Title: Selection on standing genetic variation mediates convergent evolution in extremophile fish.

Short running title: (45 characters) Genomic convergence in sulfidic fishes

Authors: Kara Ryan[1](kara.ryan@wsu.edu), Ryan Greenway[2,4] (ryan.greenway@uni-konstanz.de), Jake Landers[1] (jake.landers@wsu.edu), Lenin Arias-Rodriguez[3] (leninariasrodriguez@hotmail.com) , Michael Tobler[4] (tobler@ksu.edu), Joanna L. Kelley[1,5] (jokelley@ucsc.edu)

[1] School of Biological Sciences, Washington State University, 100 Dairy Road, Pullman, WA 99164, USA

[2] Zoology and Evolutionary Biology, Department of Biology, University of Konstanz, Universitätsstraße 10, 78464 Konstanz, Germany

[3] División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco (UJAT), C.P. 86150, Villahermosa, Tabasco, México

[4] Division of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS 66506, USA

[5] Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 130 McAllister Way, Santa Cruz, CA 95060, USA

Abstract:

Hydrogen sulfide is a toxic gas that disrupts numerous biological processes, including energy production in the mitochondria, yet fish in the *Poecilia mexicana* species complex have independently evolved sulfide tolerance several times. Despite clear evidence for convergence at the phenotypic level in these fishes, it is unclear if the repeated evolution of hydrogen sulfide tolerance is the result of similar genomic changes. To address this gap, we used a targeted capture approach to sequence genes associated with sulfide processes and toxicity from five sulfidic and five nonsulfidic populations in the species complex. By comparing sequence variation in candidate genes to a reference set, we identified similar population structure and differentiation, suggesting that patterns of variation in most genes associated with sulfide processes and toxicity are due to demographic history and not selection. But the presence of tree discordance for a subset of genes suggests that several loci are evolving divergently between ecotypes. We identified two differentiation outlier genes that are associated with sulfide detoxification in the mitochondria that have signatures of selection in all five sulfidic populations. Further investigation into these regions identified long, shared haplotypes among sulfidic populations. Together, these results reveal that selection on standing genetic variation in putatively adaptive genes may be driving phenotypic convergence in this species complex.

Keywords: 4-6 words

Adaptation | Convergent evolution | Hydrogen sulfide | Poeciliidae | Population genomics

Introduction

Convergent evolution, the independent evolution of similar traits in multiple lineages (Losos, 2011), can occur at various levels of biological organization, from phenotype to nucleotide position. However, convergence at one level does not necessitate convergence in the underlying mechanism (Rosenblum et al., 2014). In some cases, convergent phenotypes are largely driven by the same amino acid change (e.g. echolocation Liu et al., 2010; Rossiter et al., 2011), but in others, there is little convergence at the amino acid or gene level (e.g. hemoglobin Natarajan et al., 2016). Moreover, the degree to which convergent phenotypes are the result of selection on standing variation or on *de novo* mutation remains unclear. In some cases, the source of adaptive variation is largely *de novo* mutation (e.g. pigmentation in deer mice, Linnen et al., 2009) but in others, the majority source is standing variation (Alves et al., 2019; Haenel et al., 2019; Jones et al., 2012; Lai et al., 2019; N. M. Reid et al., 2016). Determining how often convergent phenotypes are driven by similar genomic changes will provide a better understanding of the repeatability of evolution (Gould, 1990; Rosenblum et al., 2014) and the relative contributions of selection on standing variation and *de novo* mutation in adaptation.

Extreme environments provide valuable systems to explore how selection may drive convergence across multiple biological levels. The strong selective regimes found in extreme environments often promote the evolution of convergent phenotypes (Tobler et al., 2015; Xu et al., 2020). These systems are particularly valuable when there are naturally replicated environments that have been independently colonized by multiple populations, allowing the investigation of how selection has independently shaped

variation in the genome (Tobler et al., 2018). Freshwater springs in the Río Grijalva basin in Southern Mexico are an example of such a system. Springs in this area contain naturally occurring hydrogen sulfide (H_2S) at concentrations orders of magnitude higher than those considered lethal for most animals (Tobler et al., 2016, 2018). Despite the highly toxic conditions, sulfide springs in four river drainages have been independently colonized by fish populations from the *Poecilia mexicana* species complex, including *Poecilia mexicana mexicana, Poecilia thermalis, and Poecilia sulphuraria* (Palacios et al., 2013; Tobler et al., 2018). Additionally, closely related populations inhabit nonsulfidic streams within the same drainages (Fig. 1), providing a unique comparative framework to explore evolutionary processes that potentially give rise to convergent adaptations (Tobler et al., 2018). The sulfide spring populations from the Puyacatengo and Tacotalpa drainages show evidence for recent divergence between ecotypes, estimated at ~10,000 years ago in Tacotalpa and ~300 years ago in Puyacatengo (Brown et al., 2018). In contrast, sulfide spring populations in the Pichucalco and Ixtapangajoya drainages are hypothesized to have diverged earlier, with estimates from the Pichucalco drainage suggesting a timing of divergence of ~19,000 years ago (Greenway et al., 2021). Sulfide spring populations are locally adapted and exhibit convergent phenotypes related to morphology (Greenway et al., 2019; Riesch et al., 2016; Tobler & Hastings, 2011), life history traits (Riesch et al., 2014; Riesch, et al., 2010) and physiology (Greenway et al., 2020; Pfenninger et al., 2014; Plath et al., 2013; Tobler et al., 2011, 2016). These naturally replicated systems, therefore, provide a unique opportunity to explore the genomic basis of convergent adaptive phenotypes, from nucleotide position to pathway.

The main mechanism of toxicity imposed by H_2S involves the inhibition of aerobic respiration. In the mitochondria, H_2S binds to Complex IV (cytochrome c oxidase, COX), inhibiting oxidative phosphorylation (OxPhos) and halting aerobic energy production even at micromolar concentrations (Cooper & Brown, 2008; Hill et al., 1984). In addition to disrupting energy production, elevated H_2S concentrations can produce harmful effects by modulating ion channels (García-Bereguiaín et al., 2008), modifying oxygen transport proteins (Pietri et al., 2011), interacting with transcription factors (Budde & Roth, 2010) and signaling molecules (Calvert et al., 2009), and disrupting posttranslational modification of proteins (Mustafa et al., 2009). Despite the presence of a highly conserved detoxification pathway across eukaryotes, the sulfide:quinone oxidoreductase (SQR) pathway in the mitochondria (Hildebrandt & Grieshaber, 2008; Libiad et al., 2014), environmental exposure to H_2S is still potentially lethal for most animals (Lagoutte et al., 2010). Although the biochemical action of H_2S is well understood, the ways in which these proteins may be modified for adaptation to H_2S-rich environments are largely unknown. At a molecular level, the strong selective pressure imposed by constant exposure to H_2S is predicted to drive adaptive modification of genes associated with OxPhos and H_2S detoxification. These genes are prime targets for natural selection due to their importance in cell survival, susceptibility to H_2S, and their highly conserved nature across taxa, which provides an opportunity to test for convergence at a molecular level. However, because of the genomic complexity and redundancy of these processes—more than 200 genes are associated with H_2S tolerance, H_2S detoxification, or OxPhos—it is unclear to what extent the convergent evolution of sulfide tolerance is a result of convergence at the genomic level.

There is evidence for genomic convergence at the gene and nucleotide levels in subsets of populations of *P. mexicana*, that seems to arise from a combination of selection on standing variation and *de novo* mutation. For example, genes associated with the sulfide:quinone oxidoreductase pathway in the mitochondria, including *sulfide:quinone oxidoreductase* (*sqor*) and persulfide dioxygenase (*ethe1*) show evidence for selection on standing variation (Brown et al., 2018; Greenway et al., 2020; Pfenninger et al., 2015; Tobler et al., 2018). Additionally, these sulfur detoxification genes are differentially expressed between ecotypes (Brown et al., 2018; Kelley et al., 2016; Passow, et al., 2017; Tobler et al., 2014). Evidence for shared adaptive changes in subunits of OxPhos have been identified (Brown et al., 2018; Greenway et al., 2020; Kelley et al., 2016; Pfenninger et al., 2014), including evidence for selection acting on *de novo* mutations in mitochondrially encoded subunits of OxPhos complexes (Greenway et al., 2020; Pfenninger et al., 2014).

While previous studies support the importance of regions associated with H_2S detoxification and OxPhos, it remains untested whether selection has acted on shared regions of the genome across all sulfidic populations of this species complex. Furthermore, it is unclear if convergence is the result of selection on standing variation or *de novo* mutation. To address these questions, we used a targeted exon capture approach to sequence candidate genes associated with H_2S toxicity and detoxification from five sulfidic and five nonsulfidic populations of *P. mexicana*. Using these data, we tested 1) whether the relationship among populations at H_2S candidate genes differs from the relationship at background genes, and 2) whether a subset of genes associated with H_2S detoxification has been targeted by selection in all drainages. This

study highlights the importance of selection on standing genetic variation in the repeated evolution of complex traits in extreme environments.

Materials and Methods

Samples and sequencing

A total of 200 individuals were sampled from five sulfidic and five nonsulfidic populations (20 individuals per population) from the Pichucalco (Pich 1 and Pich 2), Ixtapangajoya (Ixta), Puyacatengo (Puya), and Tacotalpa (Taco) drainages in the Río Grijalva basin, Mexico (Fig. 1, Table S1). Sampling included populations of *P. sulphuraria* (Pich sulfidic), *P. thermalis* (Ixta sulfidic), and *P. mexicana mexicana* (all other populations).

Probes were designed for capture sequencing by Rapid Genomics to target a total of 415 nuclear-encoded genes, comprised of 250 candidate genes and 165 background genes (Table S2). The 166 candidate genes associated with sulfide detoxification and sulfur processing were identified using Gene Ontology (GO) terms (Table S3). The candidate set also included 84 nuclear-encoded OxPhos genes, identified using a BLASTn search of the *P. mexicana* reference genome for genes encoding subunits of OxPhos from Zhang & Broughton (2013). The background set contained 73 housekeeping genes from Zhang & Broughton (2013) that are highly expressed in all cell types and involved in critical functions, providing an appropriate comparator to OxPhos genes (Amsterdam et al., 2004; Warrington et al., 2000), and 92 additional nuclear-encoded genes involved in mitochondrial functions (excluding OxPhos and sulfide-related genes identified above) selected from the MitoCarta2.0 database (Calvo et al., 2016; Pagliarini et al., 2008) at random.

DNA was extracted from muscle tissue preserved in RNAlater (Ambion, Inc.) using the Gentra Puregene Tissue Kit following the manufacturer's protocol for purifying DNA from 5–10 mg of tissue with the following modifications: (i) tissues were homogenized using a micro pestle, (ii) centrifugation was carried out for 3.5 mins following the addition of protein precipitations solution, and (iii) centrifugation was performed for 2 mins following the addition of isopropanol. DNA was quantified using a Qubit fluorometer and was visualized on a 1% agarose gel.

Library preparation was performed by Rapid Genomics utilizing the Illumina high-throughput workflow and proprietary chemistry. Briefly, DNA was sheared to a mean fragment length of 400 base pairs (bp). The resulting fragments were end-repaired, followed by the incorporation of Illumina unique dual-indexed adapters and PCR enrichment. Probes from Rapid Genomics set RG_3101 were hybridized to the libraries and enriched for the targets of interest. Sequencing was performed on an Illumina HiSeq system with paired-end 150 bp reads. The resulting raw data were demultiplexed using Illumina's BCLtoFastq.

Quality control, variant calling, and filtering

Reads were quality checked using FastQC v0.11.9 (Andrews, 2010), and outputs were summarized using MultiQC v1.11 (Ewels et al., 2016). Reads were trimmed using the Cutadapt (M. Martin, 2011) wrapper TrimGalore v0.6.6 (https://github.com/FelixKrueger/TrimGalore) with parameters --stringency 5, --length 40, and default parameters trimming adaptors and reads with a quality score less than 20. Trimmed reads were aligned to the *P. mexicana* genome (NCBI accession:

GCA_001443325.1; (Warren et al., 2018) using BWA-MEM v0.7.17 (Li, 2013). The resulting SAM files were converted to BAM files using Samtools v1.8 (Li et al., 2009). The BAM files were then sorted and duplicates were marked using Picard Tools v2.21.4 (*Broadinstitute/Picard*, 2014/2022) *SortSam* and *MarkDuplicates*. Variants were called according to the Genome Analysis Tool Kit (GATK) v4.2.5.0 (Van der Auwera & O'Connor, 2020) best practices for data pre-processing for variant discovery and germline short variant discovery (single nucleotide polymorphisms (SNPs) and insertions/deletions (Indels). GATK's *HaplotypeCaller* was used in gvcf mode to generate intermediate per-sample Genomic Variant Call Format (GVCF) files, which were then consolidated using *GenomicsDBImport*. Samples were then joint genotyped using GATK *GenotypeGVCFs*, retaining invariant sites using the -allSites parameter and combined using *CombineVariants.*

The resulting VCF files were filtered for sequencing depth and missingness (--minDP 20, -- max_missing 0.9) using VCFtools v0.1.16 (Danecek et al., 2011). Additionally, VCFtools was used to remove a nonsulfidic individual from the Pichucalco drainage that was identified as a first-generation hybrid. To generate an all-sites VCF with proper variant filtering, we separated the files into variant (--mac 1) and invariant (--maf 0) sites using VCFtools. We filtered the variant file for quality and minor allele frequency (--minQ 30 --maf 0.01) using VCFtools. Additionally, we removed loci that were significantly out of Hardy Weinberg Equilibrium within each population ($P < 0.001$) using dDocent Perl script *filter_hwe_by_pop.pl* (Puritz, et al., 2014; Puritz, Matz, et al., 2014). The resulting filtered variant sites VCF were then concatenated to the invariant sites using BCFtools v1.10.2 *concat* (Danecek et al., 2021) and intersected with the

original bed file containing the targeted regions for probe design to remove off-target variants and split the data to target and background using BEDTools v2.27.1 (Quinlan & Hall, 2010). The resulting filtered all-sites VCF file was converted to phylip format for phylogenetic analysis using python script vcf2phylip.py (https://github.com/edgardomortiz/vcf2phylip). In addition to an all-sites VCF, we filtered the candidate and background VCF files to retain only biallelic SNPs using VCFtools and remove SNPs found in high LD (+prune -l 0.8 -w 1000) using BCFtools. Unless otherwise stated, analyses were performed using the VCF containing biallelic, LD-pruned SNPs.

Analysis of evolutionary relationships among populations

To investigate population structure, a principal component analysis (PCA) was performed using Plink2 (Chang et al., 2015). Iqtree2 v2.1.3 (Minh et al., 2020) Ultra-Fast Bootstrap (Minh et al., 2013) (-B 1000 -bnni) approach was used to generate a maximum likelihood tree for the background and candidate gene sets using the unpruned, all-sites phylip file. ADMIXTURE v1.3.0 (Alexander et al., 2009) was used to investigate individual ancestry. Pong v1.5 (Behr et al., 2016) was used to visualize ADMIXTURE clustering.

Analysis of population genetic differentiation

Estimates of F_{ST}, heterozygosity, and nucleotide diversity (π) for each gene set were calculated using Stacks v2.59 *populations* (Catchen et al., 2013). Fixed differences between each comparison were identified by first filtering for private alleles and then

filtering for a maximum minor allele frequency of 0 using VCFtools. Pixy v1.2.5 (Korunes & Samuk, 2021) was used to summarize nucleotide diversity and estimate genetic differentiation at the gene level between sulfidic and nonsulfidic populations from the same drainage using the all-sites VCF. First, Nei and Li's nucleotide diversity (π) (Nei & Li, 1979) was estimated for all populations, and the difference in nucleotide diversity between pairs ($\Delta\pi_{ecotype}$) was calculated by subtracting π of sulfidic from nonsulfidic populations (π_{NS}-π_{S}), such that loci with decreased nucleotide diversity in sulfidic populations would result in $+\Delta\pi_{ecotype}$. Next, we estimated relative population differentiation between pairs using Weir and Cockerham's estimate of F_{ST} (Weir & Cockerham, 1984). Because this estimator can result in a negative value for populations that contain more variation within, we replaced all negative F_{ST} values with 0. Estimates of F_{ST} can be influenced by differences in nucleotide diversity within populations (Cruickshank & Hahn, 2014), therefore we also calculated d_{xy} (Nei & Li, 1979) as an absolute measure of population differentiation.

We first identified outlier loci putatively under selection using a multivariate approach in Minotaur v0.0.1 (Verity et al., 2017). We used the distributions of F_{ST}, d_{xy}, and $\Delta\pi_{ecotype}$ per gene to calculate the Mahalanobis distance (Mahalanobis, 1936), and loci with greater-than-expected differentiation based on a 95 % confidence intervals were considered putatively under selection. We then compared these drainage-specific gene lists to identify shared outlier genes. In addition to the per-gene approach, we identified outlier SNPs. We used VCFtools to calculate Weir and Cockerham's F_{ST} between all sulfidic and nonsulfidic populations on a per-site basis using a VCF containing biallelic SNPs that were not LD pruned. Outlier SNPs were identified using a

99.5% cutoff. SNPs were annotated using SNPeff v5.0e (Cingolani et al., 2012), and the potential effects of high F_{ST} nonsynonymous mutations were assessed using PolyPhen2 (Adzhubei et al., 2013).

Haplotype Network

To generate haplotype networks, we split the non-LD pruned VCF containing biallelic SNPs by population using VCFtools and phased the sites using the Popgen Pipeline Platform script vcf_phase.py using the Beagle v5.1 algorithm (Browning et al., 2018; Browning & Browning, 2007). The resulting VCFs were split into individual VCFs using BCFtools *query* and *view.* For each individual, a fasta file containing each haplotype was generated using BCFtools *consensus* (using the -H 1 and -H 2 parameter for haplotype 1 and haplotype 2, respectively). The header of the resulting fasta files was fixed to match the appropriate sample and haplotype before concatenating. Each gene-specific fasta file was then aligned using mafft v7.429 (Katoh & Standley, 2013) using default parameters. Haplotype networks were then generated for each gene using R v4.1.2 packages pegas v1.1 (Paradis, 2010) and ape v5.6.2 (Paradis & Schliep, 2019) using a statistical parsimony network (TCS) approach (Clement et al., 2000; Templeton et al., 1992) with singleton haplotypes filtered prior to network generation. The resulting network was visualized using Cytoscape (Shannon et al., 2003) v3.9.1.

Results

Targeted capture sequencing

Targeted capture sequencing resulted in 0.2–1.9 million 150-bp paired-end reads per sample, with an average of 0.7 million reads and an average Phred score greater than 35. GC content in reads ranged from 43–47%. Of the 839,328 bp targeted, 784,629 were retained after filtering, of which 9,446 (1.2%) were single nucleotide polymorphisms (SNPs). LD filtering resulted in a final set of 7,277 SNPs. Genes associated with H_2S had more SNPs in high LD (26.1%) when compared to OxPhos (19.1%) and background (19.2%) genes (Table 1).

Similar population structure inferred from background and candidate genes

To investigate the relationship among populations within each gene set, we used admixture to cluster individuals based on ancestry. The background set supported a best *K* of 7, while both the OxPhos and sulfide sets supported a best *K* of 8. Nonetheless, the relationship among populations was similar (Fig. S1, Fig. 2A). Variation in the best *K* was the result of nonsulfidic individuals from Pich clustering as a single population in the background set and as two populations in the OxPhos and sulfide sets (Fig. 2A). Generally, sulfide spring populations were recovered as distinct clusters, but Puya sulfidic and nonsulfidic individuals clustered as a single population in all gene sets (Fig. 2A).

PCA of LD-filtered SNPs separated *P. sulphuraria (Pich 1 and Pich 2, sulfidic)* and *P. thermalis (Ixta, sulfidic)* individuals from *P. mexicana* individuals (all others) along PC axis 1, which explained 30.3–34.9 % of variation depending on the gene set (Fig. 2B). PC axis 2 separated *P. mexicana* individuals by ecotype and explained 11.8–13.2 % of variation. Interestingly, the relationship among individuals from Puya and

299 sulfidic individuals from Taco varied across gene sets (Fig. 2B). For example, Puya
300 individuals clustered as a distinct group from Taco sulfidic individuals in both the
301 background and the sulfide set (Fig. 2B). In contrast, sulfidic individuals from Taco and
302 Puya clustered as a single group in the OxPhos set (Fig. 2B). PC axis 3 and 4 showed
303 similar clustering patterns between the sulfide and Oxphos sets, but interestingly, the
304 background set clustered sulfidic individuals from Taco with nonsulfidic individuals from
305 Pich (Fig. S2).

306 In addition to clustering analyses, we estimated relative differentiation using Weir
307 and Cockerham F_{ST} for each pairwise comparison. In all gene sets, we found the
308 highest F_{ST} between the Taco sulfidic population and the sulfidic populations from Pich
309 and Ixta and the lowest F_{ST} between sulfidic and nonsulfidic populations from Puya (Fig.
310 S3). Additionally, F_{ST} tended to be lower between comparisons of nonsulfidic
311 populations than between sulfidic populations or between populations of differing
312 ecotypes across all gene sets (Fig. S4). Observed and expected heterozygosity (H_O/H_E)
313 was reduced in all sulfidic populations when compared to the adjacent nonsulfidic
314 population, except for Puya, but this pattern was consistent across gene sets (Table
315 S4). Similarly, nucleotide diversity (π) was lower in sulfidic populations compared to
316 nonsulfidic populations (Table S5). We also compared distributions of per gene F_{ST}, d_{xy},
317 and Δπ between ecotypes from the same drainage. Within a drainage, the distributions
318 of variation in both sulfur processing and OxPhos genes were similar to the background
319 (Fig. S5). But between drainages, these distributions varied greatly (Fig. S5) Due to the
320 overall similarity with the background, our results suggest that most of the variation in

genes associated with OxPhos and sulfide processing is a result of demographic processes and not selection acting on many potentially adaptive loci.

Tree discordance between background and potential target genes

Despite limited evidence of selection acting on many loci across our targeted genes, phylogenetic analysis revealed tree discordance between background, sulfide processing, and OxPhos genes. The topology of the highest supported maximum likelihood tree of the background set supported previous studies (Brown et al., 2019; Greenway et al., 2020; Palacios et al., 2013) that suggest three independent colonizations of sulfidic springs—a more ancient colonization by the *P. sulphuraria* clade as well as two more recent colonizations by *P. mexicana* in the Taco and Puya drainages (Fig. 3). In contrast, the sulfide set clustered all populations by ecotype, contradicting the expected population tree (Fig. 3). The OxPhos set was similar to the background set, except that the Taco sulfidic individuals cluster with both Puya populations instead of as sister taxa with the nonsulfidic population from the same drainage (Fig. 3).

Evidence for selection in sulfide processing genes

A small subset of loci associated with sulfide processes were outliers and therefore putatively under selection in all sulfidic populations. The distribution of differentiation (F_{ST} and d_{xy}) and $\Delta\pi_{ecotype}$ were similar among gene sets but varied among drainages, suggesting variation in demographic history (Fig. S5). We identified six Mahalanobis outlier gene regions putatively under selection that were shared among all comparisons

(Fig. S6), of which five were associated with sulfide processes (Table S6). These regions included both copies of *ethe1*, galactose-specific lectin nattectin-like, ladderlectin-like, and C-type lectin domain family 10 member A-like. The final outlier locus was a background gene, succinate-CoA ligase ADP-forming subunit beta (*sucla2*). Other notable outlier genes associated with sulfide processes shared among some, but not all comparisons, included solute carrier family 26 member 1 (*slc26a1*) in the Pich 1, Ixta, Puya, and Taco comparisons, mercaptopyruvate sulfurtransferase (*mpst*) in Pich 2, Ixta, Puya and Taco, and *sqor* in the Ixta, Puya and Taco comparisons. Additionally, the OxPhos-associated gene cytochrome c oxidase subunit 8A (*cox8a*) was considered an outlier in the Pich 1, Pich 2, Puya, and Taco comparisons.

In addition to identifying shared outlier genes between ecotypes within the same drainage, we identified 45 highly differentiated SNPs between all sulfidic and nonsulfidic individuals based on a 99.5 % empirical cutoff (Fig. 4, Table S7). Of these 45 outlier SNPs, 35 were associated with sulfide processes, seven in OxPhos genes, and three in background genes (Table S7). Notably, the top ten most differentiated SNPs (F_{ST} 0.87–0.95) were located in two sulfide detoxification genes, *sqor* and *ethe1* (Fig. 4A). Four were in the 3' UTR of *ethe1.a,* and two were nonsynonymous mutations in *sqor.* The nonsynonymous mutations identified in *sqor* included a change from alanine to valine and from arginine to lysine, but both amino acid changes had a low predicted impact on the structure and function of the protein according to the Polyphen2 score (0.003–0.036 for Ala to Val, 0 for Arg to Lys). Of the remaining top outlier SNPs, three were synonymous mutations in *sqor* and one was a synonymous mutation in *ethe1* (Table S7). In addition to *sqor* and *ethe1*, we found four solute carrier family genes (slc13a1,

slc25a35, slc26a1, and slc26a5) that contained highly differentiated sites (F_{ST} 0.73-0.77, Table S7). Of the seven outlier SNPs found in OxPhos genes, six were in cytochrome c oxidase assembly homolog COX15. We identified three nonsynonymous positions in *cox15*, including an amino acid change from serine to cysteine that was predicted to be possibly damaging (PolyPhen2 = 0.952, Table S7). We identified three background genes that each contained a single outlier SNP, hypoxia inducible domain family member 1A (*higd1a*), 2-oxoglutarate dehydrogenase (*ogdh*), and ribosomal protein S15 (*rps15*) (Table S7).

To better understand sequence variation in regions putatively under selection, we generated haplotype networks for outlier genes of interest. Both *ethe1.a* and *sqor* showed a reduced number of sulfidic haplotypes compared to nonsulfidic haplotypes, evidence for a monophyletic origin (Fig. 4b), and the topology of these networks varied greatly from background genes (Fig. S7). There is evidence of haplotype sharing between sulfidic and nonsulfidic individuals from Puya in *sqor* (~15,000bp region) and among Puya sulfidic, Taco sulfidic, and Puya nonsulfidic individuals in *ethe1.a* (~5,000bp region) (Fig. 4b). However, this low level of sharing is not surprising given the results of Admixture, PCA, and tree discordance, which suggest ongoing gene flow among these populations (Fig. 2–3). Similar to *ethe1.a* and *sqor,* we see a partitioning of haplotypes by ecotype in the *cox15* haplotype network and a low level of sharing between ecotypes (Fig. S8).

Discussion

Although convergent phenotypic evolution is common, it remains unclear whether convergent phenotypes typically arise through similar genomic changes (Kitano et al., 2022; A. Martin & Orgogozo, 2013). Furthermore, it is unknown how often these similar genomic changes are the result of *de novo* mutation, standing ancestral variation, or introgressed loci (Rosenblum et al., 2014). Convergent adaptive traits that result from selective pressures imposed by extreme environments provide an opportunity to explore how strong selection may shape convergence at various levels of biological organization. In this study, we utilized a naturally replicated extreme environment—sulfide-rich springs that harbor populations of fishes that have independently adapted to highly toxic conditions—to test hypotheses about the role of genomic convergence in the independent evolution of sulfide tolerance. In addition to identifying drainage-specific genes under selection, we identified two candidate genes associated with H_2S detoxification in the mitochondria, *sqor* and *ethe1*, that show evidence for selection on shared variation. This study suggests that the convergent evolution of H_2S tolerance in the *P. mexicana* species complex is the result of both shared and unique genomic changes.

Many regions putatively under selection are associated with H_2S detoxification.

H_2S is detoxified in the mitochondria via a series of enzymatic reactions associated with the SQR pathway (Libiad et al., 2014; Olson, 2018). This pathway begins with SQOR binding H_2S followed by a series of reactions involving a group of enzymes, including ETHE1, that oxidizes H_2S to a variety of excretable compounds that allow elimination of oxidized sulfur molecules from the system (Hildebrandt & Grieshaber, 2008; Libiad et

al., 2014; Olson, 2018). Our findings highlight the importance of ETHE1 and SQOR in adaptation to H_2S-rich environments. Previous studies have shown that *ethe1* is differentially expressed between ecotypes (Brown et al., 2018; Kelley et al., 2016; Passow, Brown et al., 2017; Passow, et al., 2017). Consistent with this, we found highly differentiated SNPs in the 3' UTR region of *ethe1*, which may play a role underlying these patterns of differential gene regulation and expression (Mayr, 2019). Additionally, previous studies have shown that sulfidic populations have higher SQOR activity and lower endogenous levels of H_2S than nonsulfidic populations at increasing levels of H_2S exposure, suggesting an increased detoxification ability (Greenway et al., 2020). Our analyses identified multiple, highly differentiated positions between all sulfidic and nonsulfidic individuals in coding regions of *sqor*, with two variants leading to differences in the encoded amino acids. Our findings raise the question of whether these amino acid changes are involved in, or potentially directly responsible for, observed increases in SQOR activity in sulfidic populations. Future work to categorize the activity of this enzyme as well as understanding the individual and combinatoric effects of these amino acid changes will be necessary to understand how these changes impact SQOR function and activity.

In addition to strong patterns of differentiation in two key detoxification enzymes, we see repeated patterns of increased differentiation in several other candidate genes, including solute carrier genes. Epistasis could be an important driver of these patterns. For example, mutations that increase the capacity to detoxify H_2S may only be beneficial if they occur when the genomic background contains alleles that allow for an increased capacity to remove the byproducts of detoxification out of the cell.

435 Theoretically, there could be an epistatic relationship between *sqor, ethe1,* and solute
436 carrier families, such as slc13 (Bergeron et al., 2013) and slc26 (Alper & Sharma, 2013)
437 that can remove detoxification byproducts out of the cell, limiting the number of
438 mutational paths as seen in other adaptations (Weinreich et al., 2006).

439 Although we see limited evidence for selection on genes associated with OxPhos
440 compared to H_2S detoxification genes, previous studies that included mitochondrially
441 encoded subunits of OxPhos have identified parallel amino acid changes associated
442 with increased cox resistance (Greenway et al., 2020; Pfenninger et al., 2014). It is
443 likely that both H_2S regulation and resistance are crucial in adapting to this extreme
444 environment, despite our results showing limited evidence for selection in nuclear-
445 encoded OxPhos genes, beyond *cox8a* and *cox15.* This result could be explained by
446 only needing adaptive modifications to the reactive core of COX, which is
447 mitochondrially encoded, to gain H_2S resistance. Future work to determine if
448 mitonuclear coevolution is necessary for H_2S tolerance or if modification of only
449 mitochondrial genes is sufficient.

450

451 **Sources of shared variation among drainages**

452 Adaptive variation underlying convergent phenotypes may arise independently among
453 populations as the result of selection on *de novo* mutations (in the same or different
454 genes), or may be shared via selection on introgressed or ancestral variation (K. Reid et
455 al., 2021; Rosenblum et al., 2014; Stern, 2013). Consistent with previous studies,
456 phylogenetic and admixture analyses provide evidence for the independent evolution of
457 H_2S tolerance among sulfide spring populations (Pfenninger et al., 2014; Tobler et al.,

458 2018). In addition, the sulfidic populations in Puya and Taco are more recently diverged
459 from their nonsulfidic ancestor than those in other drainages (Pfenninger et al., 2014;
460 Tobler et al., 2018). However, our analyses support selection on shared variation
461 among sulfidic populations in a small subset of genes. This result raises questions as to
462 the source of this putatively adaptive variation. Despite geographical barriers between
463 drainages, seasonal flooding could provide a mechanism for the movement of fish
464 between drainages and the potential for the export of adaptive alleles from a source
465 population, known as the transporter hypothesis (Schluter & Conte, 2009). Recent
466 simulation work provided quantitative proof-of-concept for the transporter hypothesis,
467 showing that only a few individuals are needed for the export of adaptive alleles
468 between populations in similar habitats through gene flow with connected populations in
469 other habitat types (Galloway et al., 2020). We identified shared putatively adaptive
470 haplotypes, greater than 15kb in length, that could be indicative of recent introgression.
471 In a recent study, Todesco et al. (2020) identified recently introgressed adaptive
472 haplotype blocks greater than 1Mb in sunflowers. Given our study used exon capture
473 data, which limits our ability to detect long haplotypes, we were unable to determine if
474 the source of adaptive variation was recent introgression or standing ancestral variation.
475 Distinguishing between the two potential sources of shared adaptive variation remains
476 an important unanswered question in this system.
477 Evolutionary biologists have long been fascinated by examples of repetitive
478 evolutionary trajectories. But even with the expansion of low-cost sequencing, it remains
479 unclear how often convergent phenotypes are driven by similar genomic changes. Our
480 study adds to the growing body of literature that suggests selection on shared variation

may be an important driver of phenotypic convergence within closely related populations and species (Brown et al., 2019; Waters & McCulloch, 2021). Future work using whole genome data in this species complex is necessary to elucidate the relative role of selection on introgressed and standing variants. Furthermore, understanding the source of adaptive alleles across the family Poeciliidae, which have independently evolved sulfide tolerance almost 20 times (Tobler et al., 2018), will provide important insight into fundamental questions of the predictability and repeatability of evolution.

Acknowledgments

Funding was provided by National Science Foundation under Grant No IOS-1931657 and the Army Research Office (W911NF-15-1-0175, W911NF-16-1-0225). This work used computational and storage servers associated with the Kamiak computing cluster at Washington State University. Thanks to Blair Perry and the Kelley lab members for manuscript feedback.

Data Accessibility

Data deposition: All scripts and vcf files associated with analyses are available on GitHub (https://github.com/kara-ry/H2S_Capture_Pmex). All sequence data are available at National Center for Biotechnology Information (NCBI) BioProject PRJNA647126.

Benefit-Sharing

This research is the result of a long-term, international collaboration. The contributions of all individuals to the research are included as co-authors or described in the acknowledgments. All genetic data have been shared with the broader public via appropriate biological databases. Our group places great importance on engaging in international scientific collaborations and contributing to institutional capacity building efforts.

Author contributions

JLK, MT, and RG conceived the study. KR and JL performed the analyses. KR wrote the manuscript. JLK, MT and RG contributed to the revisions. All author approved the final version.

References

Adzhubei, I., Jordan, D. M., & Sunyaev, S. R. (2013). Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics*, *Chapter 7*, Unit7.20. https://doi.org/10.1002/0471142905.hg0720s76

Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, *19*(9), 1655–1664. https://doi.org/10.1101/gr.094052.109

Alper, S. L., & Sharma, A. K. (2013). The SLC26 gene family of anion transporters and channels. *Molecular Aspects of Medicine*, *34*(2), 494–515. https://doi.org/10.1016/j.mam.2012.07.009

Alves, J. M., Carneiro, M., Cheng, J. Y., Lemos de Matos, A., Rahman, M. M., Loog, L., Campos, P. F., Wales, N., Eriksson, A., Manica, A., Strive, T., Graham, S. C., Afonso, S., Bell, D. J., Belmont, L., Day, J. P., Fuller, S. J., Marchandeau, S., Palmer, W. J., … Jiggins, F. M.

(2019). Parallel adaptation of rabbit populations to myxoma virus. *Science (New York, N.Y.)*, *363*(6433), 1319–1326. https://doi.org/10.1126/science.aau7285

Amsterdam, A., Nissen, R. M., Sun, Z., Swindell, E. C., Farrington, S., & Hopkins, N. (2004). Identification of 315 genes essential for early zebrafish development. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(35), 12792–12797. https://doi.org/10.1073/pnas.0403929101

Andrews, S. (2010). *FastQC A Quality Control tool for High Throughput Sequence Data*. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

Behr, A. A., Liu, K. Z., Liu-Fang, G., Nakka, P., & Ramachandran, S. (2016). pong: Fast analysis and visualization of latent clusters in population genetic data. *Bioinformatics*, *32*(18), 2817–2823. https://doi.org/10.1093/bioinformatics/btw327

Bergeron, M. J., Clémençon, B., Hediger, M. A., & Markovich, D. (2013). SLC13 family of Na+-coupled di- and tri-carboxylate/sulfate transporters. *The ABCs of Membrane Transporters in Health and Disease (SLC Series)*, *34*(2), 299–312. https://doi.org/10.1016/j.mam.2012.12.001

Broadinstitute/picard. (2022). [Java]. Broad Institute. https://github.com/broadinstitute/picard (Original work published 2014)

Brown, A. P., Arias-Rodriguez, L., Yee, M.-C., Tobler, M., & Kelley, J. L. (2018). Concordant Changes in Gene Expression and Nucleotides Underlie Independent Adaptation to Hydrogen-Sulfide-Rich Environments. *Genome Biology and Evolution*, *10*(11), 2867–2881. https://doi.org/10.1093/gbe/evy198

549 Brown, A. P., McGowan, K. L., Schwarzkopf, E. J., Greenway, R., Rodriguez, L. A., Tobler, M., &
550 Kelley, J. L. (2019). Local ancestry analysis reveals genomic convergence in extremophile
551 fishes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1777),
552 20180240. https://doi.org/10.1098/rstb.2018.0240
553 Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-
554 Generation Reference Panels. *The American Journal of Human Genetics*, *103*(3), 338–
555 348. https://doi.org/10.1016/j.ajhg.2018.07.015
556 Browning, S. R., & Browning, B. L. (2007). Rapid and Accurate Haplotype Phasing and Missing-
557 Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype
558 Clustering. *The American Journal of Human Genetics*, *81*(5), 1084–1097.
559 https://doi.org/10.1086/521987
560 Budde, M. W., & Roth, M. B. (2010). Hydrogen sulfide increases hypoxia-inducible factor-1
561 activity independently of von Hippel-Lindau tumor suppressor-1 in C. elegans. *Molecular*
562 *Biology of the Cell*, *21*(1), 212–217. https://doi.org/10.1091/mbc.e09-03-0199
563 Calvert, J. W., Jha, S., Gundewar, S., Elrod, J. W., Ramachandran, A., Pattillo, C. B., Kevil, C. G., &
564 Lefer, D. J. (2009). Hydrogen sulfide mediates cardioprotection through Nrf2 signaling.
565 *Circulation Research*, *105*(4), 365–374.
566 https://doi.org/10.1161/CIRCRESAHA.109.199919
567 Calvo, S. E., Clauser, K. R., & Mootha, V. K. (2016). MitoCarta2.0: An updated inventory of
568 mammalian mitochondrial proteins. *Nucleic Acids Research*, *44*(D1), D1251–D1257.
569 https://doi.org/10.1093/nar/gkv1003

570 Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An
571 analysis tool set for population genomics. *Molecular Ecology*, *22*(11), 3124–3140.
572 https://doi.org/10.1111/mec.12354
573 Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-
574 generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*,
575 *4*(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8
576 Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden,
577 D. M. (2012). A program for annotating and predicting the effects of single nucleotide
578 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w ^{1118} ;
579 iso-2; iso-3. *Fly*, *6*(2), 80–92. https://doi.org/10.4161/fly.19695
580 Clement, M., Posada, D., & Crandall, K. A. (2000). TCS: A computer program to estimate gene
581 genealogies. *Molecular Ecology*, *9*(10), 1657–1659. https://doi.org/10.1046/j.1365-
582 294x.2000.01020.x
583 Cooper, C. E., & Brown, G. C. (2008). The inhibition of mitochondrial cytochrome oxidase by the
584 gases carbon monoxide, nitric oxide, hydrogen cyanide and hydrogen sulfide: Chemical
585 mechanism and physiological significance. *Journal of Bioenergetics and Biomembranes*,
586 *40*(5), 533–539. https://doi.org/10.1007/s10863-008-9166-6
587 Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation
588 are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–
589 3157. https://doi.org/10.1111/mec.12796
590 Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E.,
591 Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., Durbin, R., & 1000 Genomes Project

592 Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15),
593 2156–2158. https://doi.org/10.1093/bioinformatics/btr330
594 Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A.,
595 Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and
596 BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008
597 Ewels, P., Magnusson, M., Lundin, S., & Käller, M. (2016). MultiQC: Summarize analysis results
598 for multiple tools and samples in a single report. *Bioinformatics*, *32*(19), 3047–3048.
599 https://doi.org/10.1093/bioinformatics/btw354
600 Galloway, J., Cresko, W. A., & Ralph, P. (2020). A Few Stickleback Suffice for the Transport of
601 Alleles to New Lakes. *G3 Genes|Genomes|Genetics*, *10*(2), 505–514.
602 https://doi.org/10.1534/g3.119.400564
603 García-Bereguiaín, M. A., Samhan-Arias, A. K., Martín-Romero, F. J., & Gutiérrez-Merino, C.
604 (2008). Hydrogen sulfide raises cytosolic calcium in neurons through activation of L-type
605 Ca2+ channels. *Antioxidants & Redox Signaling*, *10*(1), 31–42.
606 https://doi.org/10.1089/ars.2007.1656
607 Gould, S. J. (1990). *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton.
608 https://books.google.com/books?id=gwlVAgAAQBAJ
609 Greenway, R., Barts, N., Henpita, C., Brown, A. P., Rodriguez, L. A., Peña, C. M. R., Arndt, S., Lau,
610 G. Y., Murphy, M. P., Wu, L., Lin, D., Shaw, J. H., Kelley, J. L., & Tobler, M. (2020).
611 Convergent evolution of conserved mitochondrial pathways underlies repeated
612 adaptation to extreme environments. *Proceedings of the National Academy of Sciences*,
613 *117*(28), 16424–16430. https://doi.org/10.1073/pnas.2004223117

614 Greenway, R., Brown, A. P., Camarillo, H., Delich, C., McGowan, K. L., Nelson, J., Arias-
615 Rodriguez, L., Kelley, J. L., & Tobler, M. (2021). *Convergent adaptation and ecological*
616 *speciation result from unique genomic mechanisms in sympatric extremophile fishes* (p.
617 2021.06.28.450104). bioRxiv. https://doi.org/10.1101/2021.06.28.450104
618 Greenway, R., McNemee, R., Okamoto, A., Plath, M., Arias-Rodriguez, L., & Tobler, M. (2019).
619 Correlated divergence of female and male genitalia in replicated lineages with ongoing
620 ecological speciation. *Evolution*, *73*(6), 1200–1212. https://doi.org/10.1111/evo.13742
621 Haenel, Q., Roesti, M., Moser, D., MacColl, A. D. C., & Berner, D. (2019). Predictable genome-
622 wide sorting of standing genetic variation during parallel adaptation to basic versus
623 acidic environments in stickleback fish. *Evolution Letters*, *3*(1), 28–42.
624 https://doi.org/10.1002/evl3.99
625 Hildebrandt, T. M., & Grieshaber, M. K. (2008). Three enzymatic activities catalyze the oxidation
626 of sulfide to thiosulfate in mammalian and invertebrate mitochondria. *The FEBS Journal*,
627 *275*(13), 3352–3361. https://doi.org/10.1111/j.1742-4658.2008.06482.x
628 Hill, B. C., Woon, T. C., Nicholls, P., Peterson, J., Greenwood, C., & Thomson, A. J. (1984).
629 Interactions of sulphide and other ligands with cytochrome c oxidase. An electron-
630 paramagnetic-resonance study. *Biochemical Journal*, *224*(2), 591–600.
631 https://doi.org/10.1042/bj2240591
632 Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun,
633 M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M.
634 C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The

genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Katoh, K., & Standley, D. M. (2013). MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/molbev/mst010

Kelley, J. L., Arias-Rodriguez, L., Patacsil Martin, D., Yee, M.-C., Bustamante, C. D., & Tobler, M. (2016). Mechanisms Underlying Adaptation to Life in Hydrogen Sulfide–Rich Environments. *Molecular Biology and Evolution*, *33*(6), 1419–1434. https://doi.org/10.1093/molbev/msw020

Kitano, J., Ishikawa, A., Ravinet, M., & Courtier-Orgogozo, V. (2022). Genetic basis of speciation and adaptation: From loci to causative mutations. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1855), 20200503. https://doi.org/10.1098/rstb.2020.0503

Korunes, K. L., & Samuk, K. (2021). pixy: Unbiased estimation of nucleotide diversity and divergence in the presence of missing data. *Molecular Ecology Resources*, *21*(4), 1359–1368. https://doi.org/10.1111/1755-0998.13326

Lagoutte, E., Mimoun, S., Andriamihaja, M., Chaumontet, C., Blachier, F., & Bouillaud, F. (2010). Oxidation of hydrogen sulfide remains a priority in mammalian cells and causes reverse electron transfer in colonocytes. *Biochimica Et Biophysica Acta*, *1797*(8), 1500–1511. https://doi.org/10.1016/j.bbabio.2010.04.004

Lai, Y.-T., Yeung, C. K. L., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., Hung, C.-M., Hung, H.-Y., Yang, M.-Y., Liang, W., Hsu, Y.-C., Yao, C.-T., Dong, L.,

Lin, K., & Li, S.-H. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, *116*(6), 2152–2157. https://doi.org/10.1073/pnas.1813597116

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv:1303.3997 [q-Bio]*. http://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., & 1000 Genome Project Data Processing Subgroup. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.

Libiad, M., Yadav, P. K., Vitvitsky, V., Martinov, M., & Banerjee, R. (2014). Organization of the Human Mitochondrial Hydrogen Sulfide Oxidation Pathway * ♦. *Journal of Biological Chemistry*, *289*(45), 30901–30910. https://doi.org/10.1074/jbc.M114.602664

Linnen, C. R., Kingsley, E. P., Jensen, J. D., & Hoekstra, H. E. (2009). On the Origin and Spread of an Adaptive Allele in Deer Mice. *Science*, *325*(5944), 1095–1098. https://doi.org/10.1126/science.1175826

Liu, Y., Cotton, J. A., Shen, B., Han, X., Rossiter, S. J., & Zhang, S. (2010). Convergent sequence evolution between echolocating bats and dolphins. *Current Biology*, *20*(2), R53–R54. https://doi.org/10.1016/j.cub.2009.11.058

Losos, J. B. (2011). Convergence, Adaptation, and Constraint. *Evolution*, *65*(7), 1827–1840. https://doi.org/10.1111/j.1558-5646.2011.01289.x

Mahalanobis, P. C. (1936). On the Generalized Distance in Statistics. Journal of Genetics. *Journal of Genetics*, *41*, 159–193.

678 Martin, A., & Orgogozo, V. (2013). THE LOCI OF REPEATED EVOLUTION: A CATALOG OF GENETIC
679 HOTSPOTS OF PHENOTYPIC VARIATION. *Evolution*, *67*(5), 1235–1250.
680 https://doi.org/10.1111/evo.12081
681 Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing
682 reads. *EMBnet.Journal*, *17*(1), Article 1. https://doi.org/10.14806/ej.17.1.200
683 Mayr, C. (2019). What Are 3' UTRs Doing? *Cold Spring Harbor Perspectives in Biology*, *11*(10),
684 a034728. https://doi.org/10.1101/cshperspect.a034728
685 Minh, B. Q., Nguyen, M. A. T., & von Haeseler, A. (2013). Ultrafast Approximation for
686 Phylogenetic Bootstrap. *Molecular Biology and Evolution*, *30*(5), 1188–1195.
687 https://doi.org/10.1093/molbev/mst024
688 Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., von Haeseler, A., &
689 Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic
690 Inference in the Genomic Era. *Molecular Biology and Evolution*, *37*(5), 1530–1534.
691 https://doi.org/10.1093/molbev/msaa015
692 Mustafa, A. K., Gadalla, M. M., Sen, N., Kim, S., Mu, W., Gazi, S. K., Barrow, R. K., Yang, G.,
693 Wang, R., & Snyder, S. H. (2009). H2S signals through protein S-sulfhydration. *Science*
694 *Signaling*, *2*(96), ra72. https://doi.org/10.1126/scisignal.2000464
695 Natarajan, C., Hoffmann, F. G., Weber, R. E., Fago, A., Witt, C. C., & Storz, J. F. (2016).
696 Predictable convergence in hemoglobin function has unpredictable molecular
697 underpinnings. *Science*, *354*(6310), 336–339. https://doi.org/10.1126/science.aaf9070

698 Nei M & Li W H. (1979). Mathematical model for studying genetic variation in terms of
699 restriction endonucleases. *Proceedings of the National Academy of Sciences*, *76*(10),
700 5269–5273. https://doi.org/10.1073/pnas.76.10.5269
701 Olson, K. R. (2018). H2S and polysulfide metabolism: Conventional and unconventional
702 pathways. *Biochemical Pharmacology*, *149*, 77–90.
703 https://doi.org/10.1016/j.bcp.2017.12.010
704 Pagliarini, D. J., Calvo, S. E., Chang, B., Sheth, S. A., Vafai, S. B., Ong, S.-E., Walford, G. A.,
705 Sugiana, C., Boneh, A., Chen, W. K., Hill, D. E., Vidal, M., Evans, J. G., Thorburn, D. R.,
706 Carr, S. A., & Mootha, V. K. (2008). A mitochondrial protein compendium elucidates
707 complex I disease biology. *Cell*, *134*(1), 112–123.
708 https://doi.org/10.1016/j.cell.2008.06.016
709 Palacios, M., Arias-Rodriguez, L., Plath, M., Eifert, C., Lerp, H., Lamboj, A., Voelker, G., & Tobler,
710 M. (2013). The Rediscovery of a Long Described Species Reveals Additional Complexity
711 in Speciation Patterns of Poeciliid Fishes in Sulfide Springs. *PLOS ONE*, *8*(8), e71069.
712 https://doi.org/10.1371/journal.pone.0071069
713 Paradis, E. (2010). pegas: An R package for population genetics with an integrated–modular
714 approach. *Bioinformatics*, *26*(3), 419–420.
715 https://doi.org/10.1093/bioinformatics/btp696
716 Paradis, E., & Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and
717 evolutionary analyses in R. *Bioinformatics*, *35*(3), 526–528.
718 https://doi.org/10.1093/bioinformatics/bty633

719 Passow, C. N., Brown, A. P., Arias-Rodriguez, L., Yee, M.-C., Sockell, A., Schartl, M., Warren, W.
720 C., Bustamante, C., Kelley, J. L., & Tobler, M. (2017). Complexities of gene expression
721 patterns in natural populations of an extremophile fish (Poecilia mexicana, Poeciliidae).
722 *Molecular Ecology*, *26*(16), 4211–4225. https://doi.org/10.1111/mec.14198
723 Passow, C. N., Henpita, C., Shaw, J. H., Quackenbush, C. R., Warren, W. C., Schartl, M., Arias-
724 Rodriguez, L., Kelley, J. L., & Tobler, M. (2017). The roles of plasticity and evolutionary
725 change in shaping gene expression variation in natural populations of extremophile fish.
726 *Molecular Ecology*, *26*(22), 6384–6399. https://doi.org/10.1111/mec.14360
727 Pfenninger, M., Lerp, H., Tobler, M., Passow, C., Kelley, J. L., Funke, E., Greshake, B., Erkoc, U. K.,
728 Berberich, T., & Plath, M. (2014). Parallel evolution of cox genes in H2S-tolerant fish as
729 key adaptation to a toxic environment. *Nature Communications*, *5*(1), 3873.
730 https://doi.org/10.1038/ncomms4873
731 Pfenninger, M., Patel, S., Arias-Rodriguez, L., Feldmeyer, B., Riesch, R., & Plath, M. (2015).
732 Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic
733 habitats of a neotropical fish (Poecilia mexicana). *Molecular Ecology*, *24*(21), 5446–5459.
734 https://doi.org/10.1111/mec.13397
735 Pietri, R., Román-Morales, E., & López-Garriga, J. (2011). Hydrogen sulfide and hemeproteins:
736 Knowledge and mysteries. *Antioxidants & Redox Signaling*, *15*(2), 393–404.
737 https://doi.org/10.1089/ars.2010.3698
738 Plath, M., Pfenninger, M., Lerp, H., Riesch, R., Eschenbrenner, C., Slattery, P. A., Bierbach, D.,
739 Herrmann, N., Schulte, M., Arias-Rodriguez, L., Rimber Indy, J., Passow, C., & Tobler, M.
740 (2013). Genetic differentiation and selection against migrants in evolutionarily

741 replicated extreme environments. *Evolution; International Journal of Organic Evolution*,
742 *67*(9), 2647–2661. https://doi.org/10.1111/evo.12133
743 Puritz, J. B., Hollenbeck, C. M., & Gold, J. R. (2014). dDocent: A RADseq, variant-calling pipeline
744 designed for population genomics of non-model organisms. *PeerJ*, *2*, e431.
745 https://doi.org/10.7717/peerj.431
746 Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014).
747 Demystifying the RAD fad. *Molecular Ecology*, *23*(24), 5937–5942.
748 https://doi.org/10.1111/mec.12965
749 Quinlan, A. R., & Hall, I. M. (2010). BEDTools: A flexible suite of utilities for comparing genomic
750 features. *Bioinformatics*, *26*(6), 841–842.
751 https://doi.org/10.1093/bioinformatics/btq033
752 Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For
753 Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*(1), 357–
754 383. https://doi.org/10.1146/annurev-genom-111720-081402
755 Reid, N. M., Proestou, D. A., Clark, B. W., Warren, W. C., Colbourne, J. K., Shaw, J. R., Karchner,
756 S. I., Hahn, M. E., Nacci, D., Oleksiak, M. F., Crawford, D. L., & Whitehead, A. (2016). The
757 genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild
758 fish. *Science*, *354*(6317), 1305–1308. https://doi.org/10.1126/science.aah4993
759 Riesch, R., Plath, M., García de León, F. J., & Schlupp, I. (2010). Convergent life-history shifts:
760 Toxic environments result in big babies in two clades of poeciliids. *Die*
761 *Naturwissenschaften*, *97*(2), 133–141. https://doi.org/10.1007/s00114-009-0613-y

762 Riesch, R., Plath, M., Schlupp, I., Tobler, M., & Brian Langerhans, R. (2014). Colonisation of toxic
763 environments drives predictable life-history evolution in livebearing fishes (Poeciliidae).
764 *Ecology Letters*, *17*(1), 65–71. https://doi.org/10.1111/ele.12209
765 Riesch, R., Tobler, M., Lerp, H., Jourdan, J., Doumas, T., Nosil, P., Langerhans, R. B., & Plath, M.
766 (2016). Extremophile Poeciliidae: Multivariate insights into the complexity of speciation
767 along replicated ecological gradients. *BMC Evolutionary Biology*, *16*(1), 136.
768 https://doi.org/10.1186/s12862-016-0705-1
769 Rosenblum, E. B., Parent, Christine E., & Brandt, E. E. (2014). The Molecular Basis of Phenotypic
770 Convergence. *Annu. Rev. Ecol. Evol. Syst.*, *45*, 203–226.
771 Rossiter, S. J., Zhang, S., & Liu, Y. (2011). Prestin and high frequency hearing in mammals.
772 *Communicative & Integrative Biology*, *4*(2), 236–239.
773 https://doi.org/10.4161/cib.4.2.14647
774 Schluter, D., & Conte, G. L. (2009). Genetics and ecological speciation. *Proceedings of the*
775 *National Academy of Sciences*, *106*(Supplement 1), 9955–9962.
776 https://doi.org/10.1073/pnas.0901264106
777 Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski,
778 B., & Ideker, T. (2003). Cytoscape: A Software Environment for Integrated Models of
779 Biomolecular Interaction Networks. *Genome Research*, *13*(11), 2498–2504.
780 https://doi.org/10.1101/gr.1239303
781 Stern, D. L. (2013). The genetic causes of convergent evolution. *Nature Reviews. Genetics*,
782 *14*(11), 751–764. https://doi.org/10.1038/nrg3483

783 Templeton, A. R., Crandall, K. A., & Sing, C. F. (1992). A cladistic analysis of phenotypic
784 associations with haplotypes inferred from restriction endonuclease mapping and DNA
785 sequence data. III. Cladogram estimation. *Genetics*, *132*(2), 619–633.
786 https://doi.org/10.1093/genetics/132.2.619
787 Tobler, M., & Hastings, L. (2011). Convergent Patterns of Body Shape Differentiation in Four
788 Different Clades of Poeciliid Fishes Inhabiting Sulfide Springs. *Evolutionary Biology*,
789 *38*(4), 412–421. https://doi.org/10.1007/s11692-011-9129-4
790 Tobler, M., Henpita, C., Bassett, B., Kelley, J. L., & Shaw, J. H. (2014). H2S exposure elicits
791 differential expression of candidate genes in fish adapted to sulfidic and non-sulfidic
792 environments. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative*
793 *Physiology*, *175*, 7–14. https://doi.org/10.1016/j.cbpa.2014.04.012
794 Tobler, M., Kelley, J. L., Plath, M., & Riesch, R. (2018). Extreme environments and the origins of
795 biodiversity: Adaptation and speciation in sulphide spring fishes. *Molecular Ecology*,
796 *27*(4), 843–859. https://doi.org/10.1111/mec.14497
797 Tobler, M., Palacios, M., Chapman, L. J., Mitrofanov, I., Bierbach, D., Plath, M., Arias-Rodriguez,
798 L., García de León, F. J., & Mateos, M. (2011). Evolution in Extreme Environments:
799 Replicated Phenotypic Differentiation in Livebearing Fish Inhabiting Sulfidic Springs.
800 *Evolution*, *65*(8), 2213–2228. https://doi.org/10.1111/j.1558-5646.2011.01298.x
801 Tobler, M., Passow, C. N., Greenway, R., Kelley, J. L., & Shaw, J. H. (2016). The Evolutionary
802 Ecology of Animals Inhabiting Hydrogen Sulfide–Rich Environments. *Annual Review of*
803 *Ecology, Evolution, and Systematics*, *47*(1), 239–262. https://doi.org/10.1146/annurev-
804 ecolsys-121415-032418

Tobler, M., Riesch, R., & Plath, M. (2015). Extremophile Fishes: An Integrative Synthesis. In R. Riesch, M. Tobler, & M. Plath (Eds.), *Extremophile Fishes: Ecology, Evolution, and Physiology of Teleosts in Extreme Environments* (pp. 279–296). Springer International Publishing. https://doi.org/10.1007/978-3-319-13362-1_12

Todesco, M., Owens, G. L., Bercovich, N., Légaré, J.-S., Soudi, S., Burge, D. O., Huang, K., Ostevik, K. L., Drummond, E. B. M., Imerovski, I., Lande, K., Pascual-Robles, M. A., Nanavati, M., Jahani, M., Cheung, W., Staton, S. E., Muños, S., Nielsen, R., Donovan, L. A., … Rieseberg, L. H. (2020). Massive haplotypes underlie ecotypic differentiation in sunflowers. *Nature*, *584*(7822), 602–607. https://doi.org/10.1038/s41586-020-2467-6

Van der Auwera, G. A., & O'Connor, B. D. (2020). *Genomics in the Cloud – Using Docker, GATK, and WDL in Terra*. O'Reilly Media, Inc. https://broadinstitute.github.io/genomics-in-the-cloud/

Verity, R., Collins, C., Card, D. C., Schaal, S. M., Wang, L., & Lotterhos, K. E. (2017). minotaur: A platform for the analysis and visualization of multivariate results from genome scans with R Shiny. *Molecular Ecology Resources*, *17*(1), 33–43. https://doi.org/10.1111/1755-0998.12579

Warren, W. C., García-Pérez, R., Xu, S., Lampert, K. P., Chalopin, D., Stöck, M., Loewe, L., Lu, Y., Kuderna, L., Minx, P., Montague, M. J., Tomlinson, C., Hillier, L. W., Murphy, D. N., Wang, J., Wang, Z., Garcia, C. M., Thomas, G. C. W., Volff, J.-N., … Schartl, M. (2018). Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nature Ecology & Evolution*, *2*(4), 669–679. https://doi.org/10.1038/s41559-018-0473-y

827 Warrington, J. A., Nair, A., Mahadevappa, M., & Tsyganskaya, M. (2000). Comparison of human
828 adult and fetal expression and identification of 535 housekeeping/maintenance genes.
829 *Physiological Genomics*, *2*(3), 143–147.
830 https://doi.org/10.1152/physiolgenomics.2000.2.3.143
831 Waters, J. M., & McCulloch, G. A. (2021). Reinventing the wheel? Reassessing the roles of gene
832 flow, sorting and convergence in repeated evolution. *Molecular Ecology*, *30*(17), 4162–
833 4172. https://doi.org/10.1111/mec.16018
834 Weinreich, D. M., Delaney, N. F., DePristo, M. A., & Hartl, D. L. (2006). Darwinian Evolution Can
835 Follow Only Very Few Mutational Paths to Fitter Proteins. *Science*, *312*(5770), 111–114.
836 https://doi.org/10.1126/science.1123539
837 Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population
838 Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641
839 Xu, S., Wang, J., Guo, Z., He, Z., & Shi, S. (2020). Genomic Convergence in the Adaptation to
840 Extreme Environments. *Plant Communications*, *1*(6), 100117.
841 https://doi.org/10.1016/j.xplc.2020.100117
842

Tables and Figures

Table 1: The number of genes, exons, and sites captured by targeted sequencing and number of filtered sites, single nucleotide polymorphisms (SNPs), and linkage-disequilibrium pruned SNPs used in this analysis.

	Genes	Exons	Targeted Sites	Filtered Sites	SNPs	Pruned SNPs
Target						
Sulfide	166	1330	421,765	398,664	5,153	3,808
OxPhos	84	432	96,472	88,421	973	787
Background	165	1,341	321,091	297,544	3,320	2,682
Total	415	3,103	839,328	784,629	9,446	7,277

Figure 1: Map of study region and sites. Samples were collected from 10 sites in the Río Grijalva basin. The study site location is indicated by a yellow star in the insert map of Mexico. Shape represents drainage and color represents ecotype. This figure was adapted from Hotaling et al., 2019.

Figure 2: Analysis of population structure in background, sulfide processing, and OxPhos genes. A) Best *K* from admixture analyses of each gene set ordered by drainage, from west to east. B) PCA of unlinked SNPs for each gene set. Color represents ecotype (sulfidic in yellow, nonsulfidic in blue) and shape represents drainage of origin.

Figure 3: Tree discordance between maximum likelihood tree of background, sulfide processing, and OxPhos genes. Color represents ecotype (sulfidic in yellow, nonsulfidic in blue). Bootstrap support for populations splits shown as a percent of 1000 bootstraps.

Figure 4: Evidence for a monophyletic origin of regions putatively under selection in sulfidic populations. A) Per base pair F_{ST} between all sulfidic and all nonsulfidic individuals colored by gene set: H_2S detoxification (gold), OxPhos (red) and background (grey). Outlier cutoffs are indicated by horizontal lines (solid line 99.5%, dashed line 99.9%). B) Haplotype network for outlier genes putatively under selection *ethe1.a* (left) and *sqor* (right). Yellow shades represent sulfidic individuals' haplotype and blues represents nonsulfidic individuals' haplotype. Shade represents population, with the lightest shade representing western populations and darker shades represents eastern

872 populations (see Figure 1). Node size represents number of haplotypes found in the
873 dataset. Number of mutations between haplotypes is labeled on the branch as tick
874 marks. Note: singletons have been removed.
875