Title

Evolutionary Rate Shifts in Coding and Regulatory Regions Underpin Repeated Adaptation to Sulfidic Streams in Poeciliid Fishes.

Journal

Genome Biology and Evolution, 16(5)

Authors

De-Kayne, Rishi

Perry, Blair

McGowan, Kerry

et al.

Publication Date

2024-05-02

DOI

10.1093/gbe/evae087

Peer reviewed

Evolutionary Rate Shifts in Coding and Regulatory Regions Underpin Repeated Adaptation to Sulfidic Streams in Poeciliid Fishes

Rishi De-Kayne [1,*], Blair W. Perry[2], Kerry L. McGowan[2], Jake Landers[2], Lenin Arias-Rodriguez[3], Ryan Greenway[4], Carlos M. Rodríguez Peña[5], Michael Tobler[6,7,8], and Joanna L. Kelley [1,*]

[1]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95060, USA
[2]School of Biological Sciences, Washington State University, Pullman, WA 99164, USA
[3]División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco (UJAT), Villahermosa, México
[4]Division of Biology, Kansas State University, Manhattan, KS 66506, USA
[5]Instituto de Investigaciones Botánicas y Zoológicas, Universidad Autónoma de Santo Domingo, Santo Domingo 10105, Dominican Republic
[6]Department of Biology, University of Missouri–St. Louis, St. Louis, MO 63131, USA
[7]Whitney R. Harris World Ecology Center, University of Missouri–St. Louis, St. Louis, MO 63121, USA
[8]WildCare Institute, Saint Louis Zoo, St. Louis, MO 63110, USA

*Corresponding authors: E-mails: rdekayne@ucsc.edu; jokelley@ucsc.edu

Accepted: April 13, 2024

Abstract

Adaptation to extreme environments often involves the evolution of dramatic physiological changes. To better understand how organisms evolve these complex phenotypic changes, the repeatability and predictability of evolution, and possible constraints on adapting to an extreme environment, it is important to understand how adaptive variation has evolved. Poeciliid fishes represent a particularly fruitful study system for investigations of adaptation to extreme environments due to their repeated colonization of toxic hydrogen sulfide–rich springs across multiple species within the clade. Previous investigations have highlighted changes in the physiology and gene expression in specific species that are thought to facilitate adaptation to hydrogen sulfide–rich springs. However, the presence of adaptive nucleotide variation in coding and regulatory regions and the degree to which convergent evolution has shaped the genomic regions underpinning sulfide tolerance across taxa are unknown. By sampling across seven independent lineages in which nonsulfidic lineages have colonized and adapted to sulfide springs, we reveal signatures of shared evolutionary rate shifts across the genome. We found evidence of genes, promoters, and putative enhancer regions associated with both increased and decreased convergent evolutionary rate shifts in hydrogen sulfide–adapted lineages. Our analysis highlights convergent evolutionary rate shifts in sulfidic lineages associated with the modulation of endogenous hydrogen sulfide production and hydrogen sulfide detoxification. We also found that regions with shifted evolutionary rates in sulfide spring fishes more often exhibited convergent shifts in either the coding region or the regulatory sequence of a given gene, rather than both.

Key words: adaptation, convergent evolution, hydrogen sulfide, Poeciliidae, extremophile.

Introduction

Across the tree of life, organisms have evolved a wide variety of strategies to facilitate the colonization of new environments. While some species occupy and persist in new ecological niches through gradual or iterative phenotypic changes across environmental gradients, the survival and persistence of species in more extreme and stressful environments often involves dramatic phenotypic shifts across

Significance

While organisms have adapted to live in even the most extreme environments, the specific regions involved in adaptation and the degree to which similar adaptive changes are present across taxa are rarely understood. In our study, we identify convergent evolutionary rate shifts associated with sulfide adaptation across seven independent sulfidic and nonsulfidic contrasts within the Poeciliidae family. We show that both coding and regulatory regions exhibit signals of convergence in evolutionary rate variation in sulfidic lineages across Poeciliidae, including evolutionary rate shifts in genomic regions associated with sulfide tolerance. Additionally, evolutionary rate shifts typically involve either coding sequence or regulatory regions, rather than both, highlighting the need for genome-wide investigations into the genomic basis of convergence.

multiple traits. In these cases, environments impose strong selection since maladaptation confers highly reduced fitness or even death. To adapt to extreme environments, organisms often evolve an array of highly specialized phenotypes. In fishes alone, a number of remarkable adaptive traits facilitate the persistence of species in extreme conditions (Riesch 2015), including the evolution of antifreeze proteins in cold-adapted lineages (Davies et al. 2002), the loss of hemoglobin in oxygen-rich environments (Sidell and O'Brien 2006), and even the evolution of air-breathing behavior in hypoxic environments (Damsgaard et al. 2020). To determine how organisms adapt to extreme environments, it is essential to better understand the mechanisms by which these specialized ecological strategies evolve. With the ever-growing power of DNA sequencing and analysis approaches, it is now feasible to investigate the aspects of genome-wide variation that play a role in conferring adaptation (Xu et al. 2020). Understanding the links between genome-wide variation and adaptation can inform us about how phenotypic diversity arises and may reveal the potential for evolution in an increasingly changing environment, particularly with regard to anthropogenically altered ecosystems.

Clades in which multiple lineages have independently evolved to thrive in extreme environmental conditions are especially valuable for studying the process of adaptation. The repeated ecological and phenotypic contrasts within such systems facilitate investigations into the mechanisms underpinning adaptation by allowing us to determine the degree of convergence across instances of adaptation (Conte et al. 2012). Patterns of convergence can, in turn, inform us about the repeatability and predictability of evolution. Convergent evolution occurs where multiple independent lineages evolve similar traits in response to similar sources of selection and has been noted across the tree of life (Losos 2011). Convergence may be particularly pronounced in response to extreme stressors, which impose strong selection on populations and may constrain adaptation to a set of key genes or biochemical pathways that alter specific aspects of physiology (Greenway et al. 2020; Xu et al. 2020). Convergence can be identified across levels of biological organization ranging from shared coding sequence through to shared phenotypes. Empirical studies have highlighted that instances of convergence appear to be more common at the phenotype level, where independent lineages have similar phenotypic traits (Rosenblum et al. 2014; Härer et al. 2018; Ishikawa et al. 2022; van der Zee et al. 2022), and are rarer at the coding sequence level, where observing convergent single-nucleotide polymorphism (SNP) allele sharing across lineages, for example, remains uncommon (Tenaillon et al. 2012). Several examples of identical nucleotide substitutions have been identified across taxa, including fishes (Sugawara et al. 2002), insects (Dobler et al. 2012), and marine mammals (Foote et al. 2015), but these cases appear to be the exception rather than the rule, even in highly conserved genes where strong constraints on mutations may be expected (Natarajan et al. 2016). In addition, many studies that aim to assess convergent nucleotide variation associated with convergent phenotypic shifts still focus on understanding sequence variation in coding regions only, despite a growing number of studies that highlight the importance of convergence in regulatory regions in underpinning the repeated evolution of key traits (Sackton et al. 2019; Artur and Kajala 2021). This is perhaps unsurprising given that the fitness consequences of nucleotide changes in coding sequences are likely more deleterious than in noncoding regulatory sequences (Hoekstra and Coyne 2007), and as a result, substantially more variation can accumulate in noncoding, including regulatory regions and subsequently be a target of selection.

Assessing patterns of convergent evolution in both coding and regulatory regions is therefore critical to understanding the repeated evolution of key phenotypes. Approaches that assess evolutionary rate shifts across regions of the genome in a phylogenetically explicit context, including those implemented in RERconverge (Kowalczyk et al. 2019), provide valuable opportunities for understanding convergence in sequence changes across complex clades (Fig. 1a). These approaches can help reveal the ways in which selection shapes genomic variation by informing us about which genomic regions are evolving

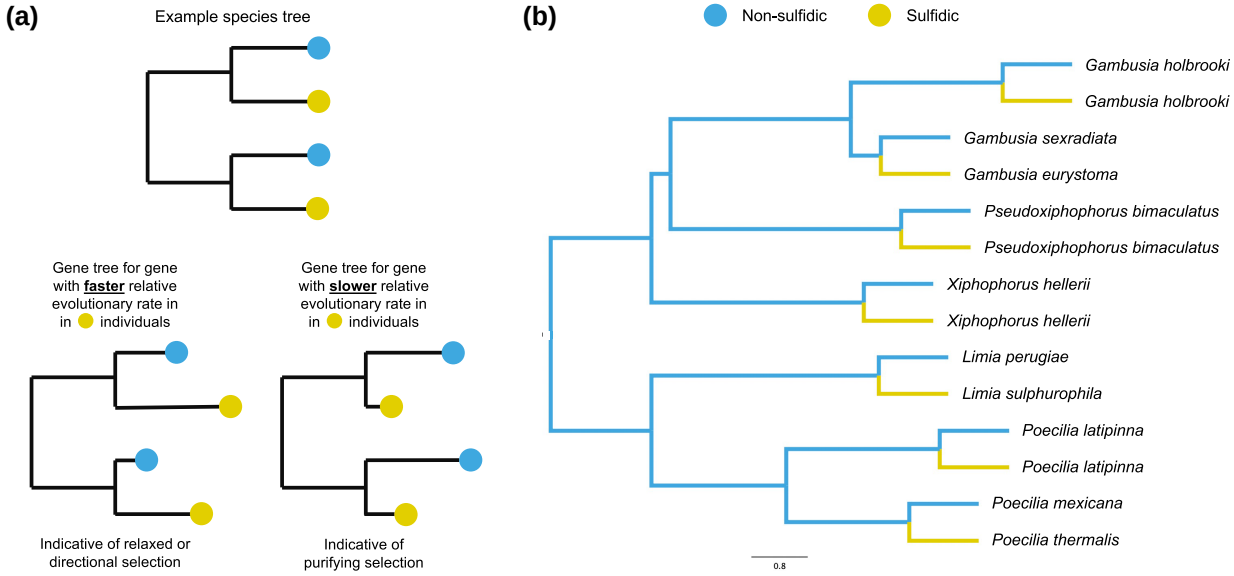

Fig. 1.—a) A schematic illustrating our analysis of convergence in relative evolutionary rate (RER). Top: example species tree. Lower left: Cartoon gene tree for a gene with a faster RER in yellow lineages, indicative of relaxed or directional selection on the yellow phenotype, where branches to individuals with the yellow phenotype within the gene tree are relatively longer. Lower right: Cartoon gene tree with a slower RER in yellow lineages, indicative of purifying selection on the yellow phenotype, where branches to individuals with the yellow phenotype are relatively shorter. b) The consensus species tree for the seven nonsulfidic–sulfidic pairs of individuals sampled in this study. The sampling spans 10 species across 5 genera, *Poecilia*, *Limia*, *Xiphophorus*, *Pseudoxiphophorus*, and *Gambusia*. Nonsulfidic branches and tips are highlighted in blue and sulfidic in yellow. All nodes were supported by 100/100 bootstrap replicates.

faster or slower than expected and the degree to which evolutionary rate shifts are correlated with specific traits across lineages. Associating variation in evolutionary rates with the evolution of convergent phenotypic traits in this way has been successfully implemented in a number of systems and has identified evolutionary rate shifts across specific regions of the genome associated with the evolution of extreme lifespans in fishes (Kolora et al. 2021), the loss of flight in birds (Sackton et al. 2019), and hairlessness in mammals (Kowalczyk et al. 2022). This approach can be used to reveal patterns of convergence not observed across other biological levels of organization. By comparing convergent evolutionary rate shifts across different genomic regions that are associated with the evolution of specific convergent adaptive phenotypes, we can also better understand the relative roles of coding and regulatory sequence evolution in adaptation.

Fishes in the family Poeciliidae provide a unique opportunity to investigate genomic convergence in response to extreme environmental conditions. Several poeciliid species have independently colonized springs rich in hydrogen sulfide (H_2S), and these sulfidic populations have subsequently diverged from nearby nonsulfidic populations, despite a lack of physical barriers that would prevent fish movement (Tobler et al. 2008; Barts et al. 2018; Greenway et al. 2020). While H_2S is endogenously produced by organisms and acts as an important signaling molecule at low concentrations (Kimura et al. 2010), it is also a potent environmental toxicant that is lethal at elevated concentrations. H_2S is damaging in part because it easily diffuses across lipid membranes and inhibits cytochrome *c* oxidase (complex IV) in the electron transport chain, with severe, and even lethal, impacts to respiration (Mathai et al. 2009; Jiang et al. 2016). H_2S also leads to the formation of reactive oxygen species and reactive sulfur species that negatively impact the same mitochondrial enzymes affected by H_2S (Giles and Jacob 2002; Vaquer-Sunyer and Duarte 2010). In addition to metabolic damage, H_2S also has indirect effects as a result of its affinity to dissolved oxygen, which it readily binds to within sulfidic springs, creating severe hypoxia, as well as a host of other potentially stressful abiotic and biotic environmental factors that are strongly correlated with H_2S within sulfide springs (Tobler et al. 2018).

The repeated evolution of H_2S adaptation in the Poeciliidae family has been previously investigated across multiple levels of biological organization. These investigations have highlighted evidence for convergence in morphology, physiology, life history, and behavior (Tobler et al. 2011; Barts et al. 2018; Greenway et al. 2020). Additionally, at the molecular level there has been substantial evidence for convergence in gene expression across poeciliid populations and species (Kelley et al. 2016; Greenway et al. 2020, 2023). These studies did not, however, observe substantial sharing of nucleotide variation across sulfidic populations. Studies have also identified signatures of positive selection on mitochondrial genes, including convergence of new

variants arising in sulfidic lineages (Pfenninger et al. 2014). More recently, sequence similarity in the coding sequence of different *Poecilia mexicana* lineages has been revealed, which likely stems from selection on standing genetic variation or migration (Brown et al. 2019; Ryan et al. 2023). Despite this previous body of work, the degree to which convergent evolutionary shifts may be observed across more distantly related H_2S-tolerant species within the Poeciliidae family (i.e. where sulfidic lineages show accelerated or decelerated evolution compared to nonsulfidic lineages), the role of genes that have a more indirect role in adaptation to sulfide spring environments, and the role of variation in regulatory regions in underpinning H_2S tolerance, remains poorly understood.

In this study, we identified signatures of convergent evolutionary rate shifts associated with sulfide tolerance across multiple species within the Poeciliidae family. We achieved this by generating and collating genomic data from seven lineages in which sulfide adaptation has independently evolved and where contemporary populations persist in nearby sulfidic and nonsulfidic streams. We calculated convergent relative evolutionary rate (RER) shifts across all seven sulfidic and nonsulfidic contrasts in a phylogenetically explicit context and across coding sequence and regulatory regions. Using RERconverge, we then correlated RER shifts across the phylogeny using a binary trait value indicating whether lineages were sulfidic or nonsulfidic. The resulting values, rho, for genes and regulatory regions reflected the degree to which evolutionary rate shifts correlated with a lineage being sulfidic rather than nonsulfidic. This approach allowed us to identify convergent evolutionary rate shifts that occurred across most sulfidic lineages, but not necessarily all seven. Regions with significant negative rho values represent instances of reduced evolutionary rates in sulfidic lineages, indicating purifying selection, whereas regions with significant positive rho values represent instances of accelerated evolution in sulfidic lineages, indicative of genomic regions under relaxed or directional selection (Fig. 1a; Kowalczyk et al. 2019). Since the correlation between RER and trait is carried out across our phylogeny, genomic regions with significant rho values have rate shifts across sulfidic lineages compared to nonsulfidic lineages. By calculating evolutionary rate shifts in regulatory regions in addition to coding regions, we were able to test whether selection in extreme environments primarily acts on regions of the genome that likely alter gene expression or on coding regions directly. We identified whether specific gene families and functional groups were significantly enriched across genomic regions with substantial evolutionary rate shifts concurrent with sulfide adaptation across species, for example, those with a direct role in detoxification, as may be expected during the evolution of H_2S tolerance.

Results

Genomic Variation and Species Relatedness Across the Poeciliidae Family

After filtering, a total of 40,919,842 SNPs remained across the 14-sample dataset, which included one sulfidic and one nonsulfidic individual from each of seven lineages (supplementary table S1, Supplementary Material online; all retained loci were biallelic across the dataset, were filtered for quality, and had no missing data across all individuals). These SNPs were used to produce a species tree and investigate RER of regulatory regions. A filtered variant file excluding 4-fold degenerate (4D) sites was used for our analysis of convergent evolutionary rate shifts in coding regions and contained 1,484,493 SNPs. To establish the evolutionary relationships between Poeciliid species in our dataset and to confirm that sulfidic and nonsulfidic populations from each species were sister lineages, we produced a consensus phylogenetic tree using SNP data from all 21,991 *Xiphophorus hellerii* genes that had polymorphic sites across our dataset (supplementary table S1, Supplementary Material online; Fig. 1b). This tree confirmed that nonsulfidic and sulfidic individuals from each species pair are indeed each other's closest relatives, as previously shown (Greenway et al. 2020). Across our species tree, *Limia* and *Poecilia* species were grouped into one monophyletic clade and *Xiphophorus*, *Gambusia*, and *Pseudoxiphophorus* into a second monophyletic clade, which is consistent with family-level analyses (Rodríguez-Machado et al. 2024). Subsequently, sulfidic and nonsulfidic individuals were differentiated from one another supporting previous suggestions that sulfide tolerance has evolved independently in each lineage (which is further supported by the presence of numerous additional nonsulfidic lineages within the Poeciliidae family not sampled here; Tobler et al. 2018).

Analysis of RER Shifts

To determine whether regions of the genome exhibited convergent RER shifts in sulfidic lineages compared to nonsulfidic lineages, we ran RERconverge (Kowalczyk et al. 2019) on coding, promoter, and enhancer regions of the genome, which resulted in a metric (rho) that represents the correlation between an evolutionary rate shift and the trait of interest (in our case whether a lineage is sulfidic or not; Fig. 2a and b; supplementary figs. S1 and S2, Supplementary Material online). This analysis identified regions of the genome that may be under relaxed or directional selection in sulfidic lineages, indicated by positive convergent RER shifts across sulfidic lineages compared to nonsulfidic lineages (significant + rho), as well as regions under purifying selection, indicated by negative convergent RER shifts across sulfidic lineages compared to nonsulfidic lineages (significant − rho).

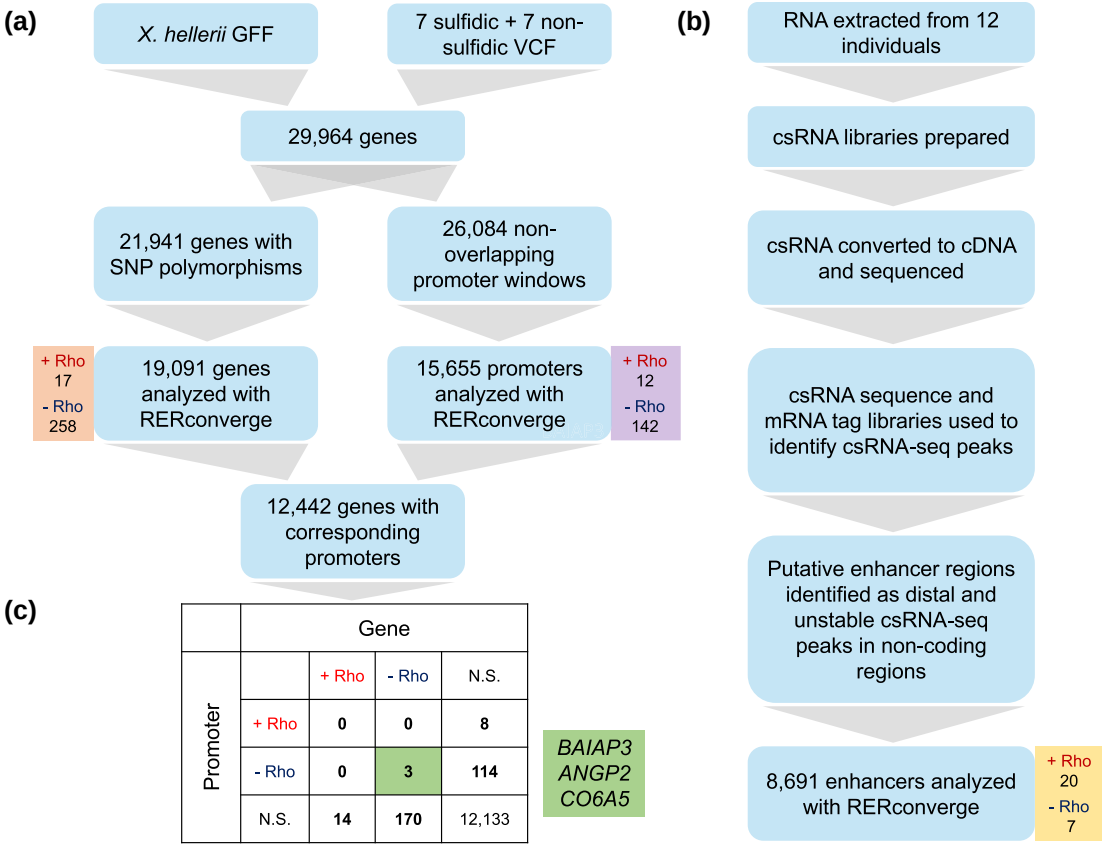

Fig. 2.—a) The pipeline for our RERconverge analysis of genes and promoter regions and the number of genes and promoter regions exhibiting significant faster (+rho) and slower (−rho) convergent evolutionary rate shifts across sulfidic lineages. b) Our capped-small RNA-seq analysis pipeline which detected transcriptional start sites for stable and unstable RNAs allowing us to identify putative enhancer regions, which were then analyzed for convergent evolutionary rate shifts. c) A summary of RER results for genes where both coding sequence and promoter regions were analyzed with RERconverge highlighting the number of instances of both, or either, coding sequence and promoter regions evolving faster or slower than expected in sulfidic lineages. The three genes where both coding sequence and promoter sequence were shown to be evolving significantly slower than expected in sulfidic lineages compared to nonsulfidic lineages are highlighted.

Coding Regions

In total, 19,091 genes met minimum variation thresholds and were tested for the evidence of convergent evolutionary rate shifts in coding regions, resulting in a range of rho values from 0.47 for the gene carbamoyl phosphate synthase (*CPSM*) (Table 1; Fig. 3a; XM_032568860.1; permuted *P*-value = 0.01, original *P* = 0.005) to −0.60 for the gene centrosomal protein of 131 kDa (Table 1; Fig. 3b; XM_032574854.1; permuted *P*-value = 0.02, original *P* = 0.0004). Across our dataset, 17 genes exhibited significant positive convergent RER shifts in coding regions, indicative of a faster evolutionary rate in sulfidic lineages, and 258 significant negative convergent RER shifts, indicative of a slower evolutionary rate (supplementary table S2, Supplementary Material online). While previous work has highlighted two particular genes that likely underpin the evolution of sulfide tolerance, including sulfide:quinone oxidoreductase (*SQOR*) and ETHE1 persulfide dioxygenase (*ETHE1*; Ryan et al. 2023), neither of these genes exhibited significant convergent evolutionary rate shifts within coding regions across our dataset (*SQOR* rho = −0.08, two orthologues of *ETHE1* rho = 0.10 and 0.05).

Promoter Regions

We also analyzed convergent evolutionary rate shifts across regulatory regions, including promoters and putative enhancers (Fig. 2a and b). For this analysis, we delineated promoter regions as the 500 base pairs (bp) upstream of each gene in the *X. hellerii* reference genome (GCA_003331165.2). In a similar manner to the analysis of convergent RER shifts across genes, a total 15,655 promoter regions met minimum variation thresholds and were analyzed using RERconverge. Rho values for promoters ranged from 0.49 for the promoter region associated with the gene ubiquitin carboxyl-terminal hydrolase (*CYLD*; Table 1; genomic window NC_045685.1:7491049-7491549; permuted

Table 1
The top three significant fastest and slowest RER genes and promoters and their corresponding RER rho

Genes			
	Gene name (gene symbol)	**UniProt number**	**RER (rho)**
Faster RER in sulfidic lineages	Carbamoyl phosphate synthase (*CPSM*)	P31327	0.47*
	Centrosomal protein of 128 kDa (*CE128*)	Q6ZU80	0.46*
	Anoctamin-7 (*ANO7*)	Q6IWH7	0.45*
Slower RER in sulfidic lineages	Centrosomal protein of 131 kDa (*CP131*)	Q9UPN4	−0.6*
	Enteropeptidase (*ENTK*)	P98073	−0.57**
	Ras-specific guanine nucleotide-releasing factor (*RGPS2*)	Q86X27	−0.53**
Promoters			
	Associated gene name (gene symbol)	**UniProt number**	**RER (rho)**
Faster RER in sulfidic lineages	Ubiquitin carboxyl-terminal hydrolase (*CYLD*)	Q9NQC7	0.49*
	F-box/LRR-repeat protein 7 (*FBXL7*)	Q9UJT9	0.42*
	N-alpha-acetyltransferase 38 (*LSMD1*)	Q9BRA0	0.41**
Slower RER in sulfidic lineages	Zona pellucida sperm-binding protein 2 (*ZP2*)	Q05996	−0.58*
	Furin (*FURIN*)	P09958	−0.49*
	Ornithine decarboxylase (*DCOR*)	P11926	−0.49*

Corresponding permuted *P*-values for each rho value are indicated by an asterisk: $*P < 0.05$, $**P < 0.01$.

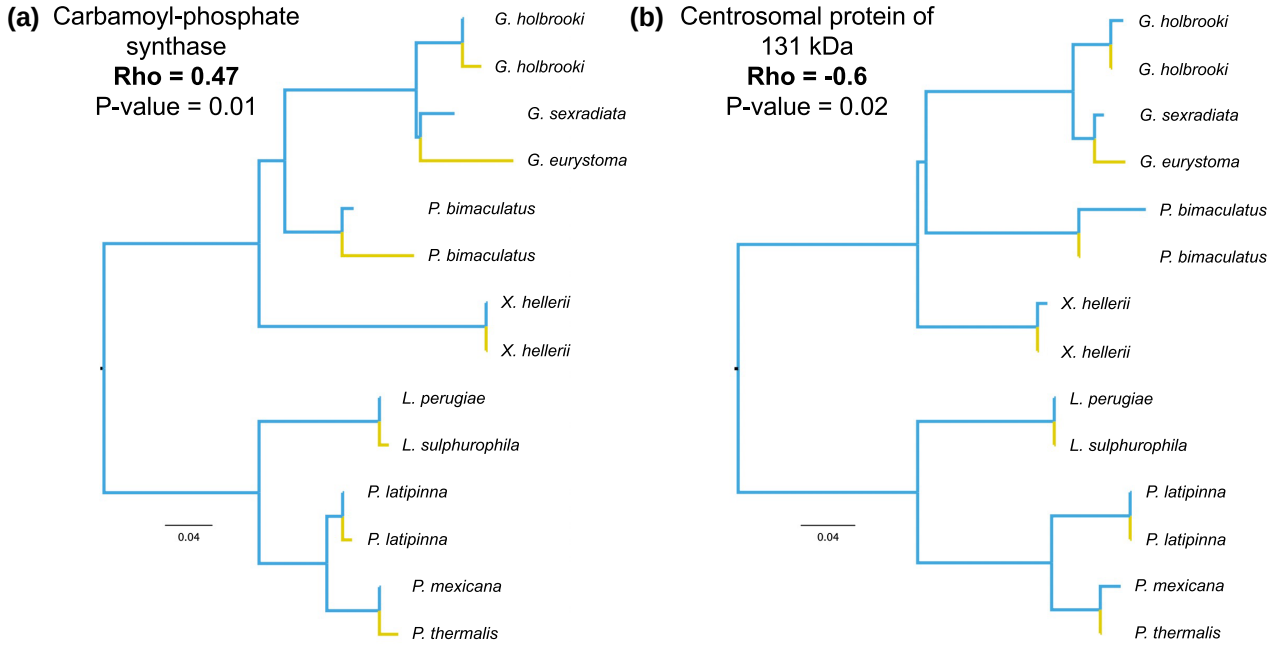

Fig. 3.—Gene trees with branch lengths reflecting evolutionary rates a) the gene *CPSM* which our analysis showed to be the most positive RER-shifted gene, evolving faster than expected and possibly reflecting relaxed or directional selection in sulfidic lineages, and b) the gene centrosomal protein of 131 kDa which was the most negatively RER-shifted gene evolving much slower than expected and possibly reflecting purifying selection in sulfidic populations across Poeciliidae.

P-value = 0.023, original *P*-value = 0.004) to −0.58 for gene zona pellucida sperm-binding protein 2 (*ZP2*; Table 1; genomic window NC_045686.1:7935884-7936384; permuted *P*-value = 0.031, original *P* = 0.0007). Across promoters, 9 were evolving significantly faster than expected in sulfidic lineages and 178 slower than expected (supplementary table S2, Supplementary Material online).

Enhancer Regions

Putative enhancer regions were identified using capped-small RNA sequencing (csRNA-seq), which captures short, nascent transcripts, including enhancer RNAs that are produced by active, accessible enhancer regions (Duttke et al. 2019; Perry et al., in review). Overall, 8,691 putative enhancer regions met the minimum variation

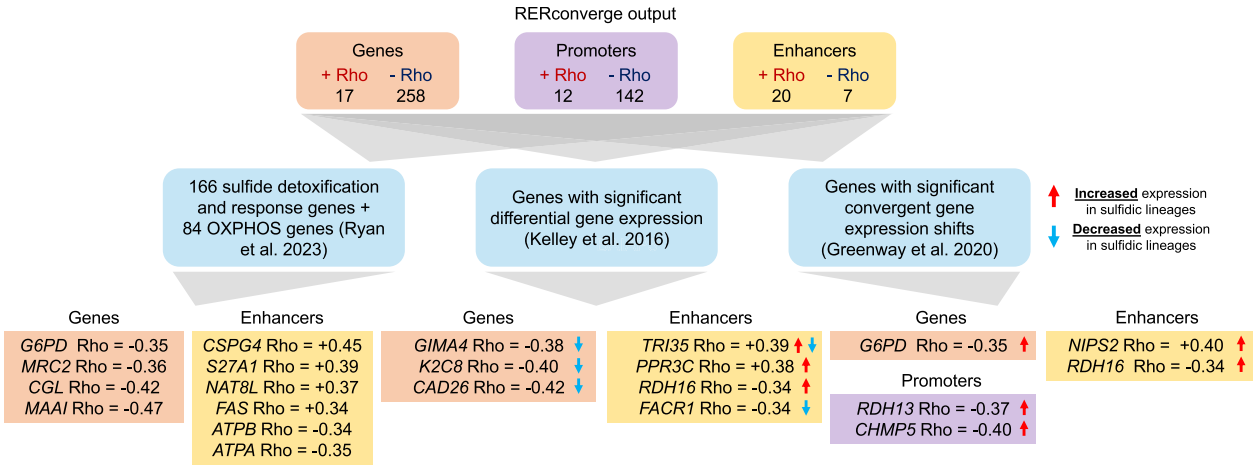

Fig. 4.—Genes and enhancers with significant RER shifts are associated with genes known to play key roles in sulfide detoxification and response, and OxPhos (Ryan et al. 2023). They also overlap with genes previously identified to have differential patterns of gene expression in sulfidic lineages (Kelley et al. 2016) and genes previously identified to have convergent shifts in gene expression in sulfidic lineages (Greenway et al. 2020). Genes and enhancer RER values highlight whether the given region has a convergent evolutionary rate shift faster (positive values) or slower (negative values) than expected. Arrows indicate whether differentially expressed genes that overlap with gene and enhancer RER shifts had increased (red up arrow) or decreased (blue down arrow) expression in sulfidic lineages.

threshold and were analyzed with RERconverge, resulting in a range of RERs from 0.45 (permuted *P*-value = 0.0078; original *P*-value = 0.0077) to −0.37 (permuted *P*-value = 0.031, original *P*-value = 0.029). In total, 20 enhancer regions had a significant positive RER shift and seven had a significant negative RER shift (supplementary table S2, Supplementary Material online). To assess potential target genes of these rate-shifted enhancers, we surveyed annotated genes with a transcription start site within 100 kilobase pairs (kbp) of each significant enhancer region. Two coding regions with significant RER shifts in sulfidic lineages are located within 100 kbp of two rate-shifted enhancer genes, *NB5R3* (coding region rho = 0.41 and coding region permuted *P* = 0.047, enhancer region rho = −0.34 and enhancer region permuted *P* = 0.046) and *FA13A* (coding region rho = 0.40 and coding region permuted *P* = 0.031, enhancer region rho = 0.34 and enhancer region permuted *P* = 0.046), while no rate-shifted promoters were within 100 kbp of a rate-shifted enhancer region.

Gene Ontology and Gene Set Enrichment of Genomic Regions with Significant RER Shifts

Of the 19,091 genes analyzed, 18,701 had confidently assigned human gene orthologues, allowing robust tests of gene ontologies (GOs) using gene set enrichment analysis (GSEA). Promoters with confident gene assignment were also tested for enrichment and included 13,182 of the original 15,655 promoter set. The GSEA for coding regions showed a significant enrichment for a number of terms, including response to interferon-beta (False discovery rate [FDR]-corrected *P*-value = 0.028, enrichment score = 0.68; biological process), endosomal sorting complexes required for transport (ESCRT) complex (FDR-corrected *P*-value = 0.045, enrichment score = −0.62; cellular component), SH2 domain binding (FDR-corrected *P*-value = 0.021, enrichment score = −0.60; molecular function), and telomeric DNA binding (FDR-corrected *P*-value = 0.030, enrichment score = −0.57; molecular function). Promoter regions with significant RER shifts were not enriched for any GO terms.

Overlap in Significant RERs for Genes and Promoters

Of the 19,091 genes analyzed, 12,442 had a corresponding promoter region that had also been analyzed using RERconverge (Fig. 2a and c; supplementary table S2, Supplementary Material online). By comparing the rho across significant RER genes and their corresponding promoters, we identified three instances where both a gene and its promoter region had significant convergent RER shifts in sulfidic lineages: BAI1-associated protein 3, angiopoietin-2, and collagen alpha-5(VI) chain (Fig. 2c; supplementary fig. S3, Supplementary Material online). All three involved slower RERs in both the coding sequence and promoter. For the remaining 122 promoter regions with a significant RER shift across sulfidic lineages, the corresponding gene did not have any significant RER shifts. Similarly, 184 genes had significant RER shifts, but their promoter did not.

Comparing Significant RER Shifts with Genes with Roles in H_2S Adaptation

While evolutionary rate shifts were identified genome-wide, without the constraints of a limited set of a priori

focal genes, post hoc characterization of rate-shifted loci using sets of genes previously shown or predicted to play a role in H_2S adaptation provides valuable context for our results. We identified rate-shifted regions that may play a direct role in sulfide detoxification and response or oxidative phosphorylation (OxPhos, the direct toxicity target of H_2S in mitochondria), using a previously published list of focal genes for sulfide adaptation in this system (Ryan et al. 2023). Interestingly, we did not find evidence for an enrichment of significant RER-shifted genes in this group of candidate genes (supplementary table S3, Supplementary Material online), suggesting that convergent rate shifts are not broadly characteristic of sulfide detoxification and OxPhos genes as a whole. However, significant RER shifts were identified for four of the candidate sulfide detoxification and response genes: glucose-6-phosphate dehydrogenase (*G6PD*) (rho = −0.35), mannose receptor C type 2 (*MRC2*) (rho = −0.36), cystathionine gamma-lyase (*CGL/CSE*) (rho = −0.42), and glutathione S-transferase zeta 1 (*GSTZ1/MAAI*) (rho = −0.47; Fig. 4; supplementary table S4, Supplementary Material online). All four significant RERs shifts were observed in coding regions (as opposed to promoter regions), and each rho value was negative, indicating that these genes are evolving more slowly than expected in sulfide-adapted lineages. An additional six genes from this candidate set are each located within 100 kbp of a significantly rate-shifted enhancers (supplementary table S4, Supplementary Material online). In contrast to the RER values of genes, which were all negative, enhancers associated with these genes had a mixture of positive RER shifts (Fig. 4; *CSPG4* rho = 0.45, *S27A1* rho = 0.39, *NAT8L* rho = 0.37, and *FAS* rho = 0.34) and negative RER shifts (ATP synthase subunit beta [*ATPB*] rho = −0.34 and ATP synthase subunit alpha [*ATPA*] rho = −0.35; supplementary table S4, Supplementary Material online).

Two previous studies identified genes with shared patterns of differential expression in multiple independent sulfide-adapted lineages in the Poeciliidae family, suggesting important roles in adaptation to H_2S-rich environments (Kelley et al. 2016; Greenway et al. 2020). While genes implicated by these two previous studies were overall not statistically enriched for convergent RER shifts (supplementary table S3, Supplementary Material online), significant convergent rate shifts were identified for a subset of these genes. Three rate-shifted genes, *GIMA4* (rho = −0.38), *K2C8* (rho = −0.40), and *CAD26* (rho = −0.42), were previously found to have decreased expression in all tested sulfidic drainages compared to their corresponding nonsulfidic drainage in Kelley et al. (2016) (Fig. 4; supplementary table S4, Supplementary Material online). Four additional genes found to be differentially expressed in this study were located within 100 kbp of enhancers with significant RER shifts, indicating a possible link between enhancer sequence evolution and changes to gene expression: *TRI35* (rho = 0.39), *PPR3C* (rho = 0.38), *RDH16* (rho = −0.34), and *FACR1* (rho = −0.34; Fig. 4). Of genes found to exhibit convergent patterns of gene expression shifts between sulfidic and nonsulfidic species in Greenway et al. (2020), we identified significant evolutionary rate shifts in the protein-coding sequence of one gene, *G6PD* (rho = −0.35) and in the promoters of retinol dehydrogenase 13 (*RDH13*) (rho = −0.37) and *CHMP5* (rho = −0.40). Two additional RER-shifted genes from this study, *NIPS2* (rho = 0.40) and *RDH16* (rho = −0.34) were located within 100 kbp of enhancers with significant RER shifts (Fig. 4).

Discussion

Determining how specific aspects of phenotypic variation have evolved is particularly important for understanding how organisms have adapted to live, and even thrive, in extreme environments. Adapting to extreme environments often involves the evolution of specialized adaptive phenotypic changes due to the strong selection pressures imposed by extreme environments, where maladaptation results in highly reduced fitness or even death. Although elucidating the basis of simple phenotypes underpinned by one or few specific genomic variants can be relatively straightforward, revealing the basis of complex multidimensional aspects of species variation remains challenging. One reason for this challenge is the hierarchical and interconnected nature of variation across levels of biological organization, the ways in which genomic variation, gene expression differences, and phenotypic variation are connected and affect each other. For example, phenotypic variation can be directly affected by genomic variation in coding regions, but also by the modulation of gene expression arising from sequence changes in regulatory regions (Stern and Orgogozo 2008; Albert and Kruglyak 2015). As a result, analyzing patterns of evolutionary change in both coding and regulatory regions is necessary for determining how adaptation proceeds. Previous work has highlighted convergence in phenotype (Tobler and Hastings 2011), gene expression (Greenway et al. 2020), and coding sequence of a select set of genes (Ryan et al. 2023) involved in sulfide adaptation across the Poeciliidae family. Despite this substantial body of previous work in the Poeciliid system, our understanding of how evolutionary rate shifts and sequence variation across the genome underpins convergent trait evolution remains lacking. Examining relative rates of evolution associated with protein-coding genes and regulatory regions across poeciliid fish lineages that have colonized H_2S-rich springs allowed us to address this knowledge gap by identifying genomic convergence that has driven repeated adaptation to an extreme environment.

Sulfidic and Nonsulfidic Populations of the Same Species Complex Are Sister Lineages

Previous studies have highlighted substantial genomic differentiation and reproductive isolation between sulfidic and nonsulfidic Poeciliid species despite there being no physical barriers to gene flow (Plath et al. 2013; Riesch et al. 2016). Our results support these previous findings in that sulfide adaptation and the establishment of reproductive isolation appears to have occurred following speciation, as different genera and species are separated by substantial levels of genomic differentiation across the family (Fig. 1b; Pfenninger et al. 2014; Brown et al. 2019; Greenway et al. 2020, 2023). Our unrooted tree contained two main clades, one containing three species of *Poecilia* and both *Limia* species and the other containing all three species of *Gambusia* as well as *X. hellerii* and *Pseudoxiphophorus bimaculatus*. Our sampling of nonsulfidic and sulfidic lineages across the Poeciliidae family allowed us to identify convergent patterns of evolution associated with multiple instances of nonsulfidic lineages evolving sulfide tolerance.

Coding and Enhancer Regions with Significant RER Shifts Were Related to Genes Involved in Sulfide Detoxification and Response and OxPhos

In total, 275 genes (1.44% of the genes investigated) and 154 promoter regions (0.98% of the promoters investigated) had significant convergent deviations in RER in sulfidic lineages. These evolutionary rate changes can potentially be attributed to selection acting on genomic regions, resulting in either an accelerated or repressed rate of mutation accumulation. Contextualizing these patterns of convergent evolutionary rate changes in sulfidic lineages is particularly complex due to the paradoxical nature of H_2S for Poeciliids. In addition to its potent properties as a toxicant that inhibits cellular respiration (Mathai et al. 2009; Jiang et al. 2016), H_2S remains an important endogenously produced cell-signaling molecule (Kimura et al. 2010). Rather than imposing a solely negative impact on organisms, where we might expect strong directional selection toward novel mechanisms that confer sulfide tolerance, biochemical mechanisms that modulate cellular processing and response to H_2S must remain active and functional, making the expected patterns of evolution and genomic variation associated with extreme levels of H_2S in sulfidic streams less clear. Regions of the genome that have necessarily accumulated adaptive changes to mitigate the toxic effects of H_2S may, therefore, be more likely to exhibit signs of increased RER, while regions underpinning core aspects of H_2S-associated cellular signaling and functioning may be expected to have reduced RER, since mutation accumulation in these regions would likely have a maladaptive deleterious effect. Across coding, promoter, and enhancer regions, far fewer genomic regions had significant positive RERs compared to negative RERs (17 positive compared to 258 negative in genes, 14 positive compared to 142 negative in promoter regions, and 7 positive compared to 20 negative in enhancer regions), indicating that a greater proportion of genomic regions may be under purifying selection than relaxed or positive, directional selection in the sulfidic lineages versus nonsulfidic lineages. This bias toward negative RER shifts may also reflect the fact that genetic variation that is neutral or nearly neutral in less stressful environments may rapidly become costly upon colonization of extreme environments.

Our analysis of RERs highlighted a number of genes with dramatic shifts in evolutionary rate in sulfidic lineages, a subset of which have been previously implicated in H_2S adaptation in Poeciliidae. Four genes with significantly reduced RER shifts overlapped with a list of candidate sulfide and OxPhos genes that were recently investigated in a study of selection on standing variation in *P. mexicana* (Ryan et al. 2023; Fig. 4). These genes including *G6PD*, *MRC2*, *CGL/CSE*, and *GSTZ1/MAAI*. The reduction in evolutionary rate in these genes across sulfidic Poeciliids is perhaps unsurprising given that some of these genes play key roles in producing and moderating cellular levels of H_2S for core processes. Due to the increased concentration of environmental H_2S, the tight regulation of intracellular H_2S production is likely critical to prevent substantial interference to cellular signaling. *CGL/CSE* is one of a number of enzymes involved in the production of endogenous H_2S within cells (Wang et al. 2013). In contrast, *G6PD* and *GSTZ1/MAAI* are both involved in the detoxification of H_2S experienced by sulfidic lineages (Olson 2018). We also found six enhancer regions that were within 100 kbp of genes in this set of sulfide and OxPhos candidate genes. Genes associated with these enhancers included *ATPA* and *ATPB*, which are both associated with OxPhos as a result of their key role in ATP synthesis and ion transport (Junge and Nelson 2015).

In addition to genes that have direct functions in mitigating sulfide toxicity, our analysis also identified genes with evolutionary rate shifts that have not been previously implicated in sulfide tolerance. The gene with the most dramatic evolutionary rate increase was CPSM, a gene that has been shown to play a key role in the detoxification of nitrogenous waste and highlighted as a target of selection in other extremophile fishes (White et al. 2020). The gene enteropeptidase was shown to have convergent reductions in RER in sulfidic fishes and has been shown to be associated with differences in digestion and diet in other fish species (Jiao et al. 2023). These genes may play important roles in sulfide adaptation due changes to the biotic environment that are also present in H_2S-rich streams, including dramatic shifts in diet to sulfide bacteria and invertebrates compared to algae in nonsulfidic environments (Tobler et al. 2015). Future work

should utilize population-level whole-genome data to differentiate the different types of selection shaping variation in evolutionary rate, for example, differentiating instances of positive selection from relaxed selection.

Coding, Promoter, and Enhancer Regions with Significant RER Shifts Overlapped with Genes with Known Expression Shifts in Sulfidic Poeciliids

Several regions with RER shifts were associated with genes previously shown to exhibit convergent shifts in gene expression in sulfidic lineages from Kelley et al. (2016) and Greenway et al. (2020; Fig. 4). We found that both coding region RER shifts and enhancer region shifts overlapped with genes found to be differentially expressed in the same direction across multiple comparisons of sulfidic to nonsulfidic lineages in Kelley et al. (2016). This included three coding regions with significant negative RER shifts that previously showed decreased expression in sulfidic lineages. Additionally, four enhancer regions that showed both faster (2 of 4) and slower (2 of 4) RER were within 100 kbp of a total of four genes with differential expression. Our comparison of regions with evolutionary rate shifts and genes with gene expression variation from Greenway et al. (2020) also highlighted a number of overlapping genes, promoters, and enhancers (Fig. 4). This overlap also included the gene *G6PD*, discussed above, as well as *RDH13*, a gene thought to play an important role in alleviating oxidative stress, which may be particularly prevalent across sulfidic lineages (Belyaeva et al. 2008). This overlap of genes, promoters, and enhancers implicated by our study and previous results support the fact that evolutionary shifts across sulfidic lineages are likely having an impact on sulfide tolerance both directly, through changes to coding sequence and protein structure, and through potential adaptive changes in gene expression that facilitate the persistence of Poeciliids in toxic H_2S-rich streams. The absence of a clear link between previously identified differentially expressed genes and the remaining promoters and enhancers analyzed may suggest that gene expression is moderated by many sequence-level changes, many of which may not be convergent across poeciliid species.

Genes and Their Associated Promoters Rarely Both Exhibit Significant Relative Evolutionary Shifts

Our results show that when a coding region exhibited a RER shift, the corresponding promoter likely did not, and vice versa. Across the 12,442 analyzed genes with corresponding promoter regions also analyzed, we found only three cases (0.02%) where both were RER outliers. These three cases involved the genes BAI1-associated protein 3, angiopoietin-2, and collagen alpha-5(VI) chain and their associated promoters. More commonly, only the protein-coding sequence or promoter of a given gene was significant. This may suggest that sulfide tolerance has involved convergent evolutionary rate shifts in coding sequence or convergent evolutionary rate shifts in regulatory regions, but rarely both. However, it is possible that for a given gene, there is a substantial degree of genetic redundancy in sulfide adaptation, such that some taxa have evolved coding sequence changes associated with some adaptive variation, while others have achieved similar phenotypic variation via promoter evolution and gene expression changes. Our approach is unable to detect these signals of partial or weak convergence. Thus, while we were able to investigate both coding regions and regulatory regions, undoubtedly some aspects of gene expression regulation may be mitigated by other regions of the genome, and as such, we may have missed other instances where both changes to coding sequence and noncoding sequence impacted adaptive variation relating to the same gene.

Conclusion

In conclusion, the repeated adaptation of Poeciliid fishes to sulfide streams has involved a number of genes known to play a role in mitigating the toxic effects of H_2S and in the endogenous production of H_2S, many of which have evolved slower than expected in sulfidic lineages, likely as a result of purifying selection stemming from the maintenance of these core cellular processes. Adaptation to extreme environments including hydrogen sulfide–rich streams involves changes to a variety of complex traits. The degree of convergence across taxa that have adapted to the same environment is likely highly variable and involves convergent evolutionary rate shifts in coding sequence, regulatory sequence, and in rare cases both.

Materials and Methods

Sample Collection

Poeciliid fishes were collected from 7 sulfidic and 7 nearby nonsulfidic sites in several locations in the Neotropics, including springs in Florida, USA; Tabasco and Chiapas, Mexico; and Independencia, Dominican Republic (supplementary table S1, Supplementary Material online; Brown et al. 2019; Greenway et al. 2020). Poeciliids included in the study spanned *Gambusia* spp., *Limia* spp., *Poecilia* spp., *Pseudoxiphophorus* spp., and *Xiphophorus* spp. (supplementary table S1, Supplementary Material online). Gill tissues were collected from sacrificed fishes, preserved in RNAlater (Invitrogen), and stored at −80°C. All experiments were approved by the Institutional Animal Care and Use Committee of Kansas State University (ACUP #3473).

Library Preparation and Sequencing

This sequencing data were combined with existing data (Brown et al. 2019) spanning two species from the genus Poecilia (*P. mexicana* and *P. thermalis*), resulting in a dataset of 14 samples, 7 sulfidic, and 7 nonsulfidic, spanning a total of 10 species (supplementary table S1, Supplementary Material online). For this study, genomic DNA was extracted from gill tissues from 12 samples using Qiagen's DNeasy Blood and Tissue Kit following all standard protocols. DNA was quantified using the Qubit 2.0 dsRNA HS Assay Kit on a Qubit fluorometer. Whole-genome libraries were prepared using IDT xGen DNA EZ library preparation kits as per the manufacturer's protocols (Integrated DNA Technologies). Briefly, DNA was fragmented, followed by end repair and A-tailing. An SPRI bead cleanup was used to purify the resulting DNA fragments. Unique dual-indexed adapters were ligated to the DNA. Libraries were quantified with a Qubit fluorometer and pooled at equimolar concentrations. The resulting pool was quantified using qPCR and the NEBNext Library Quant Kit for Illumina on a Lightcycler. Paired-end 150 bp reads were sequenced to 20× coverage for all samples on a single S4 flow cell using an Illumina NovaSeq 6000.

Mapping, Genotyping, and Filtering

All reads were mapped to the *X. hellerii* genome (https://www.ncbi.nlm.nih.gov/datasets/taxonomy/8084/; GCA_003331165.2) using bwa-mem2 (v2.2.1). Picard (v2.27.1; http://broadinstitute.github.io/picard/) and Sambamba (v0.8.2; Tarasov et al. 2015) were then used to fix read mate information, mark duplicates, sort, and index the resulting bam files. The mean genome-wide coverage of each bam file was then calculated using Mosdepth (v0.3.4; Pedersen and Quinlan 2018). Genotyping was carried out on each chromosome separately using BCFtools mpileup and bcftools call (v1.17; Li, 2011; Danecek et al. 2021). Genotype calls were then filtered using BCFtools filter for minimum depth of 7 and a minimum genotype quality score of 30. Genotype calls from different chromosomes were then combined using BCFtools concat. To reduce the potential for reference bias due to alleles being called in only the reference species *X. hellerii*, the resulting full genotypes were then filtered to exclude sites with missing data, i.e. excluding loci with <28 alleles across the 14 diploid individuals, and to retain only biallelic SNPs, resulting in a full VCF file with 40,919,842 SNPs. For our analysis of RER shifts in coding regions, we also produced a VCF file with 4D sites (those sites at which polymorphisms cannot lead to changes in amino acids) removed. To identify the level of degeneracy of each site, the script codingsitetypes.py (from https://github.com/simonhmartin/genomics_general) was run specifying the *X. hellerii* GFF file (GCF_003331165.1_Xiphophorus_hellerii-4.1_genomic.gff) and our full VCF. The resulting file, which contained each site and the degree of degeneracy, was then filtered to remove all sites labeled as 4D and the full VCF filtered to retain only these sites, leaving 1,484,493 SNPs.

Delineating Genes and Promoter Regions for Analysis

Genes for the analysis were identified using the publicly available *X. hellerii* genome annotation (GCF_003331165.1_Xiphophorus_hellerii-4.1_genomic.gff; https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003331165.1/). AGAT (v0.8.1; Dainat) was then used to produce a statistical report for the gff (agat_sq_stat_basic.pl), remove C_gene and V_gene features (agat_sp_filter_feature_from_kill_list.pl), and select only the longest isoform from each gene (agat_sp_keep_longest_isoform.pl). Genes that were not situated on chromosome-scale scaffolds were removed, leaving 21,941 genes. Promoter regions were identified across the full *X. hellerii* genome annotation and defined as the 500 bp preceding the start of the gene feature (accounting for the orientation of the gene). Any promoter regions which overlapped were removed leaving 26,084 promoter windows.

Delineating Enhancer Regions for Analysis

To get a more comprehensive understanding of regions of the genome involved in mediating expression differences rather than just coding regions, we aimed to identify whether convergent evolutionary rate shifts also occurred at enhancer sites as well as coding and promoter regions. To achieve this, we made use of csRNA-seq, an approach developed to identify transcription start sites for both stable RNA and unstable RNA (Duttke et al. 2019), providing a more detailed overview of noncoding regions that may underpin sulfide adaptation and as a result exhibit RER shifts as a result of being targets of selection.

Sample Collection for csRNA-seq

Adult *P. mexicana* and *Poecilia sulphuraria* individuals were collected via seine net from both sulfidic and nonsulfidic sites in the Pichucalco, Puyacatengo, and Tacotalpa drainages of the Río Grijalva basin in southern Mexico (all females, six total sites, $N = 2$ per site giving $N = 12$ total; supplementary table S1, Supplementary Material online) as part of a previous study (Kelley et al. 2016). Gill arches were dissected immediately following humane sacrifice, preserved in RNALater (Invitrogen, Inc.), and stored at −80 °C.

csRNA Library Preparation and Sequencing

Total RNA was extracted from gill tissue using the Qiagen miRNeasy Mini Kit following standard protocols. Total RNA concentrations were assessed with both the Qubit RNA HS Assay Kit and the Agilent 2100 BioAnalyzer using the RNA 6000 Nano Kit. Short RNAs (~20–60 nucleotides) were size-selected with gel electrophoresis. A 10% subsample of size-selected total RNA was reserved per sample to serve as the "input library" to control for exonic contamination when calling csRNA-seq peaks (Duttke et al. 2019). csRNA libraries were then generated by isolating 5′ 7-methylguanosine-capped RNA from the remaining total RNA following protocols described in Duttke et al. (2019). csRNA and input libraries were converted into cDNA libraries, amplified with PCR, size-selected to remove primer dimers, and purified with gel electrophoresis. Libraries were then pooled and sequenced on an Illumina NextSeq 500 using 75 bp single-end reads.

csRNA-seq Data Processing and Peak Calling

csRNA-seq data were processed using the HOMER csRNA-seq analysis pipeline (Heinz et al. 2010). The trim tool within homerTools (v4.11) was used to remove Illumina TruSeq adapters from the 3′ end of reads, discard reads shorter than 20 bp, and trim ends of reads with a Phred score under 20. We indexed the *X. hellerii* reference genome (NCBI: GCF_003331165.1; https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_003331165.1/) using STAR v2.7.6a and aligned trimmed reads to the reference with STAR, outputting a single primary alignment per read (Dobin et al. 2013). Tag directories were created for each mapped csRNA and input library using homerTools makeTagDirectory with the –fragLength flag set to 30 bp to reflect the average read length. A third tag directory was created for each sample using mapped mRNA libraries (described below). The findcsRNATSS.pl script from homerTools was then used with flag -ntagThreshold 7 to call csRNA-seq peaks using csRNA, input, and mRNA tag directories for each sample. Peaks were annotated using the GTF file for *X. hellerii* (NCBI: GCF_003331165.1). The mergePeaks tool in homerTools was used to create merge overlapping peaks between samples, and the annotatePeaks.pl tool in homerTools was used to annotate and quantify per-sample read counts for the resulting merged peak set. csRNA-seq peaks that were identified in only one sample were excluded in subsequent analyses.

Identification of Putative Enhancer Regions

Putative enhancer peak regions were defined as csRNA-seq peaks located in noncoding regions and annotated as both "distal" (i.e. >500 bp from the nearest gene transcription start site) and "unstable" by annotatePeaks.pl. Any overlapping features were merged, and the resulting putative enhancer regions were then used as input for RERconverge analyses as described below. Genes were considered candidate targets of an enhancer if their transcription start site was located within ±100 kb of a putative enhancer region.

mRNA Data Processing

mRNA-seq libraries were prepared and sequenced as part of a previously published study (Kelley et al., 2016; $N = 34$; supplementary table S1, Supplementary Material online). Raw reads were downloaded from NCBI (SRA: PRJNA290391) and trimmed using TrimGalore (from Cutadapt v2.7; Martin 2011) with flags –quality 24, –stringency 5, –illumina, –length 50, –clip_R1 12, and –clip_R2 12. Trimmed mRNA-seq reads were mapped to the *X. hellerii* reference using STAR. Tag directories for mRNA samples were created using homerTools makeTagDirectory with the –fragLength flag set to 150 bp and used in csRNA-seq peak calling (described above).

Gene Trees and Species Tree

To produce a species tree, we produced gene alignments using our full VCF file and the coding sequences of genes using the python script extractCDSAlignments.py from https://github.com/simonhmartin/genomics_general, with all sites not called across our 14 samples replaced by Ns to reduce the risk of reference bias. Individual gene trees were then produced from all gene alignments that contained at least 1 polymorphic site using RAxML as part of the ParGenes pipeline (v1.2.0; Morel et al. 2019) before a consensus ASTRAL tree was produced (Mirarab et al. 2014). The resulting tree was plotted using FigTree (v1.4.4; Rambaut 2012) and a version with branch lengths and bootstraps removed as a master tree topology for the RERconverge analysis. Genome-window trees for use in our RERconverge analysis were produced for three different datasets which were analyzed separately, (i) with coding regions, (ii) promoter regions, and (iii) enhancer regions. For analysis of coding regions, our VCF with 4D sites removed was used following the same extractCDSAlignments.py approach to produce gene-specific sequence alignments. Sequence alignments for promoter and enhancer regions were produced by first filtering the full VCF for each promoter or enhancer window, then converting these window-specific VCF files to sequence alignments using the scripts parseVCF.py, filterGenotypes.py, and genoToSeq.py (from https://github.com/simonhmartin/genomics_general). All three sets of genome-window alignments were filtered for the number of variant sites, with only windows containing ≥15 polymorphic sites being retained, leaving 19,091

gene regions, 15,655 promoter regions, and 8,691 enhancer regions. Across these three datasets, gene trees were then generated from each multispecies nucleotide alignment using the Phangorn (v2.11.1; Schliep 2011) function *estimatePhangornTreeAll* in R (Development Core Team 2011) specifying the master species tree topology generated using ParGenes and ASTRAL.

RERconverge

We examined the relative rates of evolution across the seven nonsulfidic and sulfidic lineages in a phylogenetically explicit context to identify genes that were evolving more quickly or more slowly than expected in sulfidic lineages. To establish a null, permuted distribution of background RERs that would allow us to identify significant outlier genes we ran RERconverge (v0.3.0) 128 times in total, each time with a permuted set of foreground, test branches. In each permutation the phylogenetic relationship between sulfidic and nonsulfidic pairs was maintained by only swapping sulfidic and nonsulfidic sister tips. In addition to our "true" test set where all sulfidic lineages were set as the foreground branches this resulted in 127 background RER values for each tested region. To calculate our empirical/permuted *P*-value from this distribution of values, we used the equation $S + 1/N + 1$, where S represents the number of permuted *P*-values as, or more, significant than our "true" test statistic and N represents the total number of tests, i.e. 128. The RERconverge function correlateWithBinaryPhenotype was used to calculate rho, the correlation between the RER shifts observed across lineages and the binary trait value of those lineages, in our analysis this denoted whether a lineage is sulfidic or not. This allowed us to detect genomic regions where most, but not necessarily all, sulfidic lineages had a shifted RERs. This analysis highlighted genomic regions where sulfidic lineages had significantly accelerated (significant positive rho) or significantly decelerated (significant negative rho) RERs compared to nonsulfidic lineages. The resulting rho values of genes, promoter regions, and enhancer regions were considered significant when both the raw *P*-value and the permuted *P*-value were < 0.05.

GO and Enriched Terms

GO was identified by first extracting protein sequences for all genes in the *X. hellerii* genome annotation GFF using agat_sp_extract_sequences.pl from AGAT (v0.8.1; Dainat). These proteins were then compared against the human proteome (https://www.uniprot.org/proteomes/UP000005640) using BLASTp (v2.7.1; Altschul et al. 1990) specifying -max_target_seqs 10 and -max_hsps 1. The resulting hits were sorted using a custom script and the top match for each gene noted. Human gene orthologues were matched to each *X. hellerii* gene and promoter region, and these were then filtered to include matches only when the BLASTp *E*-value was <0.05. To carry out a GSEA, the RERs (rho values) from RERconverge and the corresponding human gene orthologue for each *X. hellerii* was run through https://www.webgestalt.org/ for Biological Processes, Cellular components, and Molecular Functioning sets using the default parameters (supplementary table S5, Supplementary Material online). This was then repeated for the RERconverge results from promoter regions (supplementary table S5, Supplementary Material online). Gene orthology enrichment was calculated using all confidently assigned human gene orthologues for all *X. hellerii* genes as a background gene set and either the significant fast RER set (where rho was positive) or slow RER set (where rho was negative). This was repeated using the promoter set and their corresponding gene assignments.

Overlap with Focal Gene Lists and Differentially Expressed Genes from Previous Studies

To identify whether RER shifts may be directly associated with sulfide adaptation, we compared genes, promoters, and enhancers with significant RER shifts against a published list of genes involved in sulfide adaptation curated in Ryan et al. (2023). This was repeated with sets of genes that were differentially expressed in sulfidic lineages (Kelley et al. 2016) and genes with convergent gene expression shifts (Greenway et al. 2020). To determine whether the number of observed overlaps between RER-shifted genomic regions and previously published genes implicated in sulfide adaptation exceeded those expected by chance, we carried out Fisher's exact tests in R (using fisher.test).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgements

We would like to thank El Centro de Investigación e Innovación para la Enseañza y Aprendizaje and Universidad Juárez Autónoma de Tabasco, México for their support and facilitating the collection of samples. This research used resources of the shared high performance computing facility at the University of California, Santa Cruz. We would like to thank members of the Kelley and Cornejo labs for feedback on the project.

Author Contributions

M.T., L.A-R., and J.L.K. procured permits and collected samples. K.L.M. and J.L.K. prepared and sequenced samples. R.D-K. and B.W.P. performed the analyses with advice from M.T. and J.L.K. R.D-K. wrote the original draft of the

manuscript. B.W.P., K.L.M., J.L., L.A-R., M.T., and J.L.K. reviewed and edited the manuscript.

Funding

This work was supported by grants from the National Science Foundation (IOS-1931650, IOS-1931657, and IOS-2311366), the US Army Research Office (W911NF-15-1-0175 and W911NF-16-1-0225), and the Des Lee Collaborative Vision in Zoological Studies.

Data Availability

Data will be available from NCBI's GenBank (BioProject number PRJNA866745). Previously published data by Brown et al. (2019) is also available (BioProject number PRJNA473350). All bioinformatics scripts associated with this manuscript can be found at: https://github.com/RishiDeKayne/Pmex_2023/.

Ethics

All experiments were approved by the Institutional Animal Care and Use Committee of Kansas State University (ACUP #3473). Fieldwork was approved by the governments of Mexico (Fieldwork Permits: SGPA/DGVS/04315/11, SGPA/DGVS/000261/18, PRMN/DGOPA-012/2017) and the Dominican Republic (0092-11).

Literature Cited

Albert FW, Kruglyak L. The role of regulatory variation in complex traits and disease. Nat Rev Genet. 2015:16(4):197–212. https://doi.org/10.1038/nrg3891.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Artur MAS, Kajala K. Convergent evolution of gene regulatory networks underlying plant adaptations to dry environments. Plant Cell Environ. 2021:44(10):3211–3222. https://doi.org/10.1111/pce.14143.

Barts N, Greenway R, Passow CN, Arias-Rodriguez L, Kelley JL, Tobler M. Molecular evolution and expression of oxygen transport genes in livebearing fishes (Poeciliidae) from hydrogen sulfide rich springs. Genome. 2018:61(4):273–286. https://doi.org/10.1139/gen-2017-0051.

Belyaeva OV, Korkina OV, Stetsenko AV, Kedishvili NY. Human retinol dehydrogenase 13 (RDH13) is a mitochondrial short-chain dehydrogenase/reductase with a retinaldehyde reductase activity. FEBS J. 2008:275(1):138–147. https://doi.org/10.1111/j.1742-4658.2007.06184.x.

Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. Local ancestry analysis reveals genomic convergence in extremophile fishes. Philos Trans R Soc Lond B Biol Sci. 2019:374(1777):20180240. https://doi.org/10.1098/rstb.2018.0240.

Conte GL, Arnegard ME, Peichel CL, Schluter D. The probability of genetic parallelism and convergence in natural populations. Proc Biol Sci. 2012:279(1749):5039–5047. https://doi.org/10.1098/rspb.2012.2146.

Dainat J. AGAT: Another Gff Analysis Toolkit to handle annotations in any GTF/GFF format (v0.8.0). Zenodo. https://doi.org/10.5281/zenodo.3552717.

Damsgaard C, Baliga VB, Bates E, Burggren W, McKenzie DJ, Taylor E, Wright PA. Evolutionary and cardio-respiratory physiology of air-breathing and amphibious fishes. Acta Physiol (Oxf). 2020:228(3):e13406. https://doi.org/10.1111/apha.13406.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021:10(2):giab008. https://doi.org/10.1093/gigascience/giab008.

Davies PL, Baardsnes J, Kuiper MJ, Walker VK. Structure and function of antifreeze proteins. Philos Trans R Soc Lond B Biol Sci. 2002:357(1423):927–935. https://doi.org/10.1098/rstb.2002.1081.

Development Core Team R. R: A language and environment for statistical computing. 2011. https://www.R-project.org/.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dobler S, Dalla S, Wagschal V, Agrawal AA. Community-wide convergent evolution in insect adaptation to toxic cardenolides by substitutions in the Na, K-ATPase. Proc Natl Acad Sci U S A. 2012:109(32):13040–13045. https://doi.org/10.1073/pnas.1202111109.

Duttke SH, Chang MW, Heinz S, Benner C. Identification and dynamic quantification of regulatory elements using total RNA. Genome Res. 2019:29(11):1836–1846. https://doi.org/10.1101/gr.253492.119.

Foote AD, Liu Y, Thomas GW, Vinař T, Alföldi J, Deng J, Dugan S, van Elk CE, Hunter ME, Joshi V, et al. Convergent evolution of the genomes of marine mammals. Nat Genet. 2015:47(3):272–275. https://doi.org/10.1038/ng.3198.

Giles GI, Jacob C. Reactive sulfur species: an emerging concept in oxidative stress. Biol Chem. 2002:383(3-4):375–388. https://doi.org/10.1515/BC.2002.042.

Greenway R, Barts N, Henpita C, Brown AP, Arias Rodriguez L, Rodríguez Peña CM, Arndt S, Lau GY, Murphy MP, Wu L, et al. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. Proc Natl Acad Sci U S A. 2020:117(28):16424–16430. https://doi.org/10.1073/pnas.2004223117.

Greenway R, De-Kayne R, Brown AP, Camarillo H, Delich C, McGowan KL, Nelson J, Arias-Rodriguez L, Kelley JL, Tobler M. Integrative analyses of convergent adaptation in sympatric extremophile fishes. bioRxiv 450104. https://doi.org/10.1101/2021.06.28.450104, 15 December 2023, preprint: not peer reviewed.

Härer A, Meyer A, Torres-Dowdall J. Convergent phenotypic evolution of the visual system via different molecular routes: how Neotropical cichlid fishes adapt to novel light environments. Evol Lett. 2018:2(4):341–354. https://doi.org/10.1002/evl3.71.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010:38(4):576–589. https://doi.org/10.1016/j.molcel.2010.05.004.

Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007:61(5):995–1016. https://doi.org/10.1111/j.1558-5646.2007.00105.x.

Ishikawa A, Yamanouchi S, Iwasaki W, Kitano J. Convergent copy number increase of genes associated with freshwater colonization in fishes. Philos Trans R Soc Lond B Biol Sci. 2022:377(1855):20200509. https://doi.org/10.1098/rstb.2020.0509.

Jiang J, Chan A, Ali S, Saha A, Haushalter KJ, Lam WL, Glasheen M, Parker J, Brenner M, Mahon SB, et al. Hydrogen sulfide—mechanisms of toxicity and development of an antidote. Sci Rep. 2016:6(1):1–10. https://doi.org/10.1038/s41598-016-0001-8.

Jiao F, Zhang L, Limbu SM, Yin H, Xie Y, Yang Z, Shang Z, Kong L, Rong H. A comparison of digestive strategies for fishes with different feeding habits: digestive enzyme activities, intestinal morphology, and gut microbiota. Ecol Evol. 2023:13(9):e10499. https://doi.org/10.1002/ece3.10499.

Junge W, Nelson N. ATP synthase. Ann Rev Biochem. 2015:84(1):631–657. https://doi.org/10.1146/annurev-biochem-060614-034124.

Kelley JL, Arias-Rodriguez L, Patacsil Martin D, Yee MC, Bustamante CD, Tobler M. Mechanisms underlying adaptation to life in hydrogen sulfide-rich environments. Mol Biol Evol. 2016:33(6):1419–1434. https://doi.org/10.1093/molbev/msw020.

Kimura Y, Goto Y-I, Kimura H. Hydrogen sulfide increases glutathione production and suppresses oxidative stress in mitochondria. Antioxid Redox Signal. 2010:12(1):1–13. https://doi.org/10.1089/ars.2008.2282.

Kolora SRR, Owens GL, Vazquez JM, Stubbs A, Chatla K, Jainese C, Seeto K, McCrea M, Sandel MW, Vianna JA, et al. Origins and evolution of extreme life span in Pacific Ocean rockfishes. Science. 2021:374(6569):842–847. https://doi.org/10.1126/science.abg5332.

Kowalczyk A, Chikina M, Clark N. Complementary evolution of coding and noncoding sequence underlies mammalian hairlessness. eLife. 2022:11:e76911. https://doi.org/10.7554/eLife.76911.

Kowalczyk A, Meyer WK, Partha R, Mao W, Clark NL, Chikina M. RERconverge: an R package for associating evolutionary rates with convergent traits. Bioinformatics. 2019:35(22):4815–4817. https://doi.org/10.1093/bioinformatics/btz468.

Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics. 2011:27(21):2987–2993. https://doi.org/10.1093/bioinformatics/btr509.

Losos JB. Convergence, adaptation, and constraint. Evolution. 2011:65(7):1827–1840. https://doi.org/10.1111/j.1558-5646.2011.01289.x.

Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011:17(1):10–12. https://doi.org/10.14806/ej.17.1.200.

Mathai JC, Missner A, Kügler P, Saparov SM, Zeidel ML, Lee JK, Pohl P. No facilitator required for membrane transport of hydrogen sulfide. Proc Natl Acad Sci U S A. 2009:106(39):16633–16638. https://doi.org/10.1073/pnas.0902952106.

Mirarab S, Reaz R, Bayzid MS, Zimmermann T, Swenson MS, Warnow T. ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics. 2014:30(17):541–548. https://doi.org/10.1093/bioinformatics/btu462.

Morel B, Kozlov AM, Stamatakis A. ParGenes: a tool for massively parallel model selection and phylogenetic tree inference on thousands of genes. Bioinformatics. 2019:35(10):1771–1773. https://doi.org/10.1093/bioinformatics/bty839.

Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. Science. 2016:354(6310):336–339. https://doi.org/10.1126/science.aaf9070.

Olson KR. H2s and polysulfide metabolism: conventional and unconventional pathways. Biochem Pharmacol. 2018:149:77–90. https://doi.org/10.1016/j.bcp.2017.12.010.

Pedersen BS, Quinlan AR. Mosdepth: quick coverage calculation for genomes and exomes. Bioinformatics. 2018:34(5):867–868. https://doi.org/10.1093/bioinformatics/btx699.

Perry B, McGowan K, Arias-Rodriguez L, Tobler M, Kelley JL. Nascent transcription reveals regulatory changes in extremophile fishes inhabiting hydrogen sulfide-rich environments. Proc R Soc B. (in review). https://doi.org/10.1098/rspb.2019.0910.

Pfenninger M, Lerp H, Tobler M, Passow C, Kelley JL, Funke E, Greshake B, Erkoc UK, Berberich T, Plath M. Parallel evolution of *cox* genes in H_2S-tolerant fish as key adaptation to a toxic environment. Nat Commun. 2014:5(1):3873. https://doi.org/10.1038/ncomms4873.

Plath M, Pfenninger M, Lerp H, Riesch R, Eschenbrenner C, Slattery PA, Bierbach D, Herrmann N, Schulte M, Arias-Rodriguez L, et al. Genetic differentiation and selection against migrants in evolutionarily replicated extreme environments. Evolution. 2013:67(9):2647–2661. https://doi.org/10.1111/evo.12133.

Rambaut A. FigTree v1.4; 2012 [accessed 2021 Nov 16]. http://tree.bio.ed.ac.uk/software/figtree/.

Riesch R. Hydrogen sulfide-toxic habitats. In: Riesch R, Tobler M, Plath M, editors. Extremophile fishes: ecology, evolution, and physiology of teleosts in extreme environments. Heidelberg, Germany: Springer; 2015. p. 137–159.

Riesch R, Tobler M, Lerp H, Jourdan J, Doumas T, Nosil P, Langerhans RB, Plath M. Extremophile Poeciliidae: multivariate insights into the complexity of speciation along replicated ecological gradients. BMC Evol Biol. 2016:16(1):136. https://doi.org/10.1186/s12862-016-0705-1.

Rodríguez-Machado S, Elías DJ, McMahan CD, Gruszkiewicz-Tolli A, Piller KR, Chakrabarty P. Disentangling historical relationships within Poeciliidae (Teleostei: Cyprinodontiformes) using ultraconserved elements. Mol Phylogenet Evol. 2024:190:107965. https://doi.org/10.1016/j.ympev.2023.107965.

Rosenblum EB, Parent CE, Brandt EE. The molecular basis of phenotypic convergence. Ann Rev Ecol Evol Syst. 2014:45(1):203–226. https://doi.org/10.1146/annurev-ecolsys-120213-091851.

Ryan K, Greenway R, Landers J, Arias-Rodriguez L, Tobler M, Kelley JL. Selection on standing genetic variation mediates convergent evolution in extremophile fish. Mol Ecol. 2023:32(18):5042–5054. https://doi.org/10.1111/mec.17081.

Sackton TB, Grayson P, Cloutier A, Hu Z, Liu JS, Wheeler NE, Gardner PP, Clarke JA, Baker AJ, Clamp M, Convergent regulatory evolution and loss of flight in paleognathous birds. Science. 2019:364(6435):74–78. https://doi.org/10.1126/science.aat7244.

Schliep KP. Phangorn: phylogenetic analysis in R. Bioinformatics. 2011:27(4):592–593. https://doi.org/10.1093/bioinformatics/btq706.

Sidell BD, O'Brien KM. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. J Exp Biol. 2006:209(Pt 10):1791–1802. https://doi.org/10.1242/jeb.02091.

Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? Evolution. 2008:62(9):2155–2177. https://doi.org/10.1111/j.1558-5646.2008.00450.x.

Sugawara T, Terai Y, Okada N. Natural selection of the rhodopsin gene during the adaptive radiation of East African Great Lakes cichlid fishes. Mol Biol Evol. 2002:19(10):1807–1811. https://doi.org/10.1093/oxfordjournals.molbev.a004004.

Tarasov A, Vilella AJ, Cuppen E, Nijman IJ, Prins P. Sambamba: fast processing of NGS alignment formats. Bioinformatics. 2015:31(12):2032–2034. https://doi.org/10.1093/bioinformatics/btv098.

Tenaillon O, Rodríguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. The molecular diversity of adaptive convergence. Science. 2012:335(6067):457–461. https://doi.org/10.1126/science.1212986.

Tobler M, Dewitt TJ, Schlupp I, García de León FJ, Herrmann R, Feulner PG, Tiedemann R, Plath M. Toxic hydrogen sulfide and dark caves: phenotypic and genetic divergence across two abiotic

environmental gradients in Poecilia mexicana. Evolution. 2008:62(10):2643–2659. https://doi.org/10.1111/j.1558-5646.2008.00466.x.

Tobler M, Hastings L. Convergent patterns of body shape differentiation in four different clades of poeciliid fishes inhabiting sulfide springs. Evol Biol. 2011:38(4):412–421. https://doi.org/10.1007/s11692-011-9129-4.

Tobler M, Kelley JL, Plath M, Riesch R. Extreme environments and the origins of biodiversity: adaptation and speciation in sulphide spring fishes. Mol Ecol. 2018:27(4):843–859. https://doi.org/10.1111/mec.14497.

Tobler M, Palacios M, Chapman LJ, Mitrofanov I, Bierbach D, Plath M, Arias-Rodriguez L, de León FJ, Mateos M. Evolution in extreme environments: replicated phenotypic differentiation in livebearing fish inhabiting sulfidic springs. Evolution. 2011:65(8):2213–2228. https://doi.org/10.1111/j.1558-5646.2011.01298.x.

Tobler M, Scharnweber, K., Greenway, R., Passow, CN, Arias-Rodriguez L, García-De-León F. J. Convergent changes in the trophic ecology of extremophile fish along replicated environmental gradients. Freshwater Biol. 2015:60(4):768–780. https://doi.org/10.1111/fwb.12530.

van der Zee MJ, Whiting JR, Paris JR, Bassar RD, Travis J, Weigel D, Reznick DN, Fraser BA. Rapid genomic convergent evolution in experimental populations of Trinidadian guppies (*Poecilia reticulata*). Evol Lett. 2022:6(2):149–161. https://doi.org/10.1002/evl3.272.

Vaquer-Sunyer R, Duarte CM. Sulfide exposure accelerates hypoxia-driven mortalit. Limnol Oceanogr. 2010:55(3):1075–1082. https://doi.org/10.4319/lo.2010.55.3.1075.

Wang M, Guo Z, Wang S. The effect of certain conditions in the regulation of cystathionine γ-lyase by exogenous hydrogen sulfide in mammalian cells. Biochem Genet. 2013:51(7-8):503–513. https://doi.org/10.1007/s10528-013-9581-1.

White LJ, Sutton G, Shechonge A, Day JJ, Dasmahapatra KK, Pownall ME. Adaptation of the carbamoyl-phosphate synthetase enzyme in an extremophile fish. R Soc Open Sci. 2020:7(10):201200. https://doi.org/10.1098/rsos.201200.

Xu S, Wang J, Guo Z, He Z, Shi S. Genomic convergence in the adaptation to extreme environments. Plant Commun. 2020:1(6):100117. https://doi.org/10.1016/j.xplc.2020.100117.

Associate editor: Allie Graham