzhdin, 2018). Indeed, crosses between distantly related species show that hybrid gene misexpression may be associated with strong intrinsic postzygotic isolation in the form of hybrid sterility and inviability (Landry, Hartl, & Ranz, 2007; Mack, Campbell, & Nachman, 2016; Ortíz-Barrientos, Counterman, & Noor, 2007), although other studies found no association (Guerrero, Posto, Moyle, & Hahn, 2016; Kerwin & Sweigart, 2020; Wei, Clark, & Barbash, 2014). Emerging empirical evidence suggests that weak intrinsic DMIs segregate within natural populations (Corbett-detig, Zhou, Clark, Hartl, & Ayroles, 2013) and are abundant between recently diverged species, reaching hundreds of incompatibility loci within swordtail fish hybrid zones (Schumer & Brandvain, 2016; Schumer et al., 2014). Additionally, hybrid gene misexpression has been reported at early stages of divergence within a species of intertidal copepod (Barreto, Pereira, & Burton, 2015) and between young species of lake whitefish (Renaut, Nolte, & Bernatchez, 2009).

Since most studies on hybrid gene misexpression examine distantly related species pairs that exhibit strong intrinsic isolation, the role of regulatory divergence during speciation with gene flow remains largely unexplored. Furthermore, it is debated whether hybrid gene misexpression is driven largely by stabilizing selection or directional selection at early stages of species divergence. Under stabilizing selection, hybrid gene misexpression can evolve due to compensatory *cis*- and *trans*-acting variants with opposing effects on expression levels (Landry et al., 2005; Mack & Nachman, 2017; Signor & Nuzhdin, 2018; Tulchinsky, Johnson, Watt, & Porter, 2014). Compensatory evolution results in similar gene expression levels between species even though the underlying regulatory machinery has diverged (True & Haag, 2001; Wray et al., 2003). Alternatively, directional selection could favour regulatory alleles causing divergent gene expression between species that are incompatible in hybrids (Kulmuni & Westram, 2017; Pavey et al., 2010). In this scenario, the same genes showing expression divergence between species should also show misexpression in hybrids.

We examined genetic variation and gene expression divergence within an adaptive radiation to test whether genetic variants causing adaptive gene expression divergence between species may negatively interact to cause gene misexpression in F1 hybrids. If hybrid gene misexpression was influenced by adaptive divergence during ecological speciation, we predicted that (a) gene expression divergence and hybrid gene misexpression should evolve more quickly between ecologically diverged populations compared to populations adapted to similar ecological niches, (b) many of the same genes differentially expressed between species should also show misexpression in F1 hybrids, and (c) these genes should influence adaptive phenotypes and show signs of directional selection. We tested these predictions in a young (10 kya), sympatric radiation of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas.

This radiation consists of a dietary generalist and two derived specialists adapted to novel trophic niches: a molluscivore (*Cyprinodon brontotheroides*) and a scale-eater (*Cyprinodon desquamator*) (Martin & Wainwright, 2013a). All three species coexist in multiple hypersaline lake populations within the same littoral habitat. This system is one of the few examples of a multipeak adaptive landscape measured for multiple species, which was estimated using field enclosures in the wild (Martin, 2016a; Martin & Wainwright, 2013b). F2 hybrids generated from F1 hybrid intercrosses and backcrosses to all three species exhibited a continuum of phenotypes that were used to estimate relationships between hybrid fitness and phenotypic resemblance to parental types (Martin & Wainwright, 2013b). These experiments combined with recent feeding kinematic experiments showed that hybrids exhibit reduced fitness in the wild and impaired feeding performance in the laboratory (Martin & Wainwright, 2013b; St. John, Holzman, & Martin, 2020).

Here we took a genome-wide approach to identify genetic variation underlying F1 hybrid gene misexpression and found 125 genes that were misexpressed, showed high genetic differentiation between species, and were strikingly enriched for developmental functions related to trophic specialization. Our findings suggest that regulatory variation underlying adaptive changes in gene expression can interact to cause hybrid gene misexpression, which may contribute to reduced hybrid fitness and restrict gene flow between sympatric populations.

2 | METHODS

2.1 | Study system and sample collection

Our genomic data set consisted of 51 wild-caught individuals from nine isolated hypersaline lakes on San Salvador Island, Bahamas, plus seven individuals from outgroup populations across the Caribbean (see Supplemental Methods). Our total mRNA transcriptomic data set consisted of 124 *Cyprinodon* exomes from laboratory-reared embryos collected between 2017 and 2018 (Table S1). We collected fishes for breeding from two hypersaline lakes on San Salvador Island, Bahamas (Osprey Lake and Crescent Pond); Lake Cunningham, New Providence Island, Bahamas; and Fort Fisher, North Carolina, United States.

In order to understand how varying levels of genetic divergence and ecological divergence between parents affected gene expression patterns in F1 offspring, we performed 11 separate crosses falling into three categories. (a) For purebred crosses, we collected F1 embryos from breeding tanks containing multiple breeding pairs from a single location. (b) For San Salvador Island species crosses, we crossed a single individual of one species with a single individual of another species from the same lake for all combinations of the three San Salvador Island species. In order to control for maternal effects on gene expression inheritance, we collected samples from reciprocal crosses for

three of the San Salvador Island species crosses. (c) For outgroup generalist crosses, we crossed a Crescent Pond generalist male with a Lake Cunningham female and a North Carolina female (Table S1).

2.2 | Sequencing and variant discovery

Genomic resequencing libraries were prepared using TruSeq library preparation kits and sequenced on Illumina 150PE HiSeq4000. We mapped a total of 1,953,034,511 adaptor-trimmed reads to the *Cyprinodon* reference genome (Lencer, Warren, Harrison, & Mccune, 2017) with the Burrows-Wheeler Alignment Tool (Li & Durbin, 2009). We extracted RNA from a total of 348 individuals across two early developmental stages (2 days postfertilization [dpf] and 8 dpf) using RNeasy Mini Kits (Qiagen, Inc.). For 2 dpf libraries, we pooled five embryos together and pulverized them in a 1.5-ml Eppendorf tube. We used the same extraction method for samples collected at 8 dpf but did not pool larvae. Libraries were prepared using TruSeq stranded mRNA kits and sequenced on three lanes of Illumina 150 PE HiSeq4000 at the Vincent J. Coates Genomic Sequencing Center. We mapped 1,638,067,612 adaptor-trimmed reads to the reference genome using the RNAseq aligner STAR with default parameters (Dobin et al., 2013). We did not find a difference between species or outgroup populations for standard quality control measures, (Figure S1; ANOVA, $p > .1$), except for a marginal difference in transcript integrity numbers (Figure S2; ANOVA, $p = .041$) driven by slightly higher transcript quality in North Carolina generalist samples relative to other samples (Tukey post hoc test: $p = .043$). We found no significant differences among San Salvador Island generalists, molluscivores, scale-eaters and outgroups in the proportion of reads that mapped to annotated features of the *Cyprinodon* reference genome (Figure S3; ANOVA, $p = .17$). We did find that more reads mapped to features in 2 dpf samples than 8 dpf samples (Figure S4; Student's *t* test, $p < 2.2 \times 10^{-16}$).

We used the Genome Analysis Toolkit (DePristo et al., 2011) to call and refine SNP variants across 58 *Cyprinodon* genomes and across 124 *Cyprinodon* exomes. We filtered both SNP data sets to include individuals with a genotyping rate above 90% and SNPs with minor allele frequencies higher than 5%. Our final filtered genomic SNP data set included 13,838,603 variants with a mean sequencing coverage of 8.2 × per individual. We further refined our transcriptomic SNP data set using the allele-specific software WASP (v. 0.3.3) to correct for potential mapping biases that would influence tests of allele-specific expression (Degner et al., 2009; Van De Geijn, Mcvicker, Gilad, & Pritchard, 2015). We recalled SNPs using unbiased BAMs determined by WASP for a final transcriptomic SNP data set that included 413,055 variants with a mean coverage of 1,060 × across features per individual.

2.3 | Phylogenetic analyses

In order to determine the relationship between F1 hybrid misexpression, gene expression divergence between parental populations and phylogenetic distance between parental populations, we estimated a maximum-likelihood tree (Figure S5) using RAxML (Stamatakis, 2014). We excluded all missing sites and sites with more than one alternate allele from our genomic SNP data set, leaving 1,737,591 variants across 58 individuals for analyses. We performed ten separate searches with different random starting trees under the GTRGAMMA model. Node support was estimated from 1,000 bootstrap samples. We fit phylogenetic generalized least-squares (PGLS) models in R with the packages ape (Paradis & Schliep, 2019) and nlme to assess whether gene expression patterns were associated with geographic distance among populations after accounting for phylogenetic relatedness among populations and species. We excluded Osprey Lake populations from these analyses because outgroups were only crossed with Crescent Pond generalists.

2.4 | Population genomics and genome-wide association mapping

If alleles causing gene expression divergence between species affect the development of adaptive traits, and also cause gene misexpression in hybrids resulting in low fitness, we predicted that genomic regions near these genes would be strongly differentiated between species, associated with divergent ecological traits, and show signatures of positive selection. We measured relative genetic differentiation (*Fst*), within-population diversity (π) and between-population divergence (*Dxy*) across 58 *Cyprinodon* individuals using 13.8 million SNPs (Tables S2 and S3). We identified 20 kb genomic windows significantly associated with variation in oral jaw size across all populations in our data set (Figure S6). We measured upper jaw lengths and standard length for all individuals in our genomic data set using digital callipers, fit a log-transformed jaw length by log-transformed standard length linear regression to correct for body size and used the residuals for genome-wide association mapping with the software GEMMA (Zhou, Carbonetto, & Stephens, 2013). This program accounts for population structure by incorporating a genetic relatedness matrix into a Bayesian sparse linear mixed model which calculates a posterior inclusion probability (PIP) indicating the proportion of Markov chain Monte Carlo iterations in which a SNP was estimated to have a nonzero effect on phenotypic variation. We used Tajima's D statistic and the software SweeD (Pavlidis, Živković, Stamatakis, & Alachiotis, 2013) to identify shifts in the site frequency spectrum characteristic of hard selective sweeps. We performed gene ontology enrichment analyses for candidate gene sets using ShinyGo (Ge, Jung, & Yao, 2020).

2.5 | Hybrid misexpression and inheritance of gene expression patterns

We aggregated read counts with featureCounts (Liao, Smyth, & Shi, 2014) at the transcript isoform level (36,511 isoforms corresponding to 24,952 protein coding genes). Significant differential

expression between groups was determined with DESEQ2 (Love, Huber, & Anders, 2014) using Wald tests comparing normalized posterior log fold change estimates and correcting for multiple testing using the Benjamini–Hochberg procedure with a false discovery rate of 0.05 (Benjamini & Hochberg, 1995). We compared expression in F1 hybrids to expression in F1 purebred offspring to determine whether genes showed additive, dominant or transgressive patterns of inheritance in hybrids. To categorize hybrid inheritance for F1 offspring generated from a cross between a female from population A and a male from population B ($F1_{(A \times B)}$), we conducted four pairwise differential expression tests with DESEQ2: (a) $F1_{(A)}$ versus $F1_{(B),}$ (b) $F1_{(A)}$ versus $F1_{(A \times B),}$ (c) $F1_{(B)}$ versus $F1_{(A \times B)}$ and (d) $F1_{(A)} + F1_{(B)}$ versus $F1_{(A \times B)}$. Hybrid inheritance was considered additive if hybrid gene expression was intermediate between parental populations and significantly different between parental populations. Inheritance was dominant if hybrid expression was significantly different from one parental population but not the other. Genes showing misexpression in hybrids showed transgressive inheritance, meaning that hybrid gene expression was significantly higher (overdominant) or lower (underdominant) than both parental species (Figures S7–S9).

Transgressive gene expression in hybrids can result from several types of molecular interactions. Hybrid gene misregulation refers to transgressive expression that is caused by divergence in *cis*- and *trans*-regulatory machinery between parental species (Mack & Nachman, 2017). Hybrid gene misexpression is a more general term that describes transgressive expression caused by gene misregulation in addition to transgressive expression resulting from aberrant development in hybrids that leads to differences in developmental rate or the relative abundance of specific cell types (Landry et al., 2005; Mack & Nachman, 2017; Swain Lenz, Riles, & Fay, 2014). Since we did not perform experiments testing for differences in developmental rate or cellular composition between parents and hybrids, we describe transgressive expression observed in crosses between San Salvador Island species as hybrid gene misexpression. However, it is important to note that we see no evidence for differences in hatch time between these very closely related species, nor between hybrids and parental species (Lencer, Riccio, & McCune, 2016; McGirr & Martin, 2017, 2019). Furthermore, crosses between all San Salvador Island species result in fertile F1 and later generation hybrids. This contrasts with observations in other systems examining gene regulatory evolution between distantly related species pairs showing strong intrinsic postzygotic isolation (Coolon et al., 2014; Mack et al., 2016; Ranz et al., 2004). In these systems, differences in gene expression between hybrids and parents may be due to aberrant development of reproductive tissues (hybrid dysfunction).

2.6 | Parallel changes in gene expression in specialists

We looked at the intersection of genes differentially expressed between generalists versus molluscivores and generalists versus scale-eaters to determine whether specialists showed parallel changes in expression relative to generalists (McGirr & Martin, 2018). We also examined the direction of expression divergence for each gene to evaluate the significance of parallel expression evolution (Figure 4e). Specifically, we wanted to know whether the fold change in expression for genes tended to show the same sign in both specialists relative to generalists (either upregulated in both specialists relative to generalists or downregulated in both specialists). Such a pattern would indicate parallel selection on gene expression. Alternatively, under a neutral model of gene expression evolution, half of the genes differentially expressed between generalists versus molluscivores and generalists versus scale-eaters would be expected to show fold changes with the same sign and half would show fold changes with opposite signs (Figure 4e).

We wanted to determine whether significant parallelism at the level of gene expression in specialists was mirrored by parallel regulatory mechanisms. We predicted that genes showing parallel changes in specialists would show conserved expression levels in specialist hybrids if they were controlled by the same (or compatible) regulatory mechanisms, but would be misexpressed in specialist hybrids if expression was controlled by different and incompatible regulatory mechanisms. We identified genes showing conserved levels of expression in specialist hybrids (no significant difference in expression between purebred specialist F1s and specialist hybrid F1s) and genes showing misexpression in specialist hybrids. We also identified genes showing extreme Caribbean-wide misexpression in specialists. These genes were differentially expressed in specialist hybrids relative to all other samples in our data set from across the Caribbean (North Carolina to New Providence Island, Bahamas).

2.7 | Allele-specific expression

Our genomic data set included every parent used to generate F1 hybrids between populations ($n = 15$). To categorize mechanisms of regulatory divergence between two populations, we used custom R and Python scripts (github.com/joemcgirr/fishfASE) to identify SNPs that were alternatively homozygous in breeding pairs and heterozygous in their F1 offspring. We counted reads across heterozygous sites using ASEReadCounter and matched read counts to maternal and paternal alleles. We identified significant ASE using a beta-binomial test comparing the maternal and paternal counts at each gene transcript with the R package MBASED (Mayba et al., 2014).

In order to determine regulatory mechanisms controlling expression divergence between parental species, a transcript had to be included in differential expression analyses and ASE analyses. We were able to classify regulatory categories for more transcripts if breeding pairs were more genetically divergent because we could analyse more heterozygous sites in their hybrids (mean number of informative transcripts across crosses = 1,914; range = 812–3,543). For each hybrid sample and each transcript amenable to both types of analyses, we calculated H—the ratio of maternal allele counts compared to the number of paternal

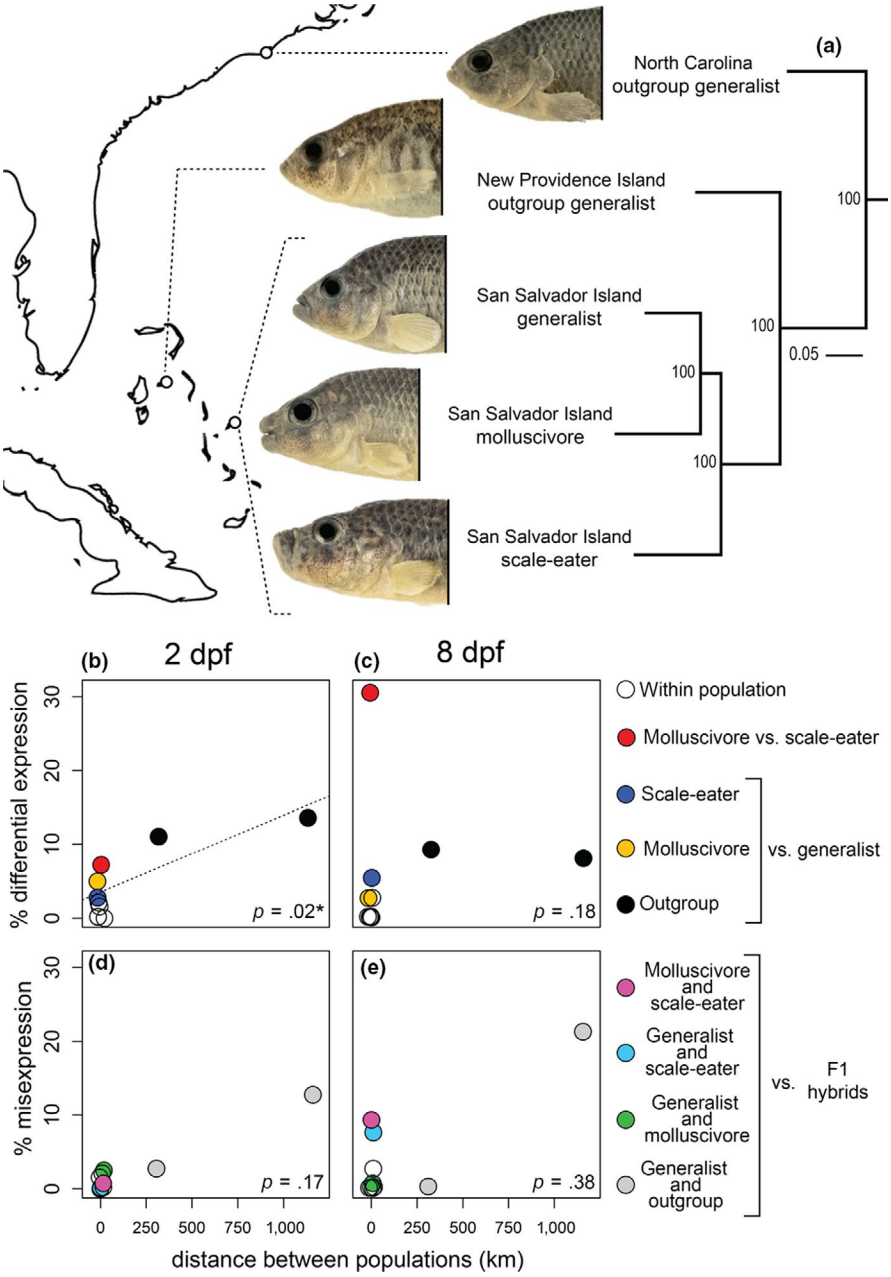


FIGURE 1 Caribbean-wide patterns of gene expression and misexpression across sympatric and allopatric populations of *Cyprinodon* pupfishes. (a) Maximum-likelihood tree estimated from 1.7 million SNPs showing phylogenetic relationships among generalist populations and specialist species (100% bootstrap support indicated at nodes). (b) Geographic distance separating populations was associated with differential gene expression levels in embryos at 2 days postfertilization (2 dpf; phylogenetic least squares $p = .02$, dotted regression line). (c) In whole larvae at 8 dpf, differential expression was not associated with geographic distance (PGLS; $p = .18$) and was higher between sympatric specialists (red) than between allopatric generalists separated by 300 and 1,000 km (black). (d and e) Hybrid gene misexpression for sympatric crosses at 2 and 8 dpf. Geographic distance was not associated with hybrid misexpression at either developmental stage (PGLS; 2 dpf $p = .17$; 8 dpf $p = .38$). Percentages in B-E were measured using Crescent Pond crosses [Colour figure can be viewed at wileyonlinelibrary.com]

allele counts in F1 hybrids, and P—the ratio of normalized read counts in purebred F1 offspring from the maternal population compared to read counts in purebred F1 offspring from the paternal population. We performed a Fisher's exact test using H and P to determine whether there was a significant *trans*-contribution to expression divergence, testing the null hypothesis that the ratio of read counts in the parental populations was equal to the ratio of parental allele counts in hybrids (Goncalves et al., 2012; Mack et al., 2016; McManus et al., 2010; Wittkopp, Haerum, & Clark, 2004). We classified expression divergence due to *cis*-regulation if a transcript showed significant ASE, significant differential expression between parental populations of purebred F1 offspring and no significant *trans*-contribution. We identified expression divergence due to *trans*-regulation if transcripts did not show ASE, were differentially expressed between parental populations of purebred F1 offspring and showed significant *trans*-contribution (Figures S10–S12).

3 | RESULTS

3.1 | Trophic specialization, not geographic distance, drives major changes in gene expression and hybrid gene misexpression

Gene expression divergence is expected to increase with increasing phylogenetic distance between closely related species and is expected to increase more rapidly when directional selection on gene expression is strong (Whitehead & Crawford, 2006). Since allopatric generalist populations are adapted to similar ecological niches

and sympatric specialist species are adapted to divergent niches (Martin, 2016b), stronger selection on gene expression in the specialist species may contribute to faster gene expression divergence between sympatric species than between allopatric generalists. However, gene expression levels among species cannot be considered to be independent and identically distributed random variables. Thus, we predicted that gene expression divergence should be higher between sympatric specialists than between allopatric generalists after controlling for genetic divergence among all populations. To test this, we determined whether isolation by distance explained patterns of gene expression divergence while controlling for phylogenetic relatedness using a maximum-likelihood tree estimated with RAxML from 1.7 million SNPs (Figure 1; Figure S5).

Overall, genetic divergence increased with geographic distance between allopatric generalist populations and was lowest between sympatric populations (Table S3; genome-wide mean *Fst* measured across 13.8 million SNPs: San Salvador generalists versus North Carolina = 0.217; versus New Providence = 0.155; versus scale-eaters = 0.106; versus molluscivores = 0.056). Geographic distance among populations was a significant predictor of the proportion of differential gene expression between populations at two days postfertilization (2 dpf) (Figure 1b; phylogenetic generalized least squares [PGLS]; $p = .02$). This is consistent with a model of gene expression evolution governed largely by stabilizing selection and drift (Whitehead & Crawford, 2006). However, at eight days postfertilization (8 dpf), when craniofacial structures of the skull begin to ossify (Lencer & McCune, 2018), geographic distance was no longer associated with differential expression (Figure 1c; PGLS; $p = .18$), which was higher between sympatric trophic specialist species on San Salvador Island than between generalist populations spanning 1,000 km across the Caribbean. Thus, differential gene expression at 8 dpf was much higher than expected due to isolation by distance, suggesting that strong directional selection on gene expression was important during ecological divergence in sympatry.

Similar to expectations for gene expression divergence between species, the extent of F1 hybrid gene misexpression likely depends on genetic divergence between parental species (Coolon et al., 2014). Thus, we predicted to find higher levels of gene misexpression in specialist F1 hybrids than allopatric generalist F1 hybrids after accounting for phylogenetic relationships. Consistent with this prediction, geographic distance between parental populations was not associated with gene misexpression in F1 hybrids at either developmental stage (Figure 1d,e; PGLS; 2 dpf $p = .17$; 8 dpf $p = .38$). This suggests that the number of genes misexpressed in Crescent Pond molluscivore × scale-eater hybrids (9.3% of genes at 8 dpf) and Crescent Pond generalist × scale-eater hybrids (7.6% of genes at 8 dpf) was higher than expected given the amount of genetic divergence observed between populations. Indeed, this amount of gene misexpression is comparable to species pairs with much greater divergence times (Coolon et al., 2014; Mack et al., 2016).

In order to further examine the relationship between hybrid misexpression and genetic divergence between parental populations, we measured *Fst* across all genes expressed in each F1 hybrid cross. For each gene region, we identified the most divergent SNP (showing the highest measure of *Fst*) across all SNPs within the reading frame of the gene and within 10 kb of the first or last exon of the gene. We also measured the mean *Fst* across all SNPs in the gene region. For the majority of all crosses at the 8 dpf stage, including all five crosses involving scale-eaters, we found a significant positive relationship between the most divergent SNP within a gene region and its fold change in expression between F1 hybrids versus parental populations (linear regression, $p < .05$: Figure 2). We also found the same relationship in the majority of 2 dpf crosses (linear regression, $p < .05$: Figure S13). We did not find a significant relationship when considering mean *Fst* at either developmental stage (linear regression, $p > .05$; Figures S14 and S15); however, this approach is less sensitive to the effects of highly differentiated SNPs when they are linked to many variants segregating at intermediate frequencies. Together, these results suggest that positive selection acting on gene expression in San Salvador Island species has contributed to patterns of gene misexpression in their F1 hybrids.

3.2 | Genes differentially expressed between species are misexpressed in F1 hybrids

Hybrid gene misexpression can result from stabilizing selection or directional selection (including divergent selection) on gene expression (Landry et al., 2007; Mack & Nachman, 2017; Signor & Nuzhdin, 2018). When stabilizing selection favours an optimal level of gene expression, hybrid gene misexpression is expected to result from epistasis between *cis*- and *trans*-compensatory variants that have accumulated between diverging lineages. In order to determine regulatory mechanisms underlying hybrid gene misexpression, we measured allele-specific expression across genes containing heterozygous sites in F1 hybrids that were homozygous in their parents. Out of 3,669 misexpressed genes amenable to this analysis, 819 (22.3%) showed allele-specific expression and were not differentially expressed between parental populations. This expression pattern is consistent with compensatory regulation underlying misexpression, indicating stabilizing selection acting on gene expression (Figures S10–S12, Table S4).

Alternatively, if directional selection on regulatory variants contributed to hybrid gene misexpression, we would expect the same genes showing differential expression between species to show misexpression in F1 hybrids. Thus, we intersected genes that were differentially expressed between San Salvador Island species with genes showing misexpression in F1 hybrids to identify two types of expression patterns consistent with directional selection on regulatory genetic variants causing adaptive expression divergence between species.

First, we found 716 genes that showed differential expression between San Salvador Island species that were also misexpressed in their F1 hybrids (Figure 3, Table S5). The majority of these genes showed at least a twofold difference in expression in each comparison (Table S6). We found that 69.8% of the 716 genes were

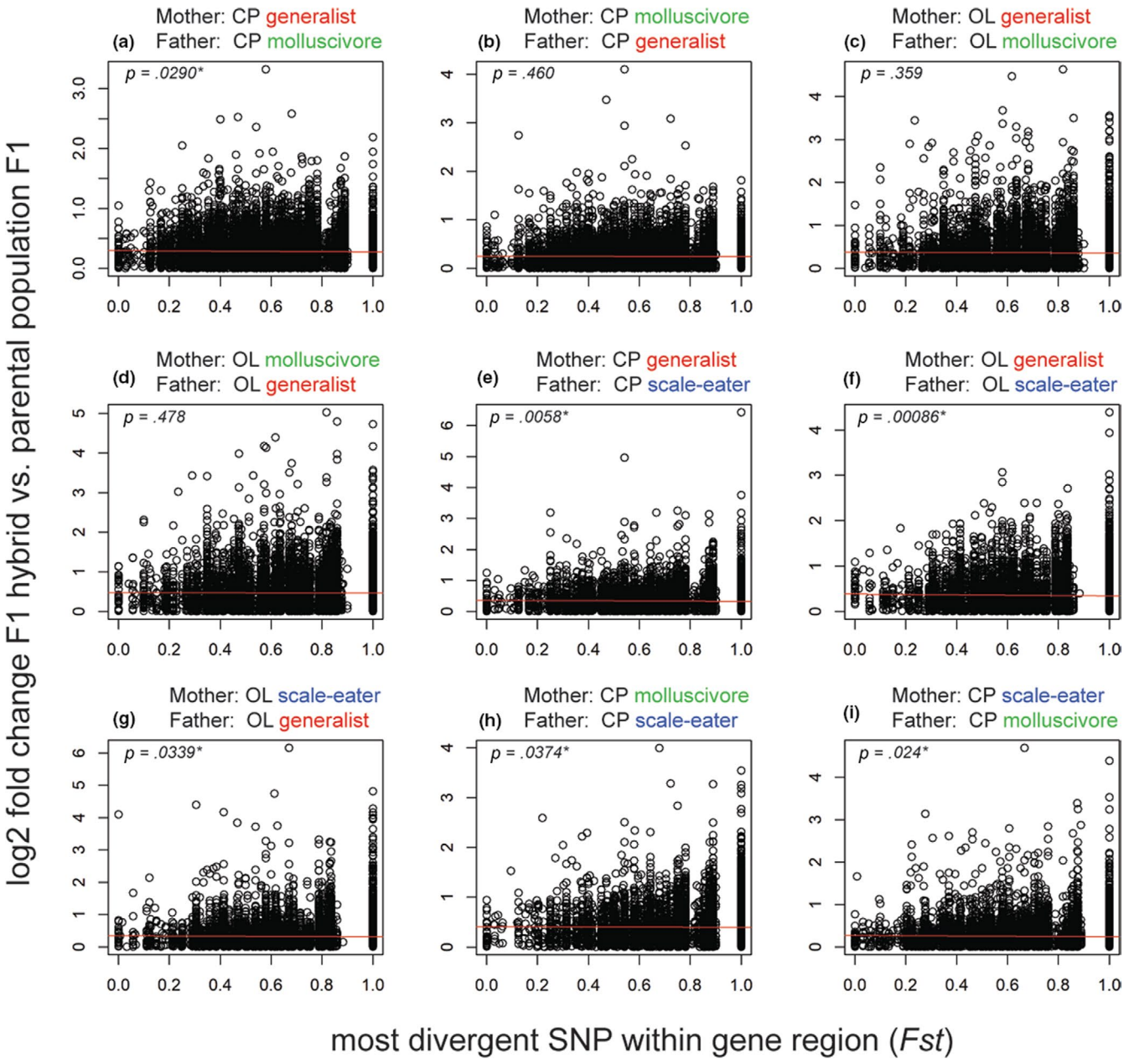


FIGURE 2 Genes show larger fold changes in expression between F1 hybrids versus. parental populations when they are near SNPs showing increased *Fst*. For the majority of crosses (a, e–i), including all crosses involving scale-eaters (e–i), we found a significant positive relationship between the most divergent SNP measured for a gene and its fold change in expression between F1 hybrids versus parental populations at 8 dpf (linear regression, $p < .05$). Absolute values shown for log2 fold changes in expression. For each gene, we identified the highest measure of *Fst* across all SNPs within the reading frame of the gene and within 10 kb of the first or last exon of the gene. CP, Crescent Pond; OL, Osprey Lake [Colour figure can be viewed at wileyonlinelibrary.com]

only misexpressed at 8 dpf in comparisons involving scale-eaters (Figure 3a–h). Additionally, nearly all of the 716 genes (712; 99.4%) were misexpressed in only one lake population. This may suggest that incompatible alleles contributing to misexpression are segregating within species and between lake populations (Corbett-detig et al., 2013). However, we also found four genes that showed differential expression between species and misexpression in their hybrids in both lake comparisons (*trim47*, *krt13*, *s100a1*, *elovl7*; Table S7).

Second, we identified genes showing parallel expression divergence in both specialist species relative to generalists that were misexpressed in specialist F1 hybrids (Figure 4). This pattern likely results from parallel expression in molluscivores and scale-eaters controlled by different genetic mechanisms (McGirr & Martin, 2018). Significantly more genes showed differential expression in both specialist comparisons than expected by chance (Figure 4a–d; Fisher's exact test, $p < 2.7 \times 10^{-5}$). Of these, 96.6% (1,206) showed the same direction of expression in specialists relative to generalists. This was much more than expected under a neutral model of gene expression evolution, where a gene would be equally likely to show expression divergence in opposite directions in specialists (Figure 4e,f;

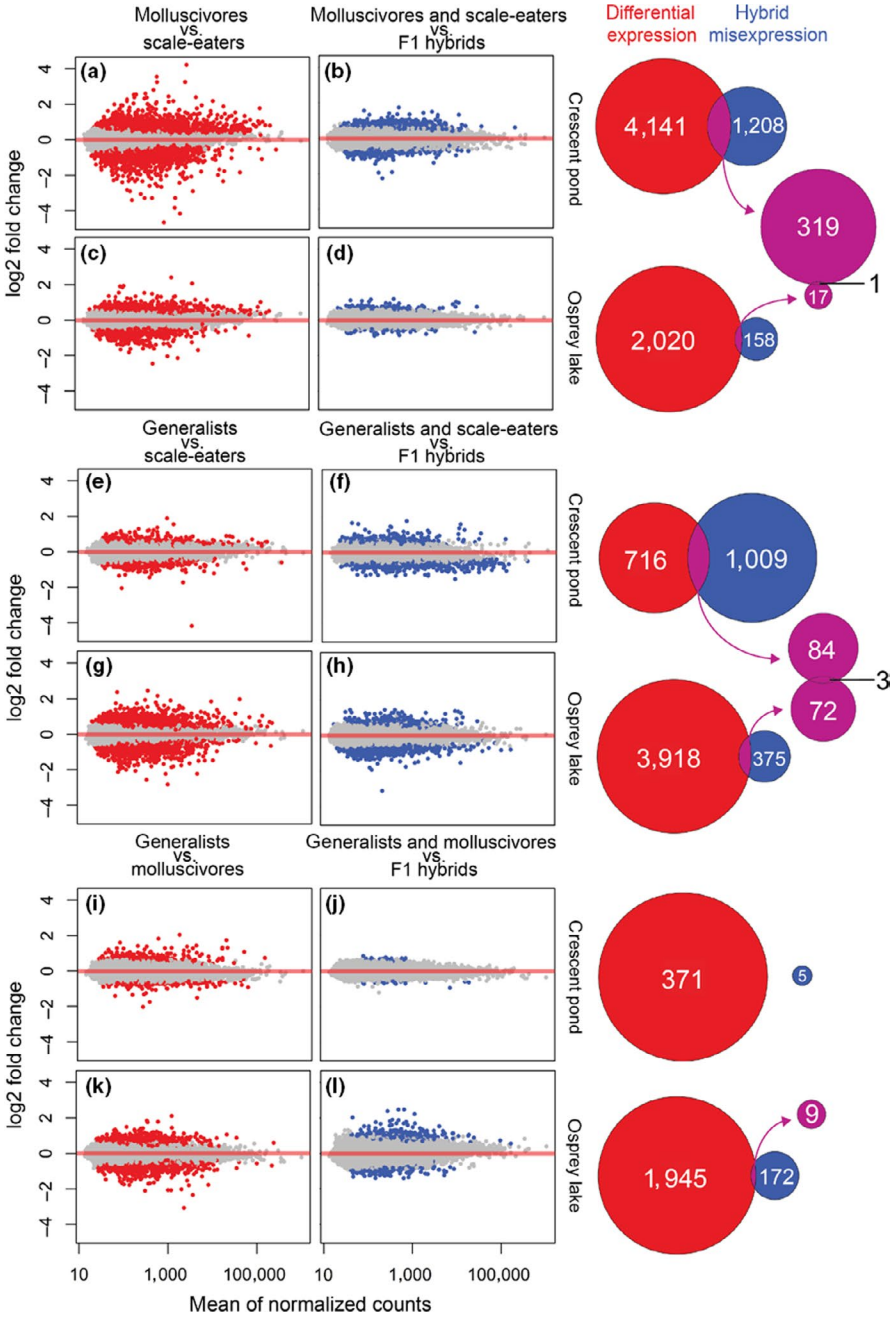


FIGURE 3 Genes differentially expressed between species are misexpressed in their F1 hybrids at 8 days postfertilization. Genes differentially expressed between San Salvador species from Crescent Pond and Osprey Lake are shown in red for molluscivore × scale-eater crosses (a–d), generalist × scale-eater crosses (e–h) and generalist × molluscivore crosses (i–l). Genes misexpressed in F1 hybrids are shown in blue. In comparisons involving reciprocal crosses (d, j, and l), we only show genes misexpressed in a single cross direction. A total of 716 genes (purple) were differentially expressed between species and also misexpressed in their F1 hybrids. Purple Venn diagrams show overlap between lake population comparisons; four genes showed differential expression and misexpression in both lake comparisons [Colour figure can be viewed at wileyonlinelibrary.com]

binomial test, $p < 1.0 \times 10^{-16}$). 45 of the 1,206 genes showing parallel expression divergence in specialists also showed misexpression in specialist F1 hybrids (Figure 4f). Eight of these genes were severely misexpressed to the extent that they were differentially expressed in hybrids relative to all other populations in our data set. For example, *sypl1* showed significantly higher expression in 8 dpf Crescent Pond molluscivore × scale-eater F1 hybrids than all other crosses spanning 1,000 km from San Salvador Island, Bahamas to North Carolina, USA ($p = 2.35 \times 10^{-4}$; Figure 4g). Overexpression of this gene is associated with epithelial–mesenchymal transition, an important process during cranial neural crest cell migration (Chen, Wu, Li, Wang, & Zhang, 2017; Kang & Svoboda, 2005). Similarly, *scn4a* showed significantly lower expression in 8 dpf Crescent Pond specialist F1 hybrids than all other crosses ($p = 5.49 \times 10^{-4}$; Figure 4h). Mutations in this gene are known to cause paramyotonia congenita, a disorder causing weakness and stiffness of craniofacial skeletal muscles (Huang, Zhang, Chang, & Guo, 2019).

3.3 | Misexpressed genes under selection influence adaptive ecological traits in trophic specialists

If hybrid gene misexpression was influenced by adaptive gene regulatory divergence between species, we predicted that these genes should show genetic signatures of selection and be associated with adaptive phenotypes. Out of 750 total candidate genes identified above as differentially expressed between populations and misexpressed in F1 hybrids, 125 (17%) were within 20 kb of SNPs that

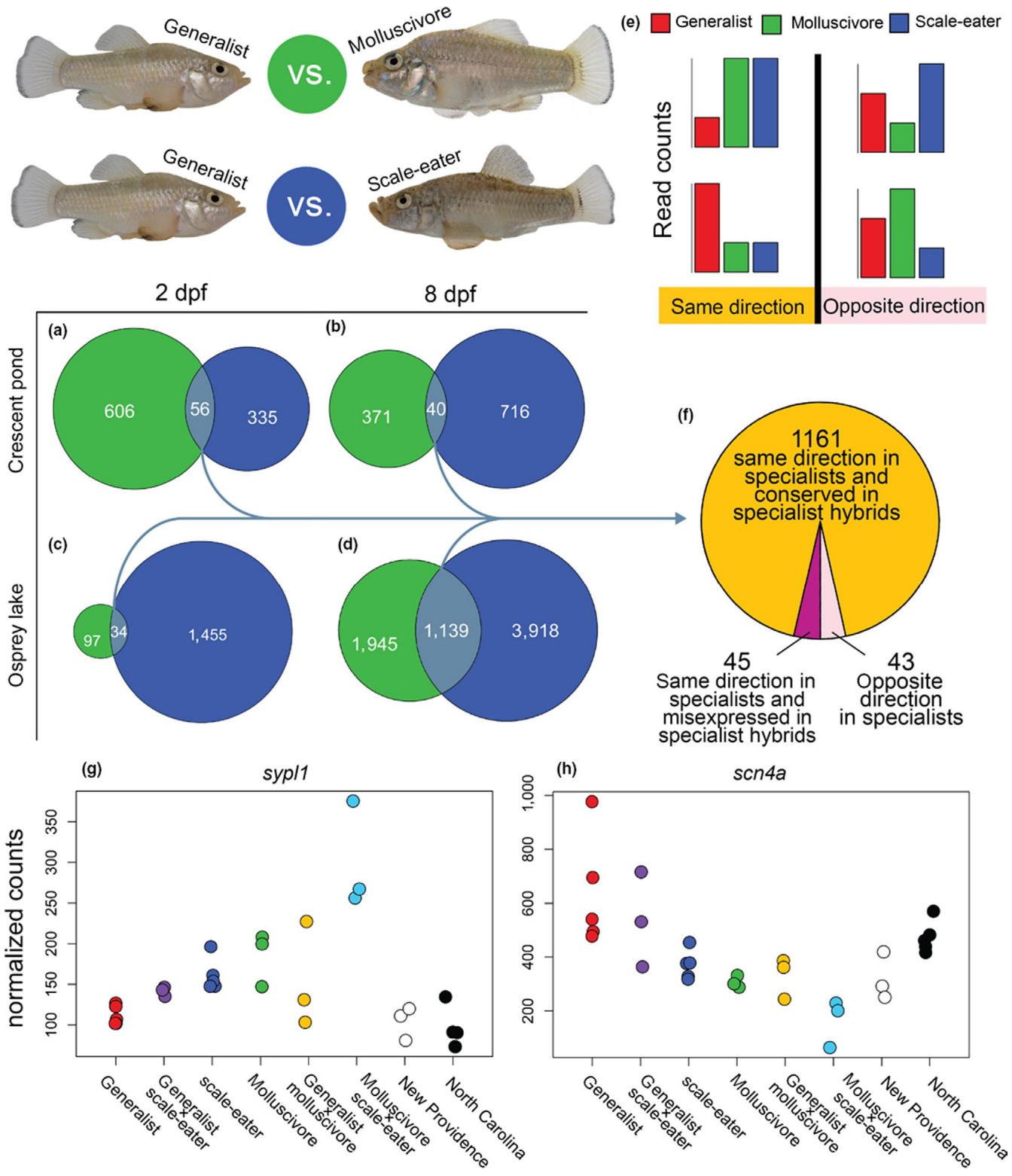


FIGURE 4 Genes showing parallel expression divergence in specialists are misexpressed in specialist hybrids. Genes differentially expressed between generalists and molluscivores (green) were compared to the set of genes differentially expressed between generalists and scale-eaters (dark blue). (a–d) Significantly more genes showed differential expression in both specialist comparisons (light blue) than expected by chance in both lakes at both developmental stages (Fisher's exact test, $p < 2.7 \times 10^{-5}$). (e) Parallel selection on gene expression in the specialists would cause more genes to show the same direction of expression divergence relative to generalists (either upregulated in both specialists relative to generalists or downregulated in both specialists) than genes showing the opposite direction of expression. (f) 96.6% of genes showed the same direction of expression in specialists, suggesting significant parallel expression divergence in specialists (binomial exact test; $p < 1.0 \times 10^{-16}$). Consistent with incompatible regulatory mechanisms underlying parallel expression in specialists, 45 of these genes were misexpressed in specialist F1 hybrids, including (g) *sypl1* and (h) *scn4a* which also showed expression levels outside the range of all other Caribbean populations examined [Colour figure can be viewed at wileyonlinelibrary.com]

were fixed between populations (*Fst* = 1) and within 20 kb windows showing high absolute genetic divergence between populations (*Dxy* ≥ genome-wide 90th percentile; range: 0.0031–0.0075; Table S3). Using this conservative threshold to define outlier differentiated regions, the overlap of 125 genes in differentiated regions was not more than expected by chance considering the overlap between noncandidate genes in differentiated regions (Fisher's exact test $p > .05$; Table S8). However, the positive relationship between highly differentiated SNPs and expression divergence in F1 hybrids predicted that genes near fixed variants were likely to show strong misexpression (Figure 2). Interestingly, these 125 candidate genes were significantly enriched for functional categories highly relevant to divergent specialist phenotypes, including head development, brain development, muscle development and cellular response to nitrogen (FDR = 0.05; Figure 5a, Table S9). We refer to these 125 genes as ecological DMI candidate genes because (a) they showed high genetic differentiation between species, (b) were enriched for developmental functions related to divergent adaptive traits (c) and showed expression patterns consistent with incompatible interactions between divergent regulatory alleles contributing to hybrid gene misexpression.

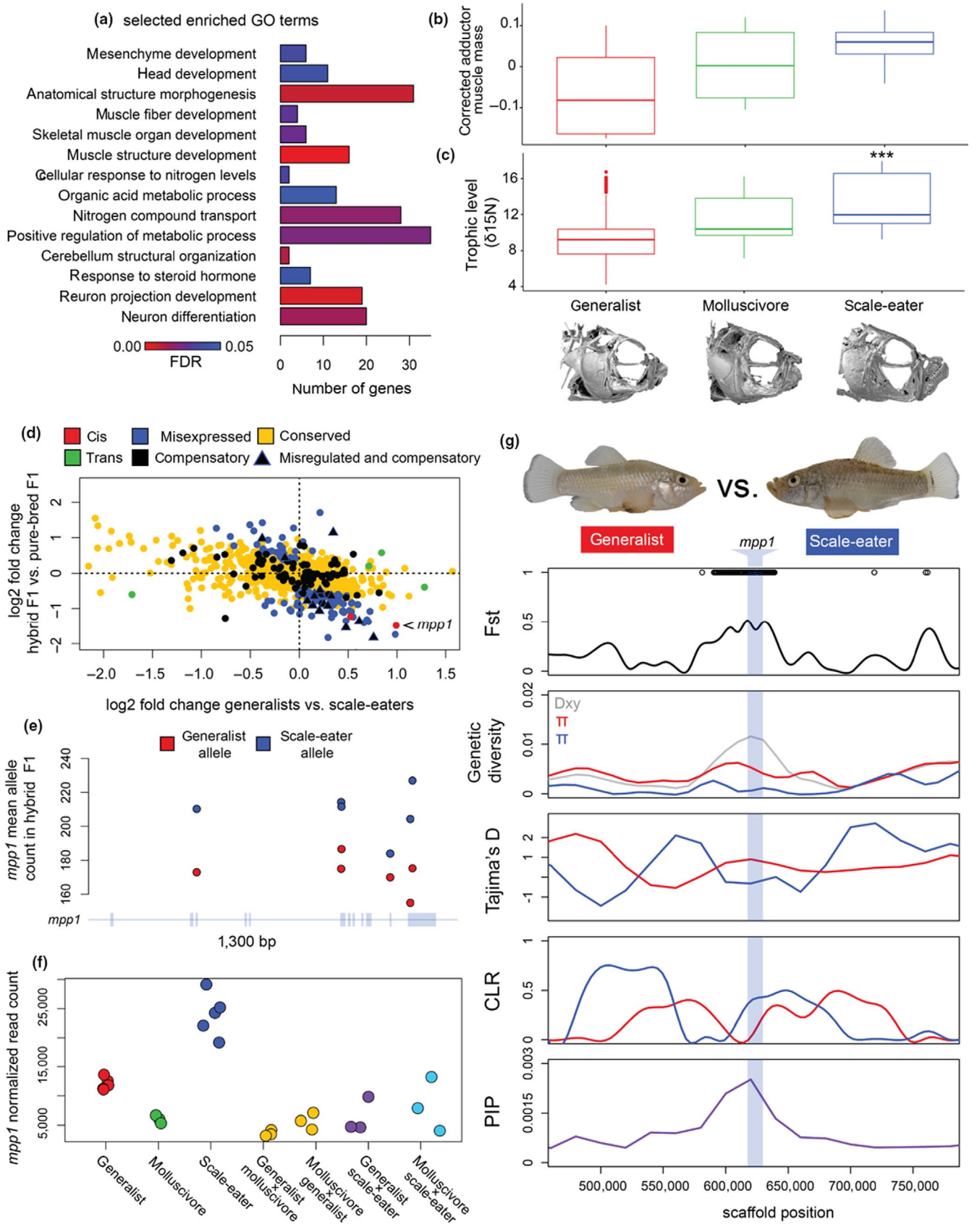
Twenty-six (20.8%) of these ecological DMI candidate genes showed strong evidence of a hard selective sweep in specialists (negative Tajima's D < genome-wide 10th percentile; range: −1.62 to −0.77; SweeD composite likelihood ratio >90th percentile by scaffold; Tables S2 and S10), and 16 of these showed at least a twofold expression difference in F1 hybrids compared to purebred F1. Several ecological DMI candidate genes have known functions that are compelling targets for divergent ecological selection. For example, the autophagy-related gene *map1lc3c* has been shown to influence growth when cells are nitrogen deprived (Otto, Wu, Kazgan, Anderson, & Kessin, 2004; Stadel et al., 2015). Given that specialists occupy higher trophic levels than generalists, as shown by stable isotope ratios (δ15N; Figure 5b), expression changes in this gene may be important adaptations to nitrogen-rich diets. Similarly, expression changes in the ten genes annotated for effects on brain development may influence divergent behavioural adaptations associated with trophic specialists, including significantly increased aggression (St John, McGirr, & Martin, 2019) and female mate preferences (West & Kodric-Brown, 2015).

Using a genome-wide association mapping method that accounts for genetic structure among populations (Zhou et al., 2013), we found that nine of the 125 genes in differentiated regions were significantly associated with oral jaw size—the most rapidly diversifying skeletal trait in this radiation (GEMMA PIP > 99th percentile; Table S11; Figure S6). For example, we found that *mpp1* was near 170 SNPs fixed between Crescent Pond generalists and scale-eaters, showed evidence of a hard selective sweep in both populations and was differentially expressed due to *cis*-regulatory mechanisms (Figure 5f–i). F1 hybrids showed a threefold decrease in expression of *mpp1* ($p = .001$; Figure 5f). Knockouts of this gene were recently shown to cause severe craniofacial defects in humans and mice (Fritz, Johnston, & Chishti, 2014). The other eight genes significantly associated with jaw size have not been previously shown to influence cranial phenotypes, but some have known functions in cell types relevant to craniofacial development (Table S11). For example, the gene *sema6c*, which shows strong signs of selection in both scale-eaters and molluscivores (Figure S16), is known to be expressed at neuromuscular junctions and is important for neuron growth and development within skeletal muscle (Svensson, Libelius, & Tågerud, 2008). Expression changes in this gene may influence the development of jaw closing muscles (adductor mandibulae), which tend to be larger in specialists relative to generalists (Figure 5c). Overall, we found candidate regulatory variants under selection that likely contribute to hybrid gene misexpression and demonstrate that genes near these variants are strikingly enriched for developmental functions related to divergent adaptive traits.

4 | DISCUSSION

By combining whole-genome sequencing with RNA sequencing across multiple developmental stages within a system of recently diverged trophic specialists and their F1 hybrids, we provide a genome-wide view of how ecological selection can influence gene misexpression in hybrids. Unlike other studies that examined hybrid gene misexpression between distantly related species pairs exhibiting strong intrinsic reproductive isolation (Kerwin & Sweigart, 2020; Landry et al., 2007; Mack & Nachman, 2017), we show that misexpression can evolve between recently diverged species that coexist in sympatry and still produce fertile hybrids. Our results are consistent with negative epistatic interactions between alleles from different parental genomes affecting 750 genes (3% of the transcriptome) that show differential expression between species and misexpression in F1 hybrids. 125 of these genes were in highly differentiated regions of the genome containing SNPs fixed between species which

FIGURE 5 Ecological divergence causes hybrid gene misexpression. (a) Fourteen selected gene ontology (GO) terms relevant to trophic specialization were significantly enriched for the set of 125 genes in highly differentiated genomic regions that showed differential expression between species and misexpression in F1 hybrids. Consistent with muscle development and nitrogen metabolism enrichment, (b) adductor mandibulae muscle mass tends to be larger in specialists and (c) stable nitrogen isotope ratios (δ15N) are significantly higher in scale-eaters, indicating that they occupy a higher trophic level (Tukey post hoc test: $p < .001^{***}$). (d) The gene *mpp1* is controlled by *cis*-regulatory divergence as shown by (e) allele-specific expression in F1 hybrids and (f) differential expression between Crescent Pond generalists versus scale-eaters and misexpression in their F1 hybrids. (g) The gene *mpp1* (light blue band) is near 170 SNPs fixed between Crescent Pond generalists versus scale-eaters (black points), shows high absolute divergence between species (*Dxy*), low within-species diversity (π) and signatures of a hard selective sweep (Tajima's D and SweeD composite likelihood ratio [CLR]), and is significantly associated with oral jaw length (PIP; GEMMA genome-wide association mapping) [Colour figure can be viewed at wileyonlinelibrary.com]



were enriched for developmental processes relevant to trophic specialization. Given that gene expression levels experience strong stabilizing selection in many organisms (Bedford & Hartl, 2009; Mack & Nachman, 2017; Signor & Nuzhdin, 2018), we speculate that misexpression of these candidate genes in F1 and later generation hybrids may disrupt developmental processes that affect the function

of adaptive traits and contribute to reproductive isolation between these nascent species.

The negative fitness consequences associated with hybrid gene misexpression have been described in several systems (Landry et al., 2007; Mack et al., 2016; Maheshwari & Barbash, 2012; Malone & Michalak, 2008; Ortíz-Barrientos et al., 2007), but most of this research has focused on genes associated with sterility and inviability between highly divergent species (but see (Renaut et al., 2009)). It is clear that these strong intrinsic postzygotic isolating barriers evolve more slowly than premating barriers (Coyne & Orr, 1989, 2004; Turissini, McGirr, Patel, David, & Matute, 2018); however, hybrid gene misexpression may also have nonlethal effects on fitness and performance that could evolve before or alongside premating isolating mechanisms. Additionally, if genes that are differentially expressed between species in developing tissues are important for adaptive trait divergence, then misexpression of those genes could contribute to abnormal phenotypes that are ecologically maladaptive (Arnegard et al., 2014; Kulmuni & Westram, 2017; Renaut et al., 2009). Consistent with this hypothesis, we previously found that hybrid gene misexpression was pervasive in tissues that develop into divergent trophic morphologies by measuring expression in craniofacial tissues, which were dissected from generalist × molluscivore F1 hybrids at an early developmental stage (McGirr & Martin, 2019).

It is difficult to demonstrate a causative link between hybrid gene misexpression and hybrid fitness without functional validation of specific genes or the use of recombinant mapping populations. Thus, it is possible that observed patterns of misexpression have little influence on hybrid fitness, or even provide a benefit similar to transgressive phenotypes observed in hybrid lineages (Rieseberg, Archer, & Wayne, 1999). However, multiple independent lines of evidence suggest that transgressive gene expression does not increase fitness in this system. First, hybrids among San Salvador Island species suffer reduced survival and growth rate in their natural field environments. Fitness measurements in the wild found that F2 hybrids showing more transgressive phenotypes exhibited the lowest survival and growth rate in field enclosures across multiple lakes and multiple independent field experiments on San Salvador Island (Martin, 2016a; Martin & Gould, 2019; Martin & Wainwright, 2013b). Second, hybrids suffer reduced performance in the laboratory. Recent feeding kinematic experiments showed that generalist × scale-eater F1 hybrids exhibited nonadditive and impaired feeding performance on scales (St. John et al., 2020).

If divergent ecological selection on adaptive traits also causes gene misexpression and subsequently reduced performance and survival of hybrids in the wild, then these ecological DMIs may promote rapid speciation, analogous to the mechanism of magic traits (Servedio, Doorn, Kopp, Frame, & Nosil, 2011). For example, whereas magic traits contribute to reproductive isolation through assortative mating as a byproduct of divergent ecological selection, these ecological DMIs contribute to isolation through gene misexpression and reduced hybrid fitness (Kulmuni & Westram, 2017). Thus, our results support a mechanism for divergent ecological selection to generate reproductive isolation as a byproduct since many adaptive traits are expected to evolve by divergent gene regulation that may come into conflict in a hybrid genetic background (Kulmuni & Westram, 2017; Pavey et al., 2010).

Mathematical models and simulations suggest that genetic incompatibilities evolve most rapidly under directional selection (Johnson & Porter, 2000; Tulchinsky et al., 2014) and evolve more slowly under stabilizing selection when compensatory *cis*- and *trans*-variants have opposing effects on expression levels (Tulchinsky et al., 2014). We see evidence for both types of selection driving misexpression. Out of the genes showing hybrid misexpression that contained heterozygous variation, 819 showed expression patterns consistent with compensatory regulation, a signature of stabilizing selection (Table S4). Alternatively, 750 misexpressed genes were differentially expressed between species, a signature of directional selection. Of these genes, 125 were in highly differentiated genomic regions containing SNPs fixed between populations, and 26 genes showed strong evidence of hard selective sweeps. (Table S10). Importantly, even more genes may have experienced soft sweeps that were not detected by our methods.

Although scale-eaters from Crescent Pond and Osprey Lake form a monophyletic group (Figure S5), we found little overlap in misexpressed genes between lakes (Figure 3). We also found a large difference in the number of genes differentially expressed between generalists and each of the specialists for each lake comparison (Figure 4a–d). This may result from selection on Caribbean-wide standing genetic variation that has similar effects on expression, as we showed previously (McGirr & Martin, 2018), and could reflect polymorphic incompatibilities segregating within species (Corbett-detig et al., 2013). We also see distinct intraspecific differences between lake populations of trophic specialists in pigmentation, maxillary protrusion and other traits (Martin & Feinstein, 2014), consistent with divergent regulatory variation underlying these adaptive phenotypes.

Identifying genetic variation that contributes to adaptive variation and studying its effect on reproductive isolation is important to understand the sequence of molecular changes leading to ecological speciation. We show that ecologically relevant genes near differentiated genetic regions between sympatric species are under selection and misexpressed in F1 hybrids. Overall, our results are consistent with previous observations that hybrid incompatibility alleles are often segregating within populations (Corbett-detig et al., 2013; Cutter, 2012; Larson et al., 2018; Reed & Markow, 2004) and that hundreds of genetic incompatibilities can contribute to reproductive isolation between species at the earliest stages of divergence (Schumer et al., 2014). We extend this emerging consensus by showing that gene misexpression may result as a byproduct of divergent ecological selection on a wide range of adaptive traits.

ACKNOWLEDGEMENTS

This study was funded by the University of North Carolina at Chapel Hill, the Miller Institute for Basic Research in Science, NSF CAREER Award 1938571 and NIH/NIDCR R01 DE027052 to CHM.

Research was also supported by a fellowship from the Triangle Center for Evolutionary Medicine and an SSE Rosemary Grant Travel Award to JAM. We thank Daniel Matute, Aaron Comeault, Chris Willett and Jennifer Coughlan for illuminating comments on the manuscript; Emilie Richards, Michelle St. John, Bryan Reatini, and Sara Suzuki for valuable discussion; The Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, supported by NIH S10 OD018174 Instrumentation Grant, for performing RNA library prep and Illumina sequencing; the Gerace Research Centre for logistics; and the Bahamian government BEST Commission for permission to conduct this research.

CONFLICT OF INTEREST

We declare no competing interests.

DATA AVAILABILITY STATEMENT

All transcriptomic raw sequence reads are available as zipped fastq files on the NCBI BioProject database. Accession: PRJNA391309. Title: Craniofacial divergence in Caribbean Pupfishes. Metadata for samples are matched to labels shown in Table S1. All R and Python scripts used for pipelines are available on Github (github.com/joemcgirr/fishfASE).

AUTHOR CONTRIBUTIONS

JAM extracted RNA samples, conducted all bioinformatic and population genetic analyses, and wrote the manuscript. JAM and CHM contributed to the conception and development of the project and revised the manuscript.

ORCID

Joseph A. McGirr https://orcid.org/0000-0003-3942-4372
Christopher H. Martin https://orcid.org/0000-0001-7989-9124

REFERENCES

Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., & Tabin, C. J. (2011). Variation of beaks in Darwin's finches. *Science*, *1462*(2004), 1462–1466. https://doi.org/10.1126/science.1098095

Arnegard, M. E., McGee, M. D., Matthews, B., Marchinko, K. B., Conte, G. L., Kabir, S., … Schluter, D. (2014). Genetics of ecological divergence during speciation. *Nature*, *511*(7509), 307–311. https://doi.org/10.1038/nature13301

Barreto, F. S., Pereira, R. J., & Burton, R. S. (2015). Hybrid dysfunction and physiological compensation in gene expression. *Molecular Biology and Evolution*, *32*(3), 613–622. https://doi.org/10.1093/molbev/msu321

Bedford, T., & Hartl, D. L. (2009). Optimization of gene expression by natural selection. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(4), 1133–1138. https://doi.org/10.1073/pnas.0812009106

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B*, *57*(1), 289–300.

Chen, D. H., Wu, Q. W., Li, X. D., Wang, S. J., & Zhang, Z. M. (2017). SYPL1 overexpression predicts poor prognosis of hepatocellular carcinoma and associates with epithelial-mesenchymal transition. *Oncology Reports*, *38*(3), 1533–1542. https://doi.org/10.3892/or.2017.5843

Coolon, J. D., Mcmanus, C. J., Stevenson, K. R., Coolon, J. D., Mcmanus, C. J., Stevenson, K. R., … Wittkopp, P. J. (2014). Tempo and mode of regulatory evolution in *Drosophila*. *Genome Research*, *24*(5), 797–808. https://doi.org/10.1101/gr.163014.113

Corbett-detig, R. B., Zhou, J., Clark, A. G., Hartl, D. L., & Ayroles, J. F. (2013). Genetic incompatibilities are widespread within species. *Nature*, *504*(7478), 135–137. https://doi.org/10.1038/nature12678

Coyne, J. A., & Orr, H. A. (1989). Patterns of speciation in *Drosophila*. *Evolution*, *43*(2), 362–381. https://doi.org/10.2307/2409213

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.

Cutter, A. D. (2012). The polymorphic prelude to Bateson – Dobzhansky – Muller incompatibilities. *Trends in Ecology & Evolution*, *27*(4), 210–219. https://doi.org/10.1016/j.tree.2011.11.004

Degner, J. F., Marioni, J. C., Pai, A. A., Pickrell, J. K., Nkadori, E., Gilad, Y., & Pritchard, J. K. (2009). Effect of read-mapping biases on detecting allele-specific expression from RNA-sequencing data. *Bioinformatics*, *25*(24), 3207–3212. https://doi.org/10.1093/bioinformatics/btp579

DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C., … Daly, M. J. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nature Genetics*, *43*(5), 491–498. https://doi.org/10.1038/ng.806

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. https://doi.org/10.1093/bioinformatics/bts635

Fritz, D. I., Hanada, T., Lu, Y., Martin Johnston, J., & Chishti, A. H. (2014). MPP1/p55 gene deletion in a hemophilia A patient with ectrodactyly and severe developmental defects. *American Journal of Hematology*, *94*(1), 29–32. https://doi.org/10.1002/ajh.25323

Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: A graphical gene-set enrichment tool for animals and plants. *Bioinformatics*, *36*(8), 2628–2629. https://doi.org/10.1093/bioinformatics/btz931

Goncalves, A., Leigh-Brown, S., Thybert, D., Stefflova, K., Turro, E., Flicek, P., … Marioni, J. C. (2012). Extensive compensatory *cis-trans* regulation in the evolution of mouse gene expression. *Genome Research*, *22*(12), 2376–2384. https://doi.org/10.1101/gr.142281.112

Guerrero, R. F., Posto, A. L., Moyle, L. C., & Hahn, M. W. (2016). Genome-wide patterns of regulatory divergence revealed by introgression lines. *Evolution*, *70*(3), 696–706. https://doi.org/10.1111/evo.12875

Huang, S., Zhang, W., Chang, X., & Guo, J. (2019). Overlap of periodic paralysis and paramyotonia congenita caused by SCN4A gene mutations. *Channels*, *13*(1), 110–119. https://doi.org/10.1080/19336950.2019.1600967

Johnson, N. A., & Porter, A. H. (2000). Rapid speciation via parallel, directional selection on regulatory genetic pathways. *Journal of Theoretical Biology*, *205*(4), 527–542. https://doi.org/10.1006/jtbi.2000.2070

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Kang, P., & Svoboda, K. K. H. (2005). Epithelial-mesenchymal transformation during craniofacial development. *Journal of Dental Research*, *84*(8), 678–690. https://doi.org/10.1177/154405910508400801

Kerwin, R. E., & Sweigart, A. L. (2020). Rampant misexpression in a *Mimulus* (monkeyflower) introgression line caused by hybrid sterility, not regulatory divergence. *Molecular Biology and Evolution*. *37*(7):2084–2098. https://doi.org/10.1093/molbev/msaa071

Kulmuni, J., & Westram, A. M. (2017). Intrinsic incompatibilities evolving as a by-product of divergent ecological selection: Considering them in empirical studies on divergence with gene flow. *Molecular Ecology*, *26*(12), 3093–3103. https://doi.org/10.1111/mec.14147

Landry, C. R., Hartl, D. L., & Ranz, J. M. (2007). Genome clashes in hybrids: Insights from gene expression. *Heredity*, *99*(5), 483–493. https://doi.org/10.1038/sj.hdy.6801045

Landry, C. R., Wittkopp, P. J., Taubes, C. H., Ranz, J. M., Clark, A. G., & Hartl, D. L. (2005). Compensatory *cis-trans* evolution and the dysregulation of gene expression in interspecific hybrids of drosophila. *Genetics*, *171*(4), 1813–1822. https://doi.org/10.1534/genetics.105.047449

Larson, E. L., Vanderpool, D., Sarver, B. A. J., Callahan, C., Keeble, S., Provencio, L. P., … Good, J. M. (2018). The evolution of polymorphic hybrid incompatibilities in house mice. *Genetics*, *209*(July), 845–859. https://doi.org/10.1534/genetics.118.300840

Lencer, E. S., & McCune, A. R. (2018). An embryonic staging series up to hatching for *Cyprinodon variegatus*: An emerging fish model for developmental, evolutionary, and ecological research. *Journal of Morphology*, *279*(11), 1559–1578. https://doi.org/10.1002/jmor.20870

Lencer, E. S., Riccio, M. L., & McCune, A. R. (2016). Changes in growth rates of oral jaw elements produce evolutionary novelty in bahamian pupfish. *Journal of Morphology*, *277*(7), 935–947. https://doi.org/10.1002/jmor.20547

Lencer, E. S., Warren, W. C., Harrison, R., & Mccune, A. R. (2017). The *Cyprinodon variegatus* genome reveals gene expression changes underlying differences in skull morphology among closely related species. *BMC Genomics*, *18*(1), 424. https://doi.org/10.1186/s12864-017-3810-7

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Mack, K. L., Campbell, P., & Nachman, M. W. (2016). Gene regulation and speciation in house mice. *Genome Research*, *26*(4), 451–461. https://doi.org/10.1101/gr.195743.115.26

Mack, K. L., & Nachman, M. W. (2017). Gene regulation and speciation. *Trends in Genetics*, *33*(1), 68–80. https://doi.org/10.1016/j.tig.2016.11.003

Maheshwari, S., & Barbash, D. A. (2012). Cis-by-trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. *PLoS Genetics*, *8*(3), e1002597. https://doi.org/10.1371/journal.pgen.1002597

Malone, J. H., & Michalak, P. (2008). Gene expression analysis of the ovary of hybrid females of *Xenopus laevis* and *X. muelleri*. *BMC Evolutionary Biology*, *8*(1), 82. https://doi.org/10.1186/1471-2148-8-82

Manceau, M., Domingues, V. S., Mallarino, R., & Hoekstra, H. E. (2011). The developmental role of Agouti in color pattern evolution. *Science*, *331*(6020), 1062–1065. https://doi.org/10.1126/science.1200684

Martin, C. H. (2016a). Context-dependence in complex adaptive landscapes: Frequency and trait-dependent selection surfaces within an adaptive radiation of Caribbean pupfishes. *Evolution*, *70*(6), 1265–1282. https://doi.org/10.1111/evo.12932.This

Martin, C. H. (2016b). The cryptic origins of evolutionary novelty: 1000-fold faster trophic diversification rates without increased ecological opportunity or hybrid swarm. *Evolution*, *70*(11), 2504–2519. https://doi.org/10.1111/evo.13046

Martin, C. H., Erickson, P. A., & Miller, C. T. (2017). The genetic architecture of novel trophic specialists: Larger effect sizes are associated with exceptional oral jaw diversification in a pupfish adaptive radiation. *Molecular Ecology*, *26*(2), 624–638. https://doi.org/10.1111/mec.13935

Martin, C. H., & Feinstein, L. C. (2014). Novel trophic niches drive variable progress towards ecological speciation within an adaptive radiation of pupfishes. *Molecular Ecology*, *23*(7), 1846–1862. https://doi.org/10.1111/mec.12658

Martin, C. H., & Gould, K. (2019). Field manipulation of competition among hybrids reveals dynamic and highly stable features of a complex fitness landscape driving adaptive radiation. *Unpublished*, *BioRxiv* doi: https://doi.org/10.1101/756908

Martin, C. H., & Wainwright, P. C. (2013a). A remarkable species flock of *Cyprinodon* pupfishes endemic to San Salvador Island, Bahamas. *Bulletin of the Peabody Museum of Natural History*, *54*(2), 231–241. https://doi.org/10.3374/014.054.0201

Martin, C. H., & Wainwright, P. C. (2013b). Multiple fitness peaks on the adaptive landscape drive adaptive radiation in the wild. *Science*, *339*(6116), 208–211. https://doi.org/10.1126/science.1227710

Mayba, O., Gilbert, H. N., Liu, J., Haverty, P. M., Jhunjhunwala, S., Jiang, Z., … Zhang, Z. (2014). MBASED: Allele-specific expression detection in cancer tissues and cell lines. *Genome Biology*, *15*(8), 1–21. https://doi.org/10.1186/s13059-014-0405-3

McGirr, J. A., & Martin, C. H. (2017). Novel candidate genes underlying extreme trophic specialization in caribbean pupfishes. *Molecular Biology and Evolution*, *34*(4), 873–888. https://doi.org/10.1093/molbev/msw286

McGirr, J. A., & Martin, C. H. (2018). Parallel evolution of gene expression between trophic specialists despite divergent genotypes and morphologies. *Evolution Letters*, *2*(2), 62–75. https://doi.org/10.1002/evl3.41

McGirr, J. A., & Martin, C. H. (2019). Hybrid gene misregulation in multiple developing tissues within a recent adaptive radiation of *Cyprinodon* pupfishes. *PLoS One*, *14*(7), e0218899. https://doi.org/10.1101/372912

McManus, C. J., Coolon, J. D., Duff, M. O., Eipper-Mains, J., Graveley, B. R., & Wittkopp, P. J. (2010). Regulatory divergence in *Drosophila* revealed by mRNA-seq. *Genome Research*, *20*(6), 816–825. https://doi.org/10.1101/gr.102491.109

Nosil, P. (2012). Ecological speciation and its alternatives. In *Ecological speciation*. New York, NY: Oxford University Press.

Orr, H. A. (1996). Anecdotal, historical and critical commentaries on genetics. *Genetics*, *144*, 1331–1335.

Ortíz-Barrientos, D., Counterman, B. A., & Noor, M. A. F. (2007). Gene expression divergence and the origin of hybrid dysfunctions. *Genetica*, *129*(1), 71–81. https://doi.org/10.1007/s10709-006-0034-1

Otto, G. P., Wu, M. Y., Kazgan, N., Anderson, O. R., & Kessin, R. H. (2004). Dictyostelium macroautophagy mutants vary in the severity of their developmental defects. *Journal of Biological Chemistry*, *279*(15), 15621–15629. https://doi.org/10.1074/jbc.M311139200

Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528. https://doi.org/10.1093/bioinformatics/bty633

Parry, J. W. L., Carleton, K. L., Spady, T., Carboo, A., Hunt, D. M., & Bowmaker, J. K. (2005). Mix and match color vision: Tuning spectral sensitivity by differential opsin gene expression in Lake Malawi cichlids. *Current Biology*, *15*(19), 1734–1739. https://doi.org/10.1016/j.cub.2005.08.010

Pavey, S. A., Collin, H., Nosil, P., & Rogers, S. M. (2010). The role of gene expression in ecological speciation. *Annals of the New York Academy of Sciences*, *1206*(1), 110–129. https://doi.org/10.1111/j.1749-6632.2010.05765.x

Pavlidis, P., Živković, D., Stamatakis, A., & Alachiotis, N. (2013). SweeD: Likelihood-based detection of selective sweeps in thousands of genomes. *Molecular Biology and Evolution*, *30*(9), 2224–2234. https://doi.org/10.1093/molbev/mst112

Ranz, J. M., Namgyal, K., Gibson, G., Hartl, D. L., Ranz, J. M., Namgyal, K., … Hartl, D. L. (2004). Anomalies in the expression profile of interspecific hybrids of *Drosophila melanogaster* and *Drosophila simulans*. *Genome Research*, *14*(3), 373–379. https://doi.org/10.1101/gr.2019804

Reed, L. K., & Markow, T. A. (2004). Early events in speciation: Polymorphism for hybrid male sterility in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(24), 9009–9012.

Renaut, S., Nolte, A. W., & Bernatchez, L. (2009). Gene expression divergence and hybrid misexpression between lake whitefish species pairs (*Coregonus* spp. *Salmonidae*). *Molecular Biology and Evolution*, *26*(4), 925–936. https://doi.org/10.1093/molbev/msp017

Rieseberg, L. H., Archer, M. A., & Wayne, R. K. (1999). Transgressive segregation, adaptation and speciation. *Heredity*, *83*(4), 363–372. https://doi.org/10.1038/sj.hdy.6886170

Schluter, D. (2000). *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.

Schumer, M., & Brandvain, Y. (2016). Determining epistatic selection in admixed populations. *Molecular Ecology*, *25*(11), 2577–2591. https://doi.org/10.1111/mec.13641

Schumer, M., Cui, R., Powell, D. L., Dresner, R., Rosenthal, G. G., & Andolfatto, P. (2014). High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *elife*, *2014*(3), 1–21. https://doi.org/10.7554/eLife.02535

Servedio, M. R., Van Doorn, G. S., Kopp, M., Frame, A. M., & Nosil, P. (2011). Magic traits in speciation: "magic" but not rare? *Trends in Ecology and Evolution*, *26*(8), 389–397. https://doi.org/10.1016/j.tree.2011.04.005

Signor, S. A., & Nuzhdin, S. V. (2018). The evolution of gene expression in *cis* and *trans*. *Trends in Genetics*, *34*(7), 532–544. https://doi.org/10.1016/j.tig.2018.03.007

St John, M. E., McGirr, J. A., & Martin, C. H. (2019). The behavioral origins of novelty: Did increased aggression lead to scale-eating in pupfishes? *Behavioral Ecology*, *30*(2), 557–569. https://doi.org/10.1093/beheco/ary196

St. John, M. E., Holzman, R., & Martin, C. H. (2020). Rapid adaptive evolution of scale-eating kinematics to a novel ecological niche. *The Journal of Experimental Biology*, *223*(6), jeb217570. https://doi.org/10.1242/jeb.217570

Stadel, D., Millarte, V., Tillmann, K. D., Huber, J., Tamin-Yecheskel, B.-C., Akutsu, M., … Behrends, C. (2015). TECPR2 cooperates with LC3C to regulate COPII-dependent ER export. *Molecular Cell*, *60*(1), 89–104. https://doi.org/10.1016/j.molcel.2015.09.010

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033

Svensson, A., Libelius, R., Tågerud, S. (2008). Semaphorin 6C expression in innervated and denervated skeletal muscle. *J Mol Histol*, *39*(1):5-13. https://doi.org/10.1007/s10735-007-9113-6

Swain Lenz, D., Riles, L., & Fay, J. C. (2014). Heterochronic meiotic misexpression in an interspecific yeast hybrid. *Molecular Biology and Evolution*, *31*(6), 1333–1342. https://doi.org/10.1093/molbev/msu098

Thompson, A. C., Capellini, T. D., Guenther, C. A., Chan, Y. F., Infante, C. R., Menke, D. B., & Kingsley, D. M. (2018). A novel enhancer near the Pitx1 gene influences development and evolution of pelvic appendages in vertebrates. *elife*, *7*, 1–21. https://doi.org/10.7554/elife.38555

True, J. R., & Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evolution and Development*, *3*(2), 109–119. https://doi.org/10.1046/j.1525-142X.2001.003002109.x

Tulchinsky, A. Y., Johnson, N. A., Watt, W. B., & Porter, A. H. (2014). Hybrid incompatibility arises in a sequence-based bioenergetic model of transcription factor binding. *Genetics*, *198*(3), 1155–1166. https://doi.org/10.1534/genetics.114.168112

Turissini, D. A., McGirr, J. A., Patel, S. S., David, J. R., & Matute, D. R. (2018). The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. *Molecular Biology and Evolution*, *35*(2), 312–334. https://doi.org/10.1093/molbev/msx271

Van De Geijn, B., Mcvicker, G., Gilad, Y., & Pritchard, J. K. (2015). WASP: Allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods*, *12*(11), 1061–1063. https://doi.org/10.1038/nmeth.3582

Wei, K. H. C., Clark, A. G., & Barbash, D. A. (2014). Limited gene misregulation is exacerbated by allele-specific upregulation in lethal hybrids between *Drosophila melanogaster* and *Drosophila simulans*. *Molecular Biology and Evolution*, *31*(7), 1767–1778. https://doi.org/10.1093/molbev/msu127

West, R. J. D., & Kodric-Brown, A. (2015). Mate choice by both sexes maintains reproductive isolation in a species flock of pupfish (*Cyprinodon* spp) in the Bahamas. *Ethology*, *121*(8), 793–800. https://doi.org/10.1111/eth.12394

Whitehead, A., & Crawford, D. L. (2006). Variation within and among species in gene expression: Raw material for evolution. *Molecular Ecology*, *15*(5), 1197–1211. https://doi.org/10.1111/j.1365-294X.2006.02868.x

Wittkopp, P. J., Haerum, B. K., & Clark, A. G. (2004). Evolutionary changes in *cis* and *trans* gene regulation. *Nature*, *430*(6995), 85–88. https://doi.org/10.1038/nature02698

Wray, G. A., Hahn, M. W., Abouheif, E., Balhoff, J. P., Pizer, M., Rockman, M. V., & Romano, L. A. (2003). The evolution of transcriptional regulation in eukaryotes. *Molecular Biology and Evolution*, *20*(9), 1377–1419. https://doi.org/10.1093/molbev/msg140

Zhou, X., Carbonetto, P., & Stephens, M. (2013). Polygenic modeling with bayesian sparse linear mixed models. *PLoS Genetics*, *9*(2), e1003264. https://doi.org/10.1371/journal.pgen.1003264

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: McGirr JA, Martin CH. Ecological divergence in sympatry causes gene misexpression in hybrids. *Mol Ecol*. 2020;29:2707–2721. https://doi.org/10.1111/mec.15512