MicroRNA Gene Regulation in Extremely Young and Parallel Adaptive Radiations of Crater Lake Cichlid Fish

Paolo Franchini 🔵 ,*[†,1] Peiwen Xiong,[†,1] Carmelo Fruciano,[1,2] Ralf F. Schneider,[‡,1] Joost M. Woltering,[1] Christopher Darrin Hulsey,[1] and Axel Meyer*,[1]

[1]Lehrstuhl für Zoologie und Evolutionsbiologie, Department of Biology, University of Konstanz, Konstanz, Germany

[2]Institut de biologie de l'Ecole normale supérieure (IBENS), Ecole normale supérieure, CNRS, PSL Université Paris, Paris, France

[‡]Present address: Marine Ecology, Helmholtz-Zentrum für Ozeanforschung Kiel (GEOMAR), Düsternbrooker Weg 20, Kiel, Germany

[†]These authors contributed equally to this work as first authors.

*Corresponding authors: E-mails: paolo.franchini@uni-konstanz.de; axel.meyer@uni-konstanz.de.

Abstract

Cichlid fishes provide textbook examples of explosive phenotypic diversification and sympatric speciation, thereby making them ideal systems for studying the molecular mechanisms underlying rapid lineage divergence. Despite the fact that gene regulation provides a critical link between diversification in gene function and speciation, many genomic regulatory mechanisms such as microRNAs (miRNAs) have received little attention in these rapidly diversifying groups. Therefore, we investigated the posttranscriptional regulatory role of miRNAs in the repeated sympatric divergence of Midas cichlids (*Amphilophus* spp.) from Nicaraguan crater lakes. Using miRNA and mRNA sequencing of embryos from five Midas species, we first identified miRNA binding sites in mRNAs and highlighted the presences of a surprising number of novel miRNAs in these adaptively radiating species. Then, through analyses of expression levels, we identified putative miRNA/gene target pairs with negatively correlated expression level that were consistent with the role of miRNA in downregulating mRNA. Furthermore, we determined that several miRNA/gene pairs show convergent expression patterns associated with the repeated benthic/limnetic sympatric species divergence implicating these miRNAs as potential molecular mechanisms underlying replicated sympatric divergence. Finally, as these candidate miRNA/gene pairs may play a central role in phenotypic diversification in these cichlids, we characterized the expression domains of selected miRNAs and their target genes via in situ hybridization, providing further evidence that miRNA regulation likely plays a role in the Midas cichlid adaptive radiation. These results provide support for the hypothesis that extremely quickly evolving miRNA regulation can contribute to rapid evolutionary divergence even in the presence of gene flow.

Key words: adaptive radiation, regulatory evolution, miRNA regulation, sympatric speciation, trophic divergence.

Introduction

Sympatric speciation, which frequently involves an evolutionary split of lineages into divergent ecological niches without geographical barriers, remains one of the most controversial topics in evolutionary biology (Mayr 1963; Jiggins 2006; Bolnick and Fitzpatrick 2007; Foote 2018). This controversy is in part due to the fact that the mechanisms generating phenotypic and genetic divergence in the face of gene flow remain poorly understood (Mayr 1963; Via 2001; Coyne 2007). However, it is clear that divergent selection regimes associated with different ecological niches should lead to the accumulation of genetic variants in diverging populations (Ravinet et al. 2017). Yet, gene flow and recombination are expected to homogenize these variants and disrupt beneficial allele combinations by breaking their physical linkage (Feder and Nosil 2010; Bird et al. 2012). Thus, sympatric speciation was long thought to be an unlikely mode of speciation (Mayr 1963; Bolnick and Fitzpatrick 2007). However, in the last few decades, several theoretical models and empirical studies have suggested that speciation with gene flow could be surprisingly common (Dieckmann and Doebeli 1999; Papadopulos et al. 2014). Furthermore, it is now clear that loci influencing adaptive divergence can exhibit linkage disequilibrium, or the nonrandom association of alleles at different loci, despite exhibiting little or no physical linkage (Feder and Nosil 2010; Flaxman et al. 2014). Within this framework, one can now address previously untraceable questions about the molecular mechanisms promoting divergence in the face of gene flow such as "Are there particular genetic architectures of adaptive traits that facilitate sympatric speciation?," "How is differential gene expression associated with trait variation?," and "What are the roles of gene regulatory mechanisms in promoting this divergence?" With the advent of the "genomic era" and the increasing availability of large-scale sequencing as well as powerful bioinformatic tools, we can now address questions about the molecular basis of sympatric divergence with unprecedented power and resolution (Seehausen et al. 2014; Foote 2018).

Historically, studies investigating the molecular mechanisms underlying adaptive evolution have focused on

protein-coding sequence variation (Coyne and Hoekstra 2007; Spielmann and Mundlos 2016). However, since the ground-breaking work of Britten and Davidson (1969), gene regulation has increasingly been acknowledged as one of the major drivers of species differentiation. Yet, the hypothesis that changes in gene expression commonly underlie phenotypic divergence has only become testable with the development of approaches that permit quantitative estimates of gene expression (Ozsolak and Milos 2011). Now, gene regulation is widely considered one of the fastest and most effective mechanisms underlying adaptive evolution (Wittkopp et al. 2003; Shapiro et al. 2004; Carroll 2008; Chan et al. 2010) and is thought to be particularly relevant in the initial stages of speciation (Pavey et al. 2010). Although nonsynonymous changes in widely expressed protein-coding sequences can produce major pleiotropic effects on the phenotype (Chan et al. 2010), gene regulatory mechanisms such as transcription factors, DNA methylation, and microRNAs (miRNAs) can produce a more fine-tuned outcome. By modifying the gene expression that facilitates adaptation and potentially reproductive isolation, gene regulation could lay the foundation for genetically based divergence even in the face of gene flow (Pavey et al. 2010; Davidson and Balakrishnan 2016; Mack et al. 2016).

It is now known that miRNAs can directly downregulate gene expression through mRNA cleavage and/or translational repression by targeted binding to mRNAs (Friedman et al. 2009; Wahid et al. 2010). Additionally, miRNAs have been shown to regulate many proteins involved in processes as disparate as cellular differentiation, cell fate determination, proliferation, and tissue development (Chen et al. 2006; Stefani and Slack 2008). As fundamental regulators of posttranscriptional gene expression, miRNAs can also contribute to environmental adaptation, phenotypic diversification, and speciation (Franchini et al. 2016; Li et al. 2016; Rastorguev et al. 2017; Xiong et al. 2018). Most protein-coding genes are regulated by miRNAs (Friedman et al. 2009) and the expression of miRNAs is often under tight temporal and spatial control (Ha and Kim 2014). Also, individual genes can be regulated by multiple miRNAs when respective binding sites are present (Friedman et al. 2009). In animals, these mature miRNAs usually bind to the 3′ untranslated region (3′ UTRs) of targeted mRNAs via Watson–Crick base pairing to a "seed" area located on the translated 5′ genomic region of an mRNA (Friedman et al. 2009). Therefore, the interactive effect of miRNA on mRNA expression is tightly linked physically to the location of the protein in the genome regardless of the genomic location of the precursor miRNA sequence. However, the precursor miRNA sequence can reside anywhere in the genome creating a "functional linkage disequilibrium" of miRNA/gene pair regulation without any necessary physical linkage of the miRNA and protein sequence. The existence of this functional linkage without any necessary physical linkage makes miRNAs particularly interesting genomic targets for adaptation and speciation in the face of the homogenizing effects of gene flow.

Cichlid fish are ideal for testing the evolutionary effects of miRNA regulation, as they have formed some of the most spectacular adaptive radiations in the animal kingdom, frequently in sympatry and often within only a few thousand generations (Verheyen et al. 2003; Barluenga et al. 2006; Irisarri et al. 2018). Additionally, the complete genomes of several lineages of these fishes are now available (Brawand et al. 2014; Malinsky et al. 2018). The *Amphilophus citrinellus* species complex (Midas cichlids) is an especially suitable candidate for genomic dissection as this group of fishes provide an especially recent replicated natural experiment (Elmer, Kusche, et al. 2010; Elmer, Lehtonen, et al. 2010). In <20,000 years, fish from the two large Nicaraguan lakes have colonized several small and geographically isolated crater lakes (some of which are <2,000 years old) and adaptively radiated to repeatedly exploit the same set of ecological niches (fig. 1 and supplementary fig. S1, Supplementary Material online). Perhaps the most spectacular divergence in the Midas radiation has occurred within crater Lakes Apoyo and Xiloá, wherein colonizing cichlids have putatively sympatrically speciated as well as independently and repeatedly evolved similar phenotypes associated with the bottom-dwelling (benthic) and open-water (limnetic) habitats of these lakes (Barluenga et al. 2006; Elmer et al. 2014). Both limnetic species differ phenotypically from the benthic species in having a more elongated body (Franchini et al. 2014) as well as in the shape and dentition of their pharyngeal jaws, which are components of the pharyngeal arches used to process food items (Liem 1973). Benthic species have enlarged pharyngeal jaws and process more durable prey compared with limnetic species (Elmer et al. 2014; Fruciano et al. 2016). Recent genomic in silico surveys in these fishes have begun to elucidate the role of gene regulation in their adaptive divergence (Franchini et al. 2016; Xiong et al. 2018) and have shown relaxed purifying selection at conserved miRNA binding sites (Franchini et al. 2016) as well as longer and faster evolving 3′ UTRs (Xiong et al. 2018). However, we know little about the role of novel miRNAs in these adaptively radiating species and whether novel or more conserved miRNA/gene pairs show differential expression in this radiation. Furthermore, determining if any miRNA/gene pairs show convergent expression associated with the repeated benthic/limnetic sympatric species divergence in Lakes Apoyo and Xiloá is essential if miRNAs have provided a molecular regulatory mechanism for replicated divergence in the face of gene flow. Finally, documenting the antagonistic expression of miRNA/gene pairs in developing phenotypes that are known to have diverged during Midas diversification would be important to elucidate if miRNA regulation does underlie some of the phenotypic divergence present in the Midas radiation.

In this study, we use extensive transcriptomic sequencing of both mRNAs and small RNAs to characterize the miRNA regulatory landscape in Midas cichlids and test whether gene regulation by miRNA could be one of the mechanisms driving their phenotypic diversification. We focus our investigation on a benthic species native to the ancestral location of Midas cichlids in Lake Nicaragua as well as two benthic–limnetic species pairs that have speciated sympatrically in crater Lakes Apoyo and Xiloá. We first identify conserved and novel miRNAs and their binding sites on protein 3′ UTRs. Then,
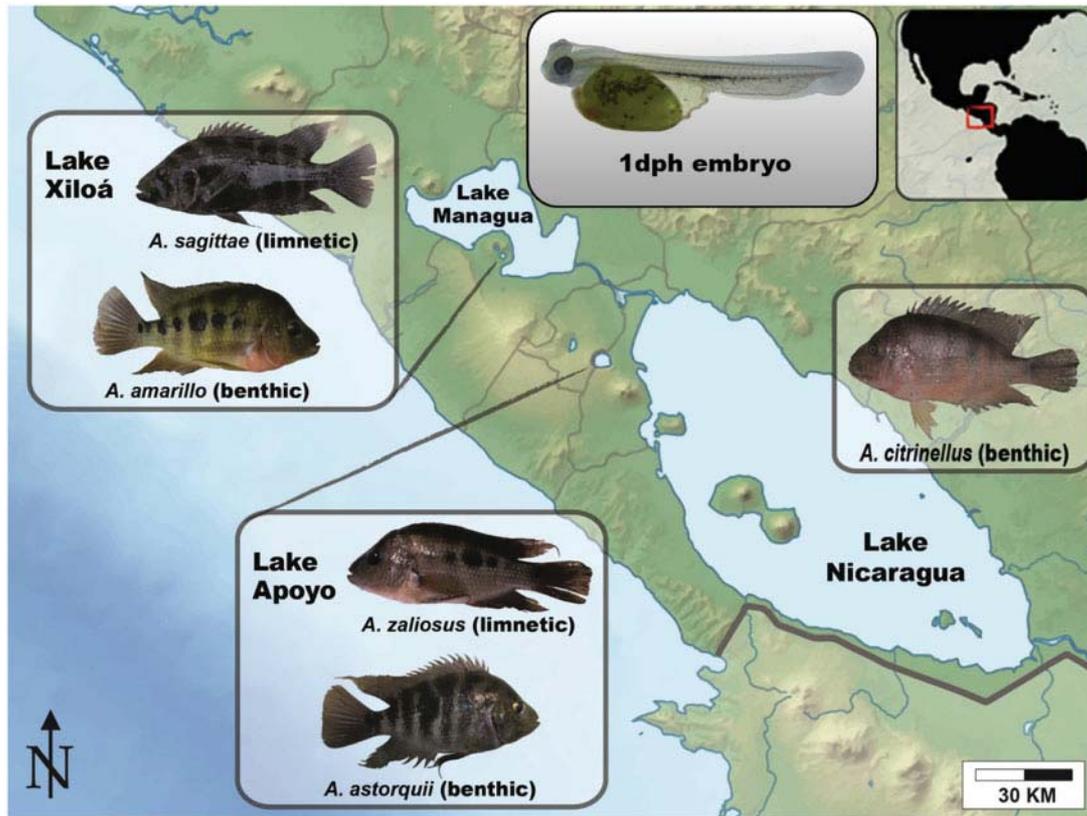
Fig. 1. Nicaraguan lakes and Midas cichlids. Photos of representative specimens of the five species of the Midas cichlid flock used in this study are shown. The upper right inset depicts the geography of the Nicaraguan lakes in the western hemisphere and a closeup of the lakes containing components of the Midas radiation are provided. For each crater lake, a species pair containing a bottom-dwelling (benthic) and open-water (limnetic) species was examined. A Midas 1 dph embryo, the developmental stage we targeted in this study, is also shown in the upper inset.

we examine expressional correlations in miRNAs and their target genes and select miRNA/gene pairs that show negative correlations between levels of miRNA and their gene target's expression. We then identify those pairs relevant to rapid divergence in lacustrine environments and divergence with gene flow by testing for DE between the putative source species and crater lake species as well as between benthic and limnetic species in the two crater lakes. Finally, we investigate the spatial expression domains of miRNA/gene pairs by in situ hybridization (ISH) to examine their expression dynamics and to obtain insight into their putative regulatory roles in the patterning of the diverging phenotypes among the sympatric Midas species.

Results

The filtered 299.4 million (M) paired-end high-quality reads (from 47.5 to 76.1 M reads per species) were used to assemble the Midas transcriptome using both a reference-guided and a de novo approach. The final set of 134,228 sequences, obtained following the workflow described in the Materials and Methods section, were processed by the clustering algorithm implemented in Corset. Using the mapping information of the fifteen 1-day posthatching (1 dph) samples on the transcript set, Corset reconstructed 31,436 clusters/genes and generated the raw count table as input for the DE analysis.

DE analysis was performed between the five species in a pairwise manner (i.e., five species, ten comparisons). Considering all the 10 species pair comparisons, we detected 15,024 genes with significant DE (false discovery rate [FDR] corrected P value < 0.05), with an average of 1,502.4 per comparison, ranging from 188 (*A. sagittae* vs. *A. amarillo*) to 3,671 DE genes (*A. astorquii* vs. *A. amarillo*) (the number of DE genes in each species pair comparisons is reported in fig. 2a). A principal component analysis (PCA) conducted on the expression of all the genes suggests a strong phylogenetic signal (fig. 2b).

Small RNA-Seq reads from the 15 samples (5 species, 3 samples per species) were used to identify miRNAs genes using the Midas genome as reference. After adapter removal, the length distributions of the reads showed a clear peak at 22 nt, indicating that mature miRNAs, usually 21–24 nt in length (Lu et al. 2008), were successfully enriched during library construction. For each species, the number of identified mature miRNAs ranged from 178 to 289, encoded by 241–369 miRNA genes. The reason for the observed discrepancy between the numbers of miRNAs and miRNA genes can be found in miRNA biogenesis, as the same mature miRNAs can be encoded by different miRNA genes (the sequence of the total 349 mature miRNAs predicted in our data set is provided in supplementary table 3, Supplementary Material online). We found 147 mature miRNAs shared among these
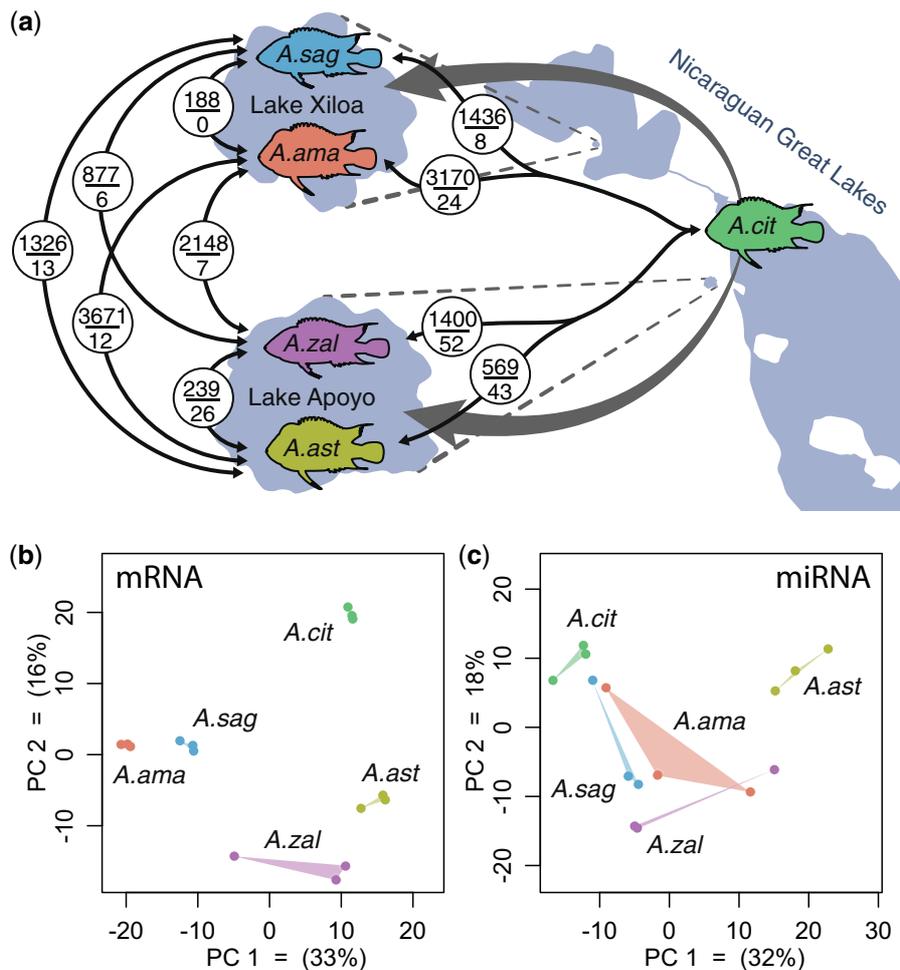
Fig. 2. Gene expression patterns across sampled Midas species. A cartoon of Midas cichlid divergence shows (*a*) the number of differentially expressed (DE) mRNAs (upper number in each circle) and differential expressed miRNAs (lower number) in pairwise species comparisons (connected with arrows). Numbers of DE mRNAs are highest in between-lake comparisons, which reflects their phylogenetic divergence. Although the limnetic *Amphilophus zaliosus* (*A. zal*) and *A. sagittae* (*A. sag*) both have ~1,400 mRNA transcripts differentiating it from the source population represented by *A. citrinellus* (*A. cit*), the Lake Apoyo benthic species *A. astorquii* (*A. ast*) has relatively few DE transcripts relative to *A. citrinellus* and compared with *A. amarillo* (*A. ama*) from Lake Xiloá. (*b*) A scatterplot of individuals' mRNA PC1 and PC2 scores. PC1 most prominently differentiates Lake Apoyo and the source population *A. citrinellus* from the two Lake Xiloá species. PC2 separates the source population from all crater lake species, particularly those of Lake Apoyo. (*c*) A scatterplot of individuals' miRNA PC1 and PC2 scores. PC1 most clearly separates *A. citrinellus* from *A. astorquii*, whereas PC2 rather separates *A. astorquii* and *A. citrinellus*, from *A. zaliosus*. Although the Lake Xiloá species are only slightly distinguishable in the miRNA PCA (supplementary fig. S1, Supplementary Material online), in general, there is ample mRNA and miRNA divergence for putative negatively correlated pairs to differentiate all of these recently diverged species.

cichlid fishes, whereas 3–21 identical mature miRNAs showed species-specific expression (fig. 3). By comparing our miRNA data set with miRBase's miRNA data, we identified a range of 142–162 conserved miRNAs and a range of 36–121 novel miRNAs across the five focal cichlid species exhibiting variation in phylogenetic signal (supplementary fig. S2, Supplementary Material online). The PCA on the expression of the whole miRNAs' data set suggests a weaker phylogenetic signal when compared with the PCA of the gene expression, where species and lakes of origin formed clearly distinguishable clusters in the PC space (fig. 2*b*).

Using the same approach that we applied to the RNA-Seq data set, differentially expressed (DE) miRNAs were identified for each of the ten species pair comparisons that are possible

involving the five species. A total of 191 miRNAs were found to be DE (FDR corrected P value < 0.05), with an average of 19.1 per comparison, ranging from 0 (*A. sagittae* vs. *A. amarillo*) to 52 DE miRNAs (*A. citrinellus* vs. *A. zaliosus*) (the number of DE miRNAs in each species pair comparisons is reported in fig. 2*a*).

Target prediction was carried out using the whole set of identified Midas miRNAs and the longest transcripts among those belonging to a Corset's cluster. Using the program miRanda, we predicted that 9,528 transcripts were potentially targeted by 349 miRNAs (in total 44,516 miRNA/gene pairs). The target prediction analysis conducted with miRanda was supported by the program TargetScan, which retrieved 42,199 mRNA/gene target pairs identified by miRanda (94.8%). A
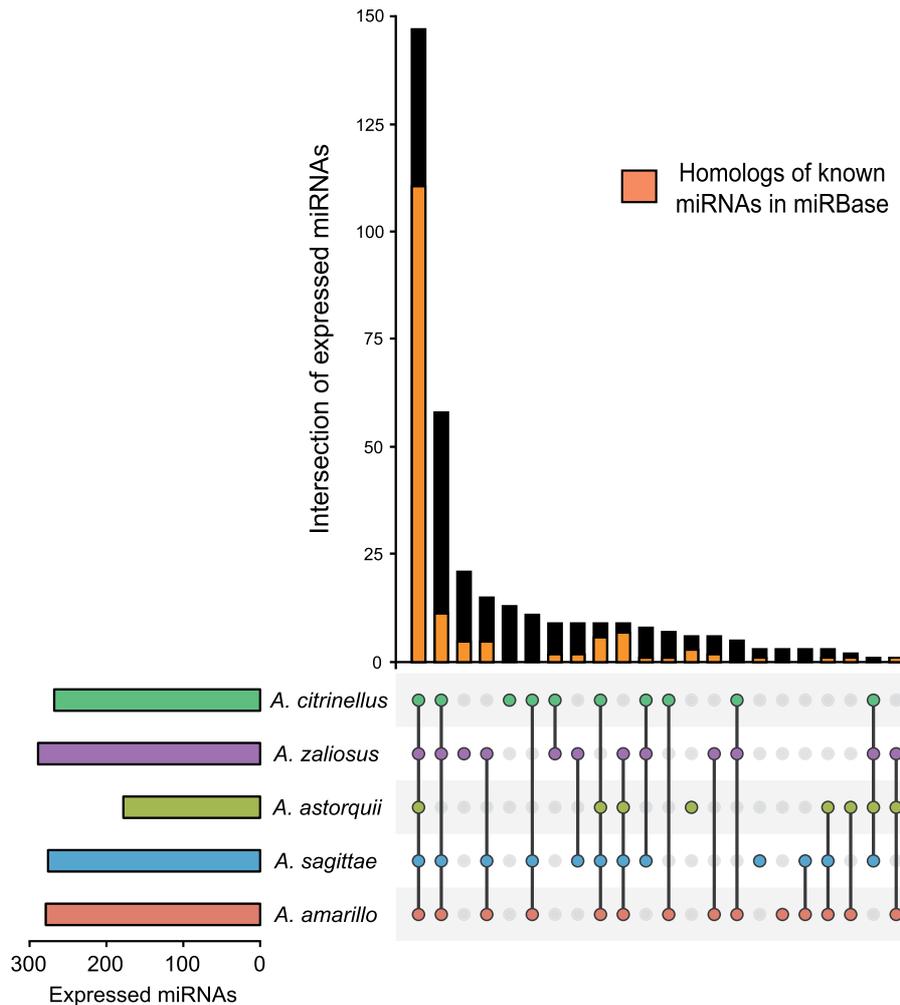
Fig. 3. Overview of miRNA numbers uniquely expressed in one species or shared among different groups within the five Midas cichlid species. Color code for the five species is the same used in figure 2. Dots depict which lineages share the set of miRNAs. Lines connect the dots of monophyletic groups that share particular miRNAs. The bar above each line connecting dots displays the number of miRNAs present in that group of cichlids. In the bar plots, the orange portion of each bar indicates the number of expressed miRNAs of our data set that are homologous to known miRNAs in miRBase. Black represents miRNAs in Midas that were not found in miRBase.

total of 2,536 transcripts were found to be targeted by only novel miRNAs (from 1 to 12 targets per gene). This gene set showed significant enrichment for three Gene Ontology (GO) terms related to cellular components (GO:0044424, "intracellular part"; GO:0005622, "intracellular"; GO:0016272, "prefoldin complex").

Using pairwise species comparisons, we detected a range from 3 (*A. astorquii* vs. *A. sagittae* and *A. zaliosus* vs. *A. sagittae*) to 58 (*A. citrinellus* vs. *A. amarillo*) miRNA/gene robustly negatively correlated pairs, for a total of 158 pairs (supplementary table 4, Supplementary Material online). As 23 of these were detected in more than one species pair comparison, the absolute number of unique robustly negatively correlated pairs identified across the whole data set was 124. As a general pattern, the comparisons between *A. citrinellus* from the great Lake Nicaragua and the crater lake species tended to show the largest number of robustly negatively correlated pairs (72%, 113 out the total 158).

The 158 robustly negatively correlated pairs involved 87 unique genes and 48 unique miRNAs. Of these, 64 genes were

targeted by a single miRNA, whereas the remaining 23 genes were targeted by multiple miRNAs—from 2 to 3 miRNAs, with the exception of 1 gene targeted by 5 DE miRNA. Notably, these five miRNAs showed the same putative effect sign on the target gene with which they were negatively correlated, as they were consistently overexpressed in *A. amarillo* when compared with *A. citrinellus*. Of the total 48 unique miRNAs, 29 had targets in more than 1 gene (from 2 to 8 target genes), whereas we identified 19 miRNAs with binding sites in a single gene (supplementary table 4, Supplementary Material online).

Using the mapping of variants against the 4,032 protein-coding loci, we obtained a character matrix of 22,560 single-nucleotide polymorphisms (SNPs). The tree inferred using SVDquartets provided reasonably robust resolution among all the individuals and species (supplementary fig. S3, Supplementary Material online). All Midas species were reconstructed as monophyletic and there was support for the two Lake Apoyo and the two Lake Xiloá species as forming monophyletic groups (100% bootstrap values), as has

Table 1. Annotation of the 13 Robustly Negatively Correlated Pairs Showing Parallel Expression in the Benthic and Limnetic Species (corrected P value < 0.05).

Midas Transcript ID	Tilapia Gene ID (Ensembl)	Gene Name	Gene Description	Midas miRNA ID	Homolog in miRBAse
STRG.10917.2	ENSONIG00000017353	ppm1e	Protein phosphatase, Mg^{2+}/Mn^{2+} dependent, 1E	mir-103	—
STRG.13060.1	ENSONIG00000016440	laptm4b	Lysosomal protein transmembrane 4 beta	mir-318	—
STRG.16240.3	ENSONIG00000011962	asb13a.1	Ankyrin repeat and SOCS box protein 13	mir-6	—
STRG.16632.1	ENSONIG00000000031	spic	Transcription factor Spi-C	mir-102	—
STRG.19865.2	ENSONIG00000015093	rab3c	RAB3C, member RAS oncogene family	miR-222-3p	miR-222-3p
STRG.20923.1	ENSONIG00000015956	zbed4	Zinc finger BED domain-containing protein 4	miR-17-5p	miR-17-5p
STRG.24389.1	ENSONIG00000019915	raraa	Retinoic acid receptor alpha	miR-24-3p	miR-24-3p
STRG.2902.1	ENSONIG00000003661	rnf144aa	Ring finger protein 144aa	mir-65	—
STRG.30402.1	ENSONIG00000017390	hla-f10al4	Class I histocompatibility antigen, F10 alpha chain	mir-345	—
STRG.30402.1	ENSONIG00000017390	hla-f10al4	Class I histocompatibility antigen, F10 alpha chain	mir-93	—
STRG.35770.2	ENSONIG00000018809	taf1b	TATA box binding protein (Tbp)-associated factor, RNA polymerase I, B	miR-203-3p	miR-203-3p
STRG.387.1	ENSONIG00000014632	pax3	Paired box 3	mir-18	—
STRG.3980.2	ENSONIG00000020316	cdh4	Cadherin-4	mir-6	—

been established before (Kautt et al. 2016). Our phylogenetic independent contrast (PIC) correlation analyses provided evidence (P value < 0.05) of negatively correlated gene expression for 851 miRNA/gene pairs (supplementary table 5, Supplementary Material online). These genes were found to be enriched for three GO terms related to protein modification (GO:0043412, "macromolecule modification"; GO:0036211, "protein modification process"; GO:0006464, "cellular protein modification process"). A similar number of miRNA/gene pairs were identified as positively correlated, as revealed by the distribution of all the correlations of PICs shown in supplementary fig. S4, Supplementary Material online). Of all the robustly negatively correlated miRNA/gene pairs we tested with multivariate analysis of variance (MANOVA), 13 were significant after controlling for multiple testing, and thus showed parallel expression in the benthic and limnetic species (the annotation of the 12 genes involved in these 13 robustly negatively correlated pairs is shown in table 1).

The first gene/miRNA robustly negatively correlated pair examined using ISH includes the well-known miRNA let-7d (our mir-69) and its predicted target immunoglobulin superfamily DCC subclass member 3 gene (igdcc3) (in our data set represented by two different transcripts: STRG.7354.1 and STRG.13143.1). Interestingly, this pair was identified both in comparisons involving A. citrinellus from Great Lake Nicaragua versus A. zaliosus in Lake Apoyo as well as in A. citrinellus versus A. Amarillo in Lake Xiloá (supplementary table 4, Supplementary Material online). Furthermore, let-7d was found to be differentially expressed in another robustly negatively correlated pair (targeting a gene of the serine threonine-protein phosphatase [psp] family; our STRG.9840.2 transcript) again involving A. citrinellus and a crater lake species, A. amarillo from Lake Xiloá. The target gene igdcc3 formed two additional robustly negatively correlated pairs with two different miRNAs in the comparisons between A. citrinellus and the two crater Lake Apoyo's species (A. citrinellus vs. A. astorqui and A. citrinellus vs. A. zaliosus) (supplementary table 4, Supplementary Material online). For

this robustly negatively correlated pair, we examined embryos both from hatching day and at 1 dph, thus allowing us to investigate the temporal and spatial expression of both the miRNA and the target gene, and to visualize the potential regulatory impact of the former on the latter. We identified a body-wide expression pattern for igdcc3 in 1 dph embryos, with more localized expression domains in the dorsal region of the head and of the trunk. As shown in figure 4, the expression of igdcc3 decreases from hatching day to 1 dph, whereas the expression of let-7d increases during development, consistent with a negative regulatory interaction. Particularly in the most anterior part of the brain, there is a clear correlation between the downregulation of igdcc3 and the upregulation of let-7d.

The second robustly negatively correlated pair we tested includes a novel Midas miRNA (mir-250) and its target (transcript STRG.40994.1 in our data set), a gene belonging to the cytochrome P450 family and with the highest homology to the medaka cyp3A40, the zebrafish cyp3A65, and the human cyp3A4 genes. This is one of the nine robustly negatively correlated pairs found in the comparisons between the crater Lake Apoyo sympatric species A. astorqui and A. zaliosus and was notably the only one involving a miRNA where no homologs in miRBase were found. At 1 dph, mir-250 had a defined expression domain in the pharyngeal arches (fig. 5) in the region where the ceratohyal and palatoquadrate develop (Woltering et al. 2018). The expression of its predicted target shows a diffuse ubiquitous pattern of expression in the head region, however with a clear downregulation in the area of mir-250 expression. Therefore, mir-250 and cyp3A40 show the complementary spatial expression dynamics expected from a negatively correlated miRNA/target pair.

Discussion

Changes in gene expression are a common source of phenotypic divergence and could promote reproductive isolation between incipient species, especially in a sympatric setting (Wittkopp et al. 2003; Pavey et al. 2010). Here, we examined
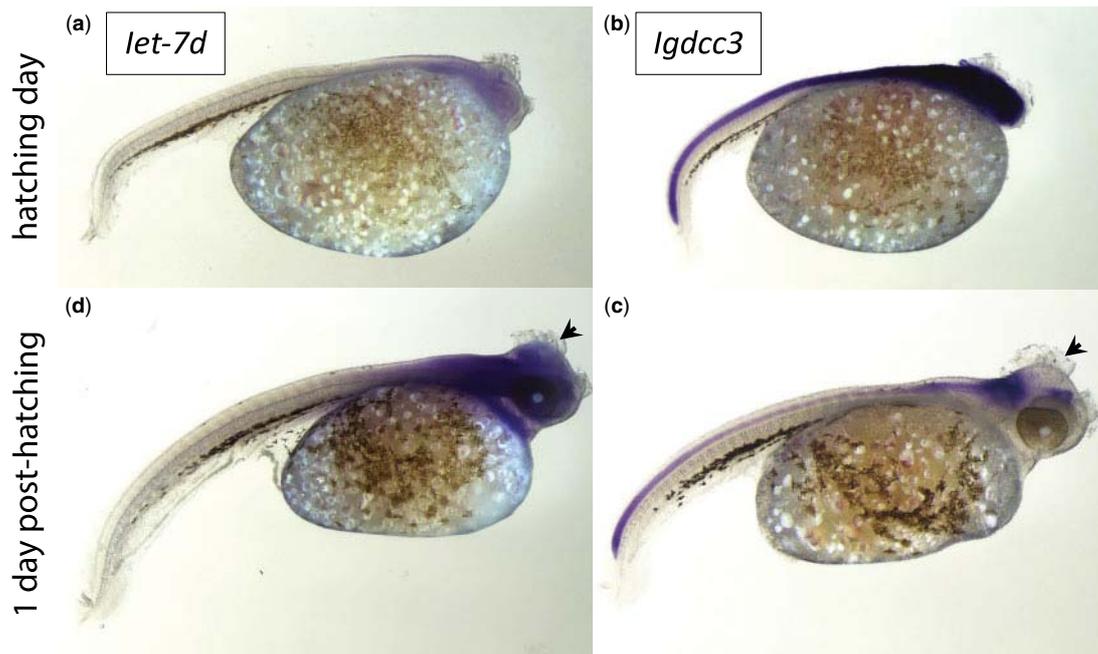
Fig. 4. ISH in *Amphilophus astorquii* of miRNA *let-7d* and its target gene *igdcc3*. When the fish embryo first hatches, the miRNA *let-7d* shows low expression in the head and in the whole dorsal area of the body (*a*). However, its target gene *igdcc3*, an immunoglobulin superfamily gene, is relatively highly expressed (*b*). At 1 dph, the expression of *let-7d* clearly increases in the embryo's head (*c*), whereas at the same developmental stage the expression of its target gene decreases (*d*). The black arrows indicate the domain were the regulatory effect of *let-7d* miRNA is most pronounced.
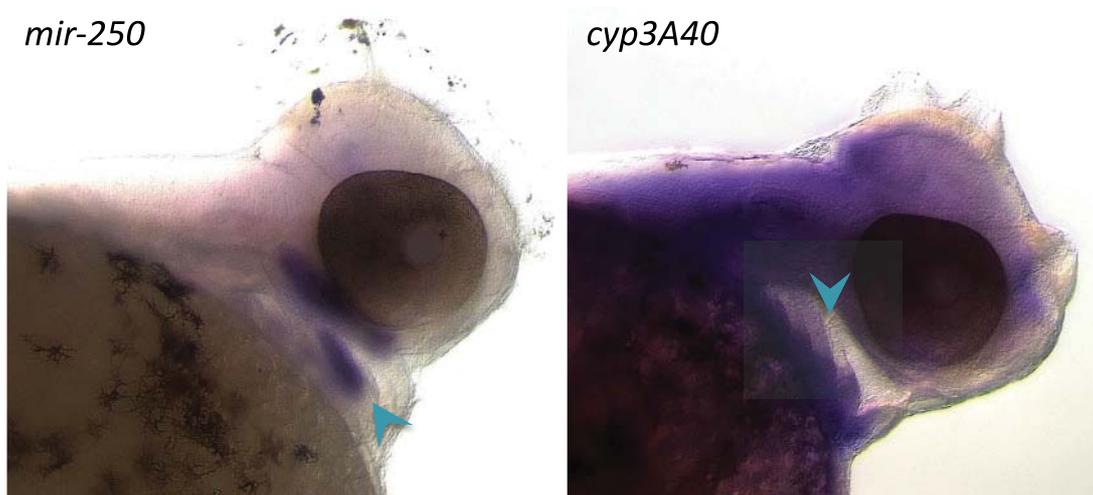

Fig. 5. ISH of miRNA *mir-250* and its target gene *cyp3A40*. This miRNA/gene negatively correlated pair was identified in the comparison between the benthic and limnetic species of crater Lake Apoyo (*Amphilophus astorqui* and *A. zaliosus*). The expression domains of a novel Midas miRNA (*mir-250*) and its target gene (a cytochrome P450 family gene, *cyp3A40*) are indicated by the blue arrows in an *A. astorquii* embryo. The upper and lower expression domains correspond to the future palatoquadrate and the ceratohyal, respectively. Here, high expression levels of *mir-250* suggest that it represses the expression of *cyp3A40* in this part of the jaw area, indicating a role of miRNA regulation in shaping the trophic apparatus.

the potential role of miRNA regulation in modulating gene expression patterns between recently diverged species from the Midas cichlid fish species complex. The extremely young Midas adaptive radiations provide replicated natural experiments that facilitate the simultaneous investigation of both the genomic mechanisms underlying sympatric speciation as well as the parallel evolution of phenotypic traits (Henning and Meyer 2014). By conducting analyses of coexpression between miRNAs and their target gene transcripts in Midas

species during early development, we first identified a large number of miRNA-regulated genes that differ between crater lake and noncrater lake Midas cichlids, consistent with a role in allopatric speciation and or phenotypic difference. Second, we also found a set of miRNA–gene pairs which show DE between all limnetic species and all benthic species, consistent with a role for these pairs during sympatric divergence. Finally, using ISH experiments to visualize the expression silencing effect of phenotypically critical miRNAs, we identified

the expression domain of a miRNA gene regulatory pair in the pharyngeal arches of a benthic–limnetic species pair.

In our study, we found a striking number of Midas miRNAs that were not present in miRBase, some of them are lake specific or even species specific (fig. 3). This result likely reflects a limited and phylogenetically unbalanced representation of miRNAs in public databases that are also often collected from only a few tissues and not throughout all developmental stages (Guerra-Assuncao and Enright 2012). However, we cannot exclude that the Midas complex in particular, or more generally cichlid fish that are one of the most species-rich lineages of vertebrates, could be characterized by a high emergence rate of novel miRNAs. The origin of these novel gene regulatory mechanisms might have facilitated the un-paralleled adaptive divergence of this group of fish (Brawand et al. 2014). Despite steadily increasing evidence on the importance of miRNA regulation in modulating morphological evolutionary change, its role in producing the extraordinary phenotypic diversity of cichlid fishes has received little attention to date (but see Loh et al. 2011; Brawand et al. 2014; Franchini et al. 2016; Xiong et al. 2018). These patterns coupled with the recently documented divergence in miRNA regulatory regions suggest that miRNAs might be an important factor underlying sympatric divergence in not only the Neotropical Midas cichlid flock (Franchini et al. 2016) but also a number of adaptively radiating lineages.

One remarkable general pattern we found in our combined analysis of gene and miRNA expression data is a high number of miRNA/gene robustly negatively correlated pairs which show different levels of expression in the comparisons between *A. citrinellus* from the Great Lake Nicaragua and the recently evolved crater lake species. This extensive variation in miRNA/gene expression levels is particularly noteworthy considering that this variation must have been produced in $<\sim$22,000 years (the age of the oldest crater lakes studied here). The most likely cause of this extensive divergence in expression levels of miRNA/gene robustly negatively correlated pairs is divergent selection due to the completely different ecological conditions that prevail between the shallow source lake and the recently formed and very deep crater Lakes Apoyo and Xiloá. This type of reorganization of gene expression might often be linked to initial adaptive responses when organisms diversify into novel environments (Pavey et al. 2010).

Although the MANOVAs highlighted a set of miRNA/gene robustly negatively correlated pairs that show generalized divergence between benthic and limnetic fish, the pairwise species comparisons failed to find clear evidence of parallelism across lakes in the miRNA/gene pairs. However, the lack of parallelism in pairwise species comparisons might merely reflect a reduced and insufficient correlational strength of designated robustly negatively correlated miRNA/gene pairs specifically in the benthic–limnetic species comparison in Lake Xiloá, likely caused by the difference in age between the two crater lakes studied here. This notion is consistent with previous independent evidence on levels of differentiation in this study system (Elmer et al. 2014; Franchini et al. 2016; Kautt et al. 2016). In fact, Lake Apoyo is the oldest crater

lake hosting Midas cichlids (ca., 22,000 years old) whereas Lake Xiloá is much younger (ca., 6,000 years old) (Kutterolf et al. 2007; Kautt et al. 2016). The divergence between benthic and limnetic species in Lake Apoyo is generally more pronounced than in Lake Xiloá, both phenotypically (Elmer, Kusche, et al. 2010) and genetically (Elmer et al. 2014; Kautt et al. 2016). Evidence even suggests that a recent admixture event has probably occurred between species from Lake Nicaragua and *A. sagittae*, the limnetic species from Lake Xiloá (Kautt et al. 2016). Additionally, our previous work on miRNA sequence variation in these species showed a pattern of relaxation of purifying selection only in Lake Apoyo (Franchini et al. 2016). Then, in light of this prior evidence, the lack of significant divergence in robustly negatively correlated pairs of Lake Xiloá—and therefore the lack of a clear-cut parallelism at the individual species level—is likely due to the younger age of the radiation in this lake and possibly secondary contact with populations of Midas from the Great Lakes.

The empirical confirmation of the antagonistic expression of miRNA/gene pairs provides a link between genomic regulatory divergence and the presence of novel phenotypes in the crater lake radiations. Functionally interacting miRNA/target pairs are typically expressed in either temporally negatively correlated domains and/or show pronounced spatial exclusivity (Woltering and Durston 2006; Zhang et al. 2007). ISH was used to visualize the spatial expression domains of one miRNA/gene robustly negatively correlated pair validated as diverging in multiple species comparisons (i.e., *A. citrinellus* vs. *A. amarillo* and *A. citrinellus* vs. *A. zaliosus*). The spatiotemporal expression dynamics of *let-7d* are compatible with the proposed gene silencing effect of this miRNA. Interestingly, this miRNA was first identified in a study of developmental timing in *Caenorhabditis elegans* (Moss et al. 1997). Although its important function in the developmental timing of larva transitions has been elucidated in invertebrates, the role of this highly conserved miRNA remains poorly understood during vertebrate development (Yermalovich et al. 2019). Here, we showed that *let-7d* likely regulates an immunoglobulin superfamily DCC subclass member 3 gene (*igdcc3*). This gene, also known as *punc*, encodes a member of the neural cell adhesion molecule family and affects motor coordination in mice (Yang et al. 2001). One of the expression domains of *igdcc3* identified in mice embryos was in the dorsal region of the head, the same area in which *igdcc3* was expressed in our fish.

Additionally, one of the nine miRNA/gene negatively correlated pairs (*mir-250/cyp3A40*) found in the crater Lake Apoyo species pair comparison was validated using ISH. This pair was chosen for further investigation because it involves the only miRNA for which we did not find homologs in miRBase and had the highest expression fold change. The gene *cyp3A40* belongs to the cytochrome P450 (*cyp*) family, which are membrane-bound hemoproteins that play a pivotal role in the detoxification of xenobiotics, cellular metabolism, and homeostasis. The *mir-250* miRNA shows a clear expression domain in the anterior pharyngeal arches of Midas cichlids, in the region where the palatoquadrate and the ceratohyal develop (Woltering et al. 2018). The identified

cyp3A gene shows diffuse expression throughout the head region, but with a clear decrease of expression where it would otherwise overlap with the expression of *mir-250*, which is consistent with the expected downregulation for a negatively correlated miRNA/gene pair. Because the pharyngeal jaw, a modified pharyngeal arch, is a key trait underlying sympatric ecological divergence and speciation mechanisms between benthic and limnetic species (Fruciano et al. 2016), the identification of this miRNA/gene pair is remarkable. This result strengthens recent evidence showing that benthic and limnetic species have distinct transcriptional profiles even at an early stage (Fruciano et al. 2019). The identification of the *mir-250*/*cyp3A40* pair represents an important step in understanding the molecular mechanisms underlying benthic/limnetic divergence and provides a clear candidate axis of regulatory divergence in these adaptively radiating fishes.

Taken together, these results provide new avenues for investigating the role of miRNA regulation during rapid evolutionary divergence in adaptive radiations in general, and Midas cichlids in particular. There are many differentially expressed miRNA/gene pairs among the Midas species suggesting that the functional linkage among these pairs has contributed to phenotypic divergence, even in the presence of gene flow. Additionally, a number of miRNAs appears to be novel to specific lineages that would provide novel genomic substrates for adaptation and divergence, potentially leading to rapid phenotypic divergence specific to those lineages containing these miRNAs. Furthermore, there are several negatively correlated pairs that appear to have diverged in association with the benthic/limnetic axis repeatedly in the crater lake Midas cichlids, suggesting that divergence in some of these pairs could be repeatable and even predictable. Finally, the validated antagonistic temporal and spatial expression of the miRNA/gene pairs suggest that even at early stages of development miRNAs could be influencing the phenotypic divergence characterizing Midas cichlids and other adaptive radiations.

Conclusions

Posttranscriptional regulation via miRNAs likely played a role in the sympatric divergence of Nicaraguan lake Midas cichlid fishes. The identification of putative miRNA/gene pairs with negatively correlated gene expression, which is consistent with the role of miRNA in downregulating mRNA, provides a set of loci that should be investigated more extensively for their role in the phenotypic divergence of these rapidly diversifying lineages. Although these fishes have been diverging for no more than a few thousand years, we found species-specific expression domains for several miRNAs and their target genes. As a number of the miRNAs and their target sites appears to be novel to particular species, the regulatory capacity of these miRNAs should be investigated further to understand their role in generating phenotypic divergence. These results provide strong evidence that miRNA regulation can contribute to rapid evolutionary divergence even in the presence of gene flow.

Materials and Methods

Sample Collection and Genomic Library Preparation

Individuals for each of five species were used to generate both RNA sequences (RNA-Seq) and microRNA sequences (miRNA-Seq). Broods from *A. astorquii* and *A. zaliosus* (crater Lake Apoyo), *A. amarillo* and *A. sagittae* (crater Lake Xiloá), and *A. citrinellus* (Lake Nicaragua) were obtained from the University of Konstanz animal facility (TFA) and sampled at 1 dph stage. In total, 15 samples (3 samples per species) were processed as single units in downstream RNA extraction and library preparation steps as described in Franchini et al. (2016). In addition to the fifteen 1 dph samples that were used for downstream DE analyses, RNA-Seq reads from 12 individuals sampled at 1-month posthatching (1 mph) (Franchini et al. 2016) were combined with the previous samples to build a more comprehensive reference transcriptome (the data set used for this study is described in supplementary table 1, Supplementary Material online).

To isolate RNA, a FastPrep-24 homogenizer (MP Biomedicals) was first used to process each sample (30 s at 4.0 M), and then the total RNA was isolated using a Qiagen RNeasy Mini Kit. Small noncoding RNA molecules were retained by using 100% ethanol in the final washing steps. RNA quality and quantity were assessed using a Bioanalyzer 2100 (Agilent Technologies, Palo Alto, CA) and a Qubit v2.0 fluorometer (Life Technologies, Darmstadt, Germany), respectively. High-quality RNA samples (RIN value > 8) were used to construct RNA-Seq and miRNA-Seq sequencing libraries. For RNA-Seq, 400 ng of RNA for each of the 43 samples (including 1 dph and 1 mph samples) were used to construct cDNA libraries with the Illumina TruSeq RNA sample preparation kit v2 according to the manufacturer's instructions (Illumina, San Diego, CA). Paired-end sequencing was performed on an Illumina HiSeq2500 with 309 cycles (151 cycles for each paired-read and 7 cycles for the barcode sequence) at the genomic facility of the University of Tufts (TUCF Genomics, Boston, MA). For miRNA-Seq, the NEBNext Small RNA Library Prep Set (New England Biolabs, Beverly, MA) was used to construct 15 barcoded miRNA libraries (three samples for each species at 1 dph developmental stage) following the manufacturer's recommendations. Single-end sequencing with 51 cycles (46 cycles for the sample read and 5 cycles for the barcode sequence) was carried out on an Illumina HiSeq2500 at TUCF Genomics.

Transcriptome Reference Assembly

The sequencing yielded a total of 490.3 million raw RNA-Seq paired-end reads (from 81.1 to 115.1 reads per sample), each 146 bp in length. The program Trimmomatic v0.33 (Bolger et al. 2014) was used to remove adapters and to filter the reads by quality with default settings, as well as to discard sequences shorter than 50 nucleotides. Reference-guided and de novo assembly were both performed using the filtered reads of the combined samples.

For the reference assembly, we used the program Stringtie v1.0.4 (Pertea et al. 2015) with the Midas genome v5.7 (unpublished version) as reference. Read mapping was

performed using TopHat v2.0.14 (Trapnell et al. 2009) in default mode in combination with Bowtie2 v2.2.3 (Langmead and Salzberg 2012) using mostly default parameters, but allowing some flexibility (–score-min L,-0.2,-0.2). All of the studied closely related Midas species reads were mapped to the same reference genome (produced from a single wild caught *A. citrinellus* individual from Lake Nicaragua). Samtools v1.2.1 (Li et al. 2009) was used to convert the Bowtie output alignment from SAM to BAM format in order to obtain the input file for Stringtie. Transcripts were extracted from the genome using the *gffread* utility implemented in the Cufflink v2.2.1 package (Trapnell et al. 2010).

Trinity v2.06 (Grabherr et al. 2011) was used to generate a de novo assembly with the PasaFly transcript reconstruction mode, k-mer size of 32, and a minimum contig length of 200 bp. To filter out transcripts with unknown function, the obtained de novo assembly was subjected to similarity searches against a custom database containing 11 teleost fish protein data sets (Amazon molly, cavefish, cod, fugu, medaka, platyfish, spotted gar, tetraodon, tilapia, stickleback, and zebrafish) and the well-annotated human and mouse protein data sets (Ensembl release 81). BlastX algorithm (Altschul et al. 1990) with a cutoff *e*-value of 1e^{-6} was used for similarity searches. Transcripts that aligned to proteins contained in the reference databases were retained for downstream analyses.

RNA-Seq DE Analysis
The final set of 134,228 sequences obtained combining the reference and the de novo assembled transcripts was further clustered using the program Corset v1.04 (Davidson and Oshlack 2014). Corset allowed us to identify the minimum numbers of clusters in the transcript data set (each cluster resembling a gene having one or more input transcripts) and at the same time, raw read counts of the fifteen 1 dph samples were assigned to each cluster. Read mapping for each sample was performed using Bowtie2 with default parameters but allowing infinite number of alignments (-*a*) as required by Corset, setting the log likelihood ratio (-*D*) to 20,000 and retaining only the transcripts with a minimum of 20 aligned reads. DE analysis was carried out on the cluster raw reads count table using the Bioconductor DESeq2 v1.8.1 (Love et al. 2014) program in R v3.2.0 (R Core Team 2015) comparing species in a pairwise fashion. The *P* values generated by the DESeq2's exact test were further corrected to control the FDR < 0.05 using the Benjamini and Hochberg (1995) procedure.

miRNA Discovery and DE Profiling
The program FASTX-Toolkit v0.0.14 (http://hannonlab.cshl.edu/fastx_toolkit/. Last accessed July 30, 2018) was used to process the miRNA raw reads by removing adapters. The reads from all individuals were merged to identify putative species-specific miRNAs using miRDeep2.0.0.8 (Friedlander et al. 2008). In detail, the filtered reads were aligned to the Midas genome using the *mapper.pl* module implemented in the miRDeep2 package with default parameters. A set of teleost fish mature miRNAs was retrieved from miRBase v21

(Kozomara and Griffiths-Jones 2011) and used as known mature miRNAs. The miRNAs showing a miRDeep2 score ≥ 10 (a value that represents the log-odds probability of a sequence being a genuine miRNA precursor versus the probability that it is a background hairpin) were retained.

To determine whether the miRNAs sequenced for this study were either already present in other species (conserved miRNAs) or not previously described (novel miRNAs), we applied a workflow developed by Brawand et al. (2014) in a study targeting five African cichlid species. Briefly, we used the *ssearch36* module implemented in the FASTA v36.3.8g package (Pearson 2016) to compare the sequences of mature miRNAs to the mature miRNAs downloaded from miRBase v21 with *e*-value cutoffs of 0.01. Using this framework, a miRNA is defined as "conserved" when a homologous miRNA can be found in the miRBase, whereas a miRNA is considered as being "novel" when no homologous miRNAs are found in the miRBase. Clearly, "novel" in this context does not necessarily mean that the focal miRNA is only novel to the cichlids examined, as these miRNAs could be present in other clades (e.g., in species not included in miRBase).

miRNA-Seq DE Analysis
The complete set of miRNAs detected in the five Midas species and the number of reads corresponding to each of the replicates were used to build the input table for the DE analysis conducted using DESeq2 as described for the RNA-Seq data. To this end, the expression of miRNAs was quantified and normalized for each individual using the *quantifier.pl* module implemented in the miRDeep2 package. As for the RNA-Seq data set, DE analysis was carried out on the obtained miRNA count table using DESeq2 and controlling the FDR at a 5% level.

Midas 3′ UTR Detection and miRNA Binding Sites Prediction
To identify the 3′ UTRs in our transcriptome data set, the longest transcript assigned to the same cluster inferred by Corset (see above) was retained and annotated. The following workflow was used to extract the sequences of 3′ UTRs: 1) the program TransDecoder v5.5.0 (https://github.com/TransDecoder/. Last accessed October 21, 2018) was used to infer potential open reading frames in each transcript, 2) the last exons (coding regions) of annotated genes of the previously described 11 teleost fishes and 2 mammals were extracted, 3) using the exons as queries, coding sequences before stop codons were identified in the previously identified longest open reading frame of each transcript using TBlast (*e*-values cutoff 1e^{-10}, matching length > 100 bp, sequence identity $> 80\%$ and containing a stop codon within ± 9 bp from the end of the match), and 4) sequences downstream of the identified stop codons were extracted from the transcripts and designated as 3′ UTRs.

Subsequently, the Midas mature miRNAs identified in this study were used to predict miRNA binding sites on the 3′ UTRs with the target prediction tool miRanda v3.3a (Enright et al. 2004) with stringent parameters, applying a minimum alignment score of 140, minimum energy threshold

of −20.0 kcal/mol, and strict alignments in the seed region (offset positions 2–8). As miRNA target prediction is prone to false positive results in animals (Pinzon et al. 2017), we used an additional target prediction tool, TargetScan v7.0 (Agarwal et al. 2015) with default settings, on the same set of transcripts to validate the analysis conducted by miRanda.

Following target prediction, to explore the regulatory landscape of "novel" miRNAs, the genes with only "novel" miRNAs binding sites (test set) were compared with the whole Nile tilapia gene set (baseline set) to identify significantly overrepresented GO terms. To test for significance, a Fisher's exact test implemented in g:Profiler ve94_eg41_p11_36d5c99 was used (g:SCS multiple testing correction method, significance threshold of 0.05) (Reimand et al. 2016).

Identification of miRNA/Gene Negatively Correlated Pairs

Typically, mature miRNAs negatively regulate protein-coding genes, by inhibiting mRNA production or protein translation. Thus, we would predict negatively correlated expression of a miRNA and its target gene to be a good indicator of a pair's functionality. Recent studies have recovered positively correlated miRNA/gene target pairs and have described how these pairs could have a potential functional role in several complex regulatory circuits (Aukerman and Sakai 2004; Lopez-Gomollon et al. 2012; de Vries et al. 2018; Kittelmann and McGregor 2019). However, in this study, we focused on negatively correlated miRNA/gene pairs, as their role in gene expression regulation is better understood and described.

To this aim, two different approaches were used to designate miRNAs and their target genes as "robustly negatively correlated pairs" or not. In the first approach, we focused on pairwise comparisons and we defined a miRNA and its target gene as a robustly negatively correlated candidate pair when both were found to be differentially expressed in a given species pairwise comparisons (DESeq2, FDR corrected P value < 0.05), and at the same time negatively correlated in their expression levels. In the second approach, since individuals and species are not evolutionarily independent (Felsenstein 1985), we performed PIC correlations (Pearson correlation, FDR corrected P value < 0.05) using the whole data set (five species) to test the empirical support for our miRNA and mRNA robustly negatively correlated pairs. To carry out this phylogenetically informed analysis, we first constructed a phylogeny for all individuals. The phylogeny was generated from a matrix of 22,560 SNPs taken from protein-coding genes recovered from all individuals and ensuring that 95% of all SNPs could be scored in all individuals. The SNPs were retrieved using the following workflow: 1) a nonredundant set of Midas cichlid genes was obtained using a reciprocal best blast hit approach (BlastN algorithm, maximum number of hits = 10, cutoff e-value of $1e^{-9}$) using the previously retrieved Corset clusters (the longest transcript among those belonging to each Cluster was used) and the tilapia coding sequences, 2) the filtered read set of each individual was aligned to the previously obtained 4,032 nonredundant set of Midas cichlid genes using bwa v0.7.15 (Li and Durbin

2009), and 3) the resulting alignments were processed using Freebayes v1.1.0 (Garrison and Marth 2012) in order to infer individuals' genotypes that were filtered using VCFtools v0.1.15 (Danecek et al. 2011) (the full set of filters applied is reported in supplementary table 2, Supplementary Material online). Six *Archocentrus centrarchus* individuals were also genotyped and included in this analysis as outgroups to polarize the resulting phylogeny. We reconstructed the tree for all individuals accounting for incomplete lineage sorting using the coalescent-based SVDquartets program (Chifman and Kubatko 2014), implemented in PAUP v4.163 (Swofford 2001). To build the phylogeny, we evaluated all possible quartet combinations. All individuals sampled were treated as independent in the coalescent model allowing these individuals to inform the tree reconstruction and mRNA/gene pair correlations. To assess confidence in nodes recovered, we generated 100 bootstrap replicate quartet trees. Using this phylogenetic comparative framework, we performed PIC correlations in R using the "ape" package (Paradis et al. 2004) for the miRNA and mRNA pairs. This allowed us to determine for each pair whether expression levels of miRNA and mRNA were negatively correlated, while accounting for the evolutionary divergence among individuals.

Subsequently, the Fisher's exact test implemented in g:Profiler was used to identify overrepresented GO terms in the genes showing significantly negatively correlated expression with corresponding miRNAs (test set) when compared with the whole Nile tilapia gene set (baseline set). Furthermore, to identify which robustly negatively correlated miRNA/gene pairs could be related to repeated divergence in benthic and limnetic forms, we performed a series of MANOVAs. For each pair identified as functional using the correlation of PICs (i.e., with a significant negative correlation between miRNA and mRNA), we used miRNA and mRNA levels as dependent variables in the MANOVA and the benthic or limnetic habitat occupied as a categorical predictor. To account for multiple tests, we controlled for the FDR at the 5% level using the binomial SGoF approach (Carvajal-Rodriguez et al. 2009).

Identification of Expression Domains by ISH

To further characterize the expression dynamics of the miRNA/target pairs identified, we performed whole mount ISH. These experiments allowed us to obtain insight into the morphological structures these miRNA/target pairs might pattern during the process of body formation. For the ISH experiments, we focused on two robustly negatively correlated pairs detected in species pair comparisons. The first pair involved divergence between *A. citrinellus* from the great Lake Nicaragua and all the species from the crater lakes, and the second involved sympatric crater lake species. This design allowed us to identify expression domains of miRNA and genes that putatively could be associated with adaptation to crater lake environment and to sympatric speciation. Critically, the second robustly negatively correlated pair was the only pair among those identified within Lake Apoyo involving a miRNA with no homolog in miRBase (see Results). For the ISH, embryos of Midas species were first fixed at

hatching day and at 1 dph overnight in 4% paraformaldehyde (PFA) in phosphate suffered saline (PBS) at 4 °C and subsequently stored in methanol at −20 °C. ISH was carried out according to Woltering and Durston (2008) and Woltering et al. (2009). Dual DIG labeled LNA probes were ordered from Exiqon (Vedbaek, Denmark) with the following sequences: dre-let7d, /5DIGN/AACCATACAACCAACTACCTCA/ 3DIG_N/; midas_mir250, /5DIGN/GCGTCTGCCGCCCTC AGTCTGCT/3DIG_N/. In the ISH procedure, these were used at 10 pmol/ml hybmix and the following hybridization temperatures were used: dre-let7d at 62 °C and midas_mir250 at 65 °C. Probes for *igdcc3* and *cyp3A40* were cloned in the pGEMT vector (promega A-3600) from *A. citrinellus* genomic DNA using the following primers: *igdcc3*, FW-CACTGGTCAATGATGGAGTGCATC, RV-GCAAAGCTTGA ACAGACGGTTG; *cyp3A40*, FW- GTTGTCAGAGCTGAC ATTCTG, RV- CTTTACTTCTCTCAAACCAGAC. DIG labeled probes were synthesized and purified using standard protocols and used at a concentration of 100 ng/ml hybmix as previously described (Woltering et al. 2009).

Acknowledgments

Author Contributions

References

Agarwal V, Bell GW, Nam JW, Bartel DP. 2015. Predicting effective microRNA target sites in mammalian mRNAs. *Elife* 4.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215(3):403–410.

Aukerman MJ, Sakai H. 2004. Regulation of flowering time and floral organ identity by a microRNA and its APETALA2-like target genes (vol 15, pg 2730, 2003). *Plant Cell* 16:555.

Barluenga M, Stolting KN, Salzburger W, Muschick M, Meyer A. 2006. Sympatric speciation in Nicaraguan crater lake cichlid fish. *Nature* 439(7077):719–723.

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol.* 57(1):289–300.

Bird CE, Fernandez-Silva I, Skillings DJ, Toonen RJ. 2012. Sympatric speciation in the post "modern synthesis" era of evolutionary biology. *Evol Biol.* 39(2):158–180.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30(15):2114–2120.

Bolnick DI, Fitzpatrick BM. 2007. Sympatric speciation: models and empirical evidence. *Annu Rev Ecol Evol Syst.* 38(1):459–487.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. *Nature* 513(7518):375–381.

Britten RJ, Davidson EH. 1969. Gene regulation for higher cells—a theory. *Science* 165(3891):349.

Carroll SB. 2008. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134(1):25–36.

Carvajal-Rodriguez A, de Una-Alvarez J, Rolan-Alvarez E. 2009. A new multitest correction (SGoF) that increases its statistical power when increasing the number of tests. *BMC Bioinformatics* 10:209.

Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitx1 enhancer. *Science* 327(5963):302–305.

Chen JF, Mandel EM, Thomson JM, Wu QL, Callis TE, Hammond SM, Conlon FL, Wang DZ. 2006. The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet.* 38(2):228–233.

Chifman J, Kubatko L. 2014. Quartet inference from SNP data under the coalescent model. *Bioinformatics* 30(23):3317–3324.

Coyne JA. 2007. Sympatric speciation. *Curr Biol.* 17(18):R787–R788.

Coyne JA, Hoekstra HE. 2007. Evolution of protein expression: new genes for a new diet. *Curr Biol.* 17(23):R1014–R1016.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. 2011. The variant call format and VCFtools. *Bioinformatics* 27(15):2156–2158.

Davidson JH, Balakrishnan CN. 2016. Gene regulatory evolution during speciation in a songbird. *G3 (Bethesda)* 6:1357–1364.

Davidson NM, Oshlack A. 2014. Corset: enabling differential gene expression analysis for de novo assembled transcriptomes. *Genome Biol.* 15:410.

de Vries S, Kukuk A, von Dahlen JK, Schnake A, Kloesges T, Rose LE. 2018. Expression profiling across wild and cultivated tomatoes supports the relevance of early miR482/2118 suppression for *Phytophthora* resistance. *Proc R Soc B Biol Sci.* 285(1873):pii: 20172560.

Dieckmann U, Doebeli M. 1999. On the origin of species by sympatric speciation. *Nature* 400(6742):354–357.

Elmer KR, Fan S, Kusche H, Luise Spreitzer M, Kautt AF, Franchini P, Meyer A. 2014. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. *Nat Commun.* 5(1):5168.

Elmer KR, Kusche H, Lehtonen TK, Meyer A. 2010. Local variation and parallel evolution: morphological and genetic diversity across a species complex of neotropical crater lake cichlid fishes. *Philos Trans R Soc B* 365(1547):1763–1782.

Elmer KR, Lehtonen TK, Kautt AF, Harrod C, Meyer A. 2010. Rapid sympatric ecological differentiation of crater lake cichlid fishes within historic times. *BMC Biol.* 8(1):60.

Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS. 2004. MicroRNA targets in *Drosophila. Genome Biol.* 5(1):R1.

Feder JL, Nosil P. 2010. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. *Evolution* 64(6):1729–1747.

Felsenstein J. 1985. Phylogenies and the comparative method. *Am Nat.* 125(1):1–15.

Flaxman SM, Wacholder AC, Feder JL, Nosil P. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol Ecol.* 23(16):4074–4088.

Foote AD. 2018. Sympatric speciation in the genomic era. *Trends Ecol Evol (Amst).* 33(2):85–95.

Franchini P, Fruciano C, Spreitzer ML, Jones JC, Elmer KR, Henning F, Meyer A. 2014. Genomic architecture of ecologically divergent body shape in a pair of sympatric crater lake cichlid fishes. *Mol Ecol*. 23(7):1828–1845.

Franchini P, Xiong P, Fruciano C, Meyer A. 2016. The role of microRNAs in the repeated parallel diversification of lineages of Midas cichlid fish from Nicaragua. *Genome Biol Evol*. 8(5):1543–1555.

Friedlander MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N. 2008. Discovering microRNAs from deep sequencing data using miRDeep. *Nat Biotechnol*. 26(4):407–415.

Friedman RC, Farh KK, Burge CB, Bartel DP. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res*. 19(1):92–105.

Fruciano C, Franchini P, Kovacova V, Elmer KR, Henning F, Meyer A. 2016. Genetic linkage of distinct adaptive traits in sympatrically speciating crater lake cichlid fish. *Nat Commun*. 7(1):12736.

Fruciano C, Meyer A, Franchini P. 2019. Divergent allometric trajectories in gene expression and coexpression produce species differences in sympatrically speciating Midas cichlid fish. *Genome Biol Evol*. 11(6):1644–1657.

Garrison E, Marth G. 2012. Haplotype-based variant detection from short-read sequencing. Preprint at arXiv: 1207.3907v2 [q-bio.GN].

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng QD, et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol*. 29(7):644–652.

Guerra-Assuncao JA, Enright AJ. 2012. Large-scale analysis of microRNA evolution. *BMC Genomics*. 13:218.

Ha M, Kim VN. 2014. Regulation of microRNA biogenesis. *Nat Rev Mol Cell Biol*. 15(8):509–524.

Henning F, Meyer A. 2014. The evolutionary genomics of cichlid fishes: explosive speciation and adaptation in the postgenomic era. *Annu Rev Genom Hum Genet*. 15:417–441.

Irisarri I, Singh P, Koblmuller S, Torres-Dowdall J, Henning F, Franchini P, Fischer C, Lemmon AR, Lemmon EM, Thallinger GG, et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. *Nat Commun*. 9(1):3159.

Jiggins CD. 2006. Sympatric speciation: why the controversy? *Curr Biol*. 16(9):R333–R334.

Kautt AF, Machado-Schiaffino G, Meyer A. 2016. Multispecies outcomes of sympatric speciation after admixture with the source population in two radiations of Nicaraguan crater lake cichlids. *PLoS Genet*. 12(6):e1006157.

Kittelmann S, McGregor AP. 2019. Modulation and evolution of animal development through microRNA regulation of gene expression. *Genes* 10(4):pii: E321.

Kozomara A, Griffiths-Jones S. 2011. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res*. 39(Database):D152–D157.

Kutterolf S, Freundt A, Perez W, Wehrmann H, Schmincke HU. 2007. Late Pleistocene to Holocene temporal succession and magnitudes of highly-explosive volcanic eruptions in west-central Nicaragua. *J Volcanol Geotherm Res*. 163(1-4):55–82.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 9(4):357–359.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 25(14):1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Proc G. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25(16):2078–2079.

Li KX, Wang LY, Knisbacher BA, Xu QQ, Levanon EY, Wang HH, Frenkel-Morgenstern M, Tagore S, Fang XD, Bazak L, et al. 2016. Transcriptome, genetic editing, and microRNA divergence substantiate sympatric speciation of blind mole rat, Spalax. *Proc Natl Acad Sci U S A*. 113(27):7584–7589.

Liem KF. 1973. Evolutionary strategies and morphological innovations: cichlid pharyngeal jaws. *Syst Biol*. 22:425–441.

Loh YHE, Yi SV, Streelman JT. 2011. Evolution of microRNAs and the diversification of species. *Genome Biol Evol*. 3:55–65.

Lopez-Gomollon S, Mohorianu I, Szittya G, Moulton V, Dalmay T. 2012. Diverse correlation patterns between microRNAs and their targets during tomato fruit development indicates different modes of microRNA actions. *Planta* 236(6):1875–1887.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 15(12):550.

Lu C, Jeong DH, Kulkarni K, Pillay M, Nobuta K, German R, Thatcher SR, Maher C, Zhang L, Ware D, et al. 2008. Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). *Proc Natl Acad Sci U S A*. 105(12):4951–4956.

Mack KL, Campbell P, Nachman MW. 2016. Gene regulation and speciation in house mice. *Genome Res*. 26(4):451–461.

Malinsky M, Svardal H, Tyers AM, Miska EA, Genner MJ, Turner GF, Durbin R. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. *Nat Ecol Evol*. 2(12):1940–1955.

Mayr E. 1963. Animal species and evolution. Animal species and their evolution. Cambridge, MA: Harvard University Press.

Moss EG, Lee RC, Ambros V. 1997. The cold shock domain protein LIN-28 controls developmental timing in *C. elegans* and is regulated by the lin-4 RNA. *Cell* 88(5):637–646.

Ozsolak F, Milos PM. 2011. RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet*. 12(2):87–98.

Papadopulos AST, Kaye M, Devaux C, Hipperson H, Lighten J, Dunning LT, Hutton I, Baker WJ, Butlin RK, Savolainen V. 2014. Evaluation of genetic isolation within an island flora reveals unusually widespread local adaptation and supports sympatric speciation. *Philos Trans R Soc B Biol Sci*. 369(1648):pii: 20130342.

Paradis E, Claude J, Strimmer K. 2004. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 20(2):289–290.

Pavey SA, Collin H, Nosil P, Rogers SM. 2010. The role of gene expression in ecological speciation. *Ann N Y Acad Sci*. 1206:110–129.

Pearson WR. 2016. Finding protein and nucleotide similarities with FASTA. *Curr Protoc Bioinformatics* 53:3.9.1–3.9.25.

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. 2015. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 33(3):290.

Pinzon N, Li B, Martinez L, Sergeeva A, Presumey J, Apparailly F, Seitz H. 2017. microRNA target prediction programs predict many false positives. *Genome Res*. 27(2):234–245.

R Core Team. 2015. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: http://www.R-project.org/

Rastorguev SM, Nedoluzhko AV, Gruzdeva NM, Boulygina ES, Sharko FS, Ibragimova AS, Tsygankova SV, Artemov AV, Skryabin KG, Prokhortchouk EB. 2017. Differential miRNA expression in the three-spined stickleback, response to environmental changes. *Sci Rep*. 7(1):18089.

Ravinet M, Faria R, Butlin RK, Galindo J, Bierne N, Rafajlovic M, Noor MAF, Mehlig B, Westram AM. 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J Evol Biol*. 30(8):1450–1477.

Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, Vilo J. 2016. g: profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res*. 44(W1):W83–W89.

Seehausen O, Butlin RK, Keller I, Wagner CE, Boughman JW, Hohenlohe PA, Peichel CL, Saetre G-P, Bank C, Brännström A, et al. 2014. Genomics and the origin of species. *Nat Rev Genet*. 15(3):176–192.

Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jonsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* 428(6984):717–723.

Spielmann M, Mundlos S. 2016. Looking beyond the genes: the role of non-coding variants in human disease. *Hum Mol Genet*. 25(R2):R157–R165.

Stefani G, Slack FJ. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol.* 9(3):219–230.

Swofford D. 2001. Phylogenetic analysis using parsimony (*and other methods), version 4.0a142. Sunderland (MA): Sinauer.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105–1111.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 28(5):511–515.

Verheyen E, Salzburger W, Snoeks J, Meyer A. 2003. Origin of the superflock of cichlid fishes from Lake Victoria. *Science* 300(5617):325–329.

Via S. 2001. Sympatric speciation in animals: the ugly duckling grows up. *Trends Ecol Evol.* 16(7):381–390.

Wahid F, Shehzad A, Khan T, Kim YY. 2010. MicroRNAs: synthesis, mechanism, function, and recent clinical trials. *Biochim Biophys Acta Mol Cell Res.* 1803(11):1231–1243.

Wittkopp PJ, Williams BL, Selegue JE, Carroll SB. 2003. *Drosophila* pigmentation evolution: divergent genotypes underlying convergent phenotypes. *Proc Natl Acad Sci U S A.* 100(4):1808–1813.

Woltering JM, Durston AJ. 2006. The zebrafish hoxDb cluster has been reduced to a single microRNA. *Nat Genet.* 38(6):601.

Woltering JM, Durston AJ. 2008. MiR-10 represses HoxB1a and HoxB3a in zebrafish. *PLoS One* 3(1):e1396.

Woltering JM, Holzem M, Schneider RF, Nanos V, Meyer A. 2018. The skeletal ontogeny of *Astatotilapia burtoni*—a direct-developing model system for the evolution and development of the teleost body plan. *BMC Dev Biol.* 18(1):8.

Woltering JM, Vonk FJ, Muller H, Bardine N, Tuduce IL, de Bakker MAG, Knochel W, Sirbu IO, Durston AJ, Richardson MK. 2009. Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Dev Biol.* 332(1):82–89.

Xiong PW, Hulsey CD, Meyer A, Franchini P. 2018. Evolutionary divergence of 3′ UTRs in cichlid fishes. *BMC Genomics.* 19(1):433.

Yang W, Li CY, Mansour SL. 2001. Impaired motor coordination in mice that lack punc. *Mol Cell Biol.* 21(17):6031–6043.

Yermalovich AV, Osborne JK, Sousa P, Han A, Kinney MA, Chen MJ, Robinton DA, Montie H, Pearson DS, Wilson SB, et al. 2019. Lin28 and let-7 regulate the timing of cessation of murine nephrogenesis. *Nat Commun.* 10(1):168.

Zhang YJ, Woltering JM, Verbeek FJ. 2007. Screen of microRNA targets in zebrafish using heterogeneous data sources: a case study for Dre-miR-10 and Dre-miR-196. Proceedings of World Academy of Science, Engineering and Technology, Vol 26, Parts 1 and 2; December 2007. p. 258.