



**Cite this article:** Perry BW, McGowan KL, Arias-Rodriguez L, Duttke SH, Tobler M, Kelley JL. 2024 Nascent transcription reveals regulatory changes in extremophile fishes inhabiting hydrogen sulfide-rich environments. *Proc. R. Soc. B* **291**: 20240412. <https://doi.org/10.1098/rspb.2024.0412>

Received: 20 February 2024

Accepted: 3 May 2024

**Subject Category:**

Genetics and genomics

**Subject Areas:**

evolution, genomics, bioinformatics

**Keywords:**

capped-small RNA-sequencing, transcriptomics, transcription factor, *Poecilia mexicana*, adaptation

**Author for correspondence:**

Joanna L. Kelley

e-mail: [jokelley@ucsc.edu](mailto:jokelley@ucsc.edu)

<sup>†</sup>Authors contributed equally.

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.7247128>.

# Nascent transcription reveals regulatory changes in extremophile fishes inhabiting hydrogen sulfide-rich environments

Blair W. Perry<sup>1,†</sup>, Kerry L. McGowan<sup>1,†</sup>, Lenin Arias-Rodriguez<sup>3</sup>,  
Sascha H. Duttke<sup>2</sup>, Michael Tobler<sup>4,5,6</sup> and Joanna L. Kelley<sup>7</sup>

<sup>1</sup>School of Biological Sciences, and <sup>2</sup>School of Molecular Biosciences, Washington State University, Pullman, WA 99164, USA

<sup>3</sup>División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco, Villahermosa, Tabasco 86150, México

<sup>4</sup>Department of Biology, and <sup>5</sup>Whitney R. Harris World Ecology Center, University of Missouri—St Louis, St Louis, MO 63121, USA

<sup>6</sup>WildCare Institute, Saint Louis Zoo, St Louis, MO 63110, USA

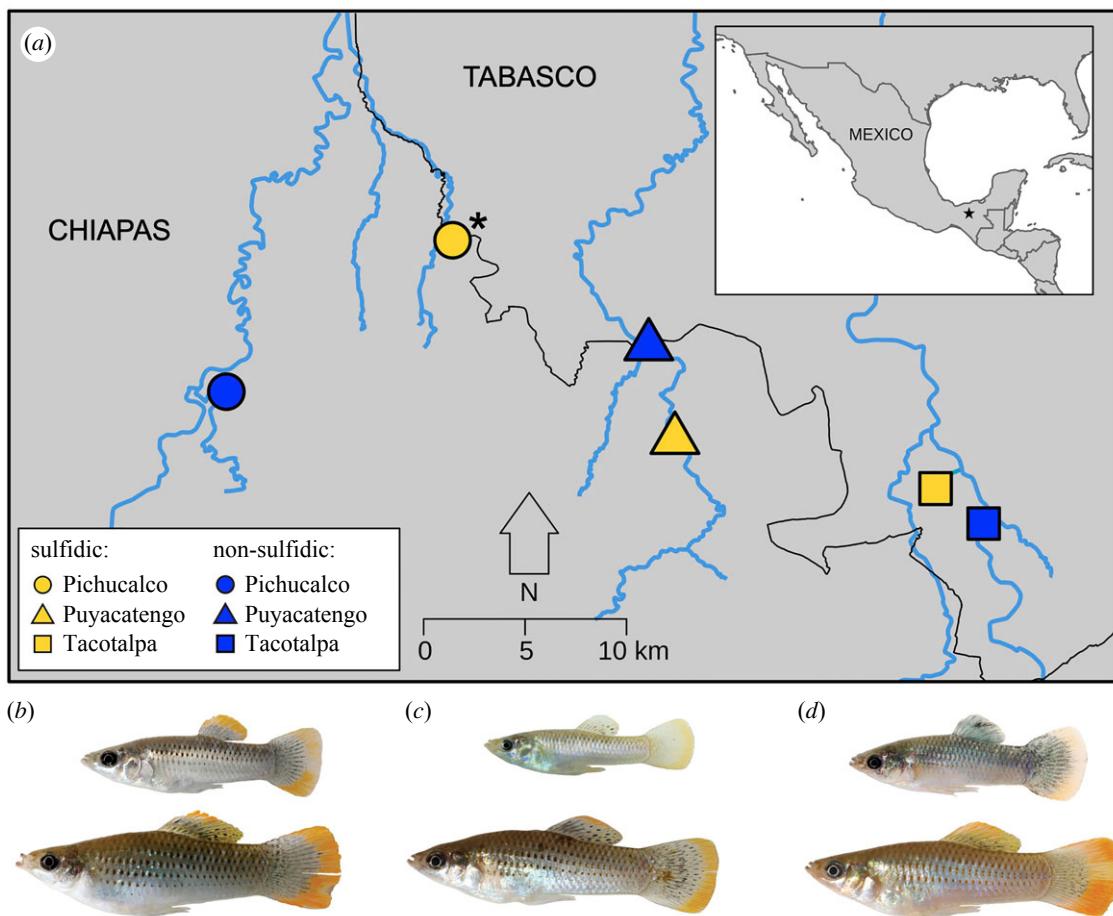
<sup>7</sup>Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, CA 95060, USA

BWP, [0000-0003-4905-7852](https://orcid.org/0000-0003-4905-7852); KLM, [0000-0001-6388-3873](https://orcid.org/0000-0001-6388-3873); JLK, [0000-0002-7731-605X](https://orcid.org/0000-0002-7731-605X)

Regulating transcription allows organisms to respond to their environment, both within a single generation (plasticity) and across generations (adaptation). We examined transcriptional differences in gill tissues of fishes in the *Poecilia mexicana* species complex (family Poeciliidae), which have colonized toxic springs rich in hydrogen sulfide (H<sub>2</sub>S) in southern Mexico. There are gene expression differences between sulfidic and non-sulfidic populations, yet regulatory mechanisms mediating this gene expression variation remain poorly studied. We combined capped-small RNA sequencing (csRNA-seq), which captures actively transcribed (i.e. nascent) transcripts, and messenger RNA sequencing (mRNA-seq) to examine how variation in transcription, enhancer activity, and associated transcription factor binding sites may facilitate adaptation to extreme environments. csRNA-seq revealed thousands of differentially initiated transcripts between sulfidic and non-sulfidic populations, many of which are involved in H<sub>2</sub>S detoxification and response. Analyses of transcription factor binding sites in promoter and putative enhancer csRNA-seq peaks identified a suite of transcription factors likely involved in regulating H<sub>2</sub>S-specific shifts in gene expression, including several key transcription factors known to respond to hypoxia. Our findings uncover a complex interplay of regulatory processes that reflect the divergence of extremophile populations of *P. mexicana* from their non-sulfidic ancestors and suggest shared responses among evolutionarily independent lineages.

## 1. Introduction

Changes in gene expression can facilitate the alteration of phenotypes. There are many examples of transcriptional differences underlying dramatic phenotypic changes, ranging from tissue regeneration [1] and diapause [2] to stress responses [3]. While evolutionary biologists have long focused on nucleotide changes in protein-coding genes to explain phenotypic evolution [4,5], foundational work has also shown that changes in the expression of protein-coding genes can drive evolutionary change and mediate adaptation [6–8]. Well known examples of adaptive gene expression changes include high-altitude muscle function [9] and coat camouflage for predator avoidance [10]. However, the regulatory mechanisms that drive changes in gene expression are often complex and difficult to study, and therefore remain poorly understood. A better understanding of gene regulatory networks and the molecular mechanisms underlying changes in gene expression is critical to help explain patterns of adaptation and plasticity [11].



**Figure 1.** (a) Map of sampling sites in the Pichucalco, Puyacatengo and Tacotalpa drainages in the Río Grijalva basin in southern Mexico. Shapes indicate drainage of origin. Colours indicate whether sites were sulfidic (yellow) or non-sulfidic (blue). In the Pichucalco drainages, sulfidic stream sites (marked with \*) are inhabited by the endemic *Poecilia sulphuraria* species, while the non-sulfidic streams in this drainage are inhabited by *Poecilia mexicana*. Both sulfidic and non-sulfidic streams in the Puyacatengo and Tacotalpa drainages are inhabited by *P. mexicana*. (b–d) Fish sampled from sulfidic (top) and non-sulfidic (bottom) streams in the (b) Pichucalco, (c) Puyacatengo and (d) Tacotalpa drainages. Fish not to scale. Photo credit: Michael Tobler.

Gene expression is largely regulated by activity in non-coding genomic regions and often involves the transcription of different types of RNA, including microRNAs (miRNAs) [12], RNA transcripts from promoter, terminal and enhancer regions [13,14] and long non-coding RNAs [15]. Additionally, the binding of transcription factor (TF) proteins in promoters and enhancers initiates and regulates transcription, such that changes in TF binding sites and TF activity can have dramatic effects on gene expression and ultimately emerging phenotypes [16,17]. Understanding the impact of regulatory activity is therefore key to explaining how responses to an environment occur at the genomic level, and how they might evolve.

Species in which variation in gene expression has been linked to adaptation offer opportunities to study mechanisms of gene regulation. Fishes in the *Poecilia mexicana* (Atlantic molly) species complex (family Poeciliidae) are one such example. *Poecilia mexicana* is a livebearing fish common in freshwater habitats throughout Mexico and Central America [18]. In southern Mexico, multiple lineages have independently colonized toxic springs rich in H<sub>2</sub>S (figure 1a; [19]). Sulfide spring populations (i.e. sulfidic ecotypes) are locally adapted and are phenotypically and genetically distinct from adjacent populations in non-sulfidic habitats [20–22]. H<sub>2</sub>S is toxic to most animals, inhibiting enzymes in the electron transport chain and lethally interfering with aerobic ATP production via oxidative phosphorylation [23,24]. H<sub>2</sub>S also reacts with dissolved oxygen in aquatic environments, creating additionally stressful hypoxic conditions [25,26]. The ability to modulate gene expression plays a major role in survival in these H<sub>2</sub>S-rich environments, with sulfidic populations showing convergent shifts in the expression of key genes involved in H<sub>2</sub>S detoxification and sulfur metabolism [21,22,27,28]. Many of these changes involve genes in the sulfide:quinone oxidoreductase (SQR) pathway, which endogenously detoxifies H<sub>2</sub>S [29]. This pathway is present in all metazoans and likely a bacterial relic from the Proterozoic aeon [30]. The expression of genes involved in the SQR pathway is consistently upregulated in sulfidic populations, as are sulfur transporter genes, and genes encoding subunits of cytochrome *c* oxidase (COX, the main toxicity target of H<sub>2</sub>S) [21,22,27,28]. Some of these genes show evolved, constitutive expression differences in sulfidic populations compared with non-sulfidic populations, whereas other genes show evolved plasticity in sulfidic populations in response to H<sub>2</sub>S [31]. However, aside from some evidence that miRNAs and DNA methylation changes affect the expression levels of many H<sub>2</sub>S-related genes in one population [32,33], the underlying regulatory processes responsible for the changes in gene expression observed in this system remain underexplored. Examining these mechanisms will provide a better understanding of the regulatory changes underlying parallel adaptation in these fishes.

To examine regulatory differences among populations in the *P. mexicana* species complex, we used capped-small RNA-sequencing (csRNA-seq), a nascent RNA sequencing method that captures short, newly initiated transcripts (approx. 20–60 nucleotides in length) [34]. In contrast to traditional RNA-sequencing approaches (i.e. RNA-sequencing with either messenger RNA (mRNA) or

ribosomal depletion methods) which capture only the mature, stable transcripts of genes, nascent RNA-sequencing methods like csRNA-seq instead capture all initiated transcripts, including unstable transcripts [34–37]. These unstable RNAs (i.e. enhancer RNAs, miRNAs, antisense transcripts and other regulatory RNAs) persist for only a short time in the cell but can have important effects on gene regulation [38–40]. csRNA-seq can pinpoint transcriptional changes in stable and unstable regulatory RNAs and identify initiation sites in the genome to base-pair (bp) resolution [34]. Capturing this positional information and identifying shared changes in regulatory RNA transcription across independently adapted populations will reveal how regulatory mechanisms alter gene expression to enable survival of *P. mexicana* in sulfidic environments. Given that sulfidic lineages are often genetically distinct from adjacent non-sulfidic populations [20–22], shared differences across independent lineages may be indicative of adaptation [41]. More broadly, this approach will provide valuable new insight into the mechanisms by which gene expression regulation may evolve in response to selection.

In this study, we sampled three population pairs in the *P. mexicana* species complex comprising both ancestral non-sulfidic and corresponding H<sub>2</sub>S-adapted lineages, and we investigated differences in patterns of nascent transcription between ecologically contrasting populations. Specifically, we asked the following questions: (1) What regulatory mechanisms underlie differences in gene expression in sulfidic populations compared with non-sulfidic populations? (2) What TF binding sites are overrepresented near differentially initiated transcription start sites of genes with gene expression shifts in sulfidic fish? (3) Are the genes coding for TFs with enriched binding sites differentially expressed between ecotypes? (4) Are differences in gene expression between ecotypes driven in part by differential activity of enhancer regions? To answer these questions, we first investigated variation across two distinct aspects of gene expression. In addition to capturing variation in transcript initiation with csRNA-seq, we also quantified shifts in gene expression using mature, stable mRNAs from coding regions using previously published RNA-sequencing (mRNA-seq) data [27]. We identified unique patterns of transcript initiation in sulfidic populations that were associated with H<sub>2</sub>S detoxification and identified TFs that may be particularly important in controlling gene expression in sulfidic environments. These analyses highlight regulatory variation between H<sub>2</sub>S-adapted populations and ancestral non-sulfidic populations and shed light on regulatory mechanisms that underlie changes to some of the most conserved metabolic pathways in metazoans, including H<sub>2</sub>S detoxification and oxidative phosphorylation.

## 2. Methods

### (a) Sample collection

Adult female fish were collected via seine net from sulfidic and non-sulfidic sites in the Pichucalco, Puyacatengo and Tacotalpa drainages of the Río Grijalva basin in southern Mexico ( $N = 5$  or 6 per site; electronic supplementary material, table S1) as part of a previous study by Kelley *et al.* [27]. In the Pichucalco drainage, sulfidic springs are inhabited by the highly endemic extremophile *Poecilia sulphuraria* (sulphur molly; figure 1b, top) [42], but non-sulfidic springs are populated by *P. mexicana* (figure 1b, bottom). In the Puyacatengo and Tacotalpa drainages, *P. mexicana* inhabits both the sulfidic and non-sulfidic sites (figure 1c,d). Fish were sacrificed, and their gill arches were extracted, preserved in RNAlater (Invitrogen), and stored at  $-80^{\circ}\text{C}$ . Gill tissues were selected because they are directly exposed to H<sub>2</sub>S in the water and were previously shown to have significant gene expression differences between the sulfidic and non-sulfidic ecotypes [27,31].

While inter-drainage variation in adaptation to H<sub>2</sub>S has been demonstrated in *Poecilia* [4,22], we were most interested in the broad patterns of gene regulation across ecotypes, as there is pre-established convergence in the differential expression of H<sub>2</sub>S-related genes in all sulfidic populations [27,31]. For this reason, we compared gill tissues sampled from all sulfidic fish with those from all non-sulfidic fish combined across drainages.

### (b) csRNA library preparation and sequencing

csRNA-seq was generated from a subset of the fish samples described above ( $N = 2$  fish per site; electronic supplementary material, table S1). Total RNA was extracted from gill tissues using Qiagen's miRNeasy Mini Kit following all standard protocols. Total RNA concentrations were estimated with the Qubit RNA HS Assay Kit and the Agilent 2100 BioAnalyzer using the RNA 6000 Nano Kit. Short RNAs (approx. 20–60 nucleotides) were isolated using size selection with gel electrophoresis. For each sample, a 10% aliquot of the size-selected RNA suspension was reserved for use as an 'input library' [43]. These aliquots were sequenced and used in the HOMER pipeline to control for exonic contamination when identifying transcription initiation sites [34]. For the remaining size-selected RNA for each sample, 5' 7-methylguanosine-capped RNA was isolated following protocols from Duttke *et al.* [34], and this constituted our csRNA libraries. csRNA and the reserved input aliquots were converted into cDNA libraries for each sample, amplified with PCR, size-selected to remove primer dimers, purified with gel electrophoresis and pooled, and single-end 75 bp reads were sequenced using an Illumina NextSeq 500.

### (c) Analysis of differential transcription initiation using csRNA-seq

We used FastQC (v. 0.11.9) to examine the quality of the raw csRNA and input RNA reads [44]. Subsequent analyses were carried out using the csRNA-seq analysis pipeline in HOMER (v. 4.11) [34]. We used HOMER's *trim* to remove Illumina TruSeq adapters from the 3' end of reads, discarded reads less than 20 bp, and trimmed the ends of reads below a Phred score of 20. We indexed the *P. mexicana* reference genome (GenBank assembly accession: GCF\_001443325.1) [45] using STAR (v. 2.7.6a) and then aligned the reads with STAR, outputting one primary alignment per read [46]. Tag directories were created for the uniquely mapped csRNA libraries and the input libraries for each sample using HOMER's *makeTagDirectory*, including the *--single* flag to account for the large number of scaffolds in the reference genome and *--fragLength* set to 30 bp to reflect the average length of the nascent RNAs captured by the csRNA-seq protocol [34]. A third tag directory was created for each sample using the mapped mRNA libraries (described below) and used to control for false positives and exonic contaminants in the next step. The csRNA, input and mRNA tag directories were input into HOMER's *findcsRNATSS.pl* to identify sites of transcription initiation, hereafter called peaks. This step of the HOMER pipeline also determines the stability of initiated transcripts by assessing whether stable mRNA transcripts map within  $-100$  to  $+500$  bp of each csRNA-seq peak.

Unstable, nascent transcripts are not processed into stable mRNAs and are thus determined by the lack of stable mRNA transcripts mapping to the adjacent region (see [34] for additional detail). We retained peaks with at least seven csRNA-sequencing reads per 10 million reads for further analysis following the default parameters. Peaks were annotated using the GTF file for *P. mexicana* (GenBank assembly accession: GCF\_001443325.1, converted from GFF using *gffread* (v. 0.9.9)). HOMER's *mergePeaks* was used to create a non-redundant list of peaks across all samples, condensing peaks that directly overlapped. The merged peaks file was used with *homerTools annotatePeaks.pl* and the *--strand + --fragLength 1* options required for csRNA-seq to quantify the raw read counts for each sample, associate each peak with the nearest annotated gene from the GFF, and identify the region of the genome they mapped to relative to this gene (i.e. in an exon, intergenic, intron, promoter or transcription start site (promoter-TSS), or transcription termination site (TTS)). The promoter-TSS region is defined as the region 1000 bp upstream to 100 bp downstream of a gene's annotated TSS in the GTF file. We identified differentially initiated peaks in R using *edgeR* (v. 4.0.12) [47] by comparing all sulfidic samples with all non-sulfidic samples while including drainage and species as covariates. Significantly differentially initiated peaks were defined as those with a false discovery rate (FDR)  $<0.05$ . Variation in transcription initiation across samples was plotted using a principal components analysis (PCA), using the 500 peaks with the highest average signal across all samples using the R package *limma* (v. 3.46.0). Hierarchically clustered heatmaps of csRNA-seq peaks were generated with the R package *pheatmap* (v. b1.0.12) using the top 5000 peaks based on overall number of counts across samples and subsequently with only significantly differentially initiated peaks.

We used a curated list of genes with Gene Ontology (GO) terms related to  $\text{H}_2\text{S}$  detoxification or response (GO:0070221, GO:0006790, GO:0044273, GO:0070813, GO:0000096, GO:0006090, GO:0006749, GO:1901687, GO:0000098, GO:0008272, GO:0006534, GO:0044272) to carry out a Fisher's exact test using *fisher.test()* in base R (v. 4.0.3), which tested whether peaks in or near genes related to  $\text{H}_2\text{S}$  detoxification and response were overrepresented in the list of significantly differentially initiated peaks.

#### (d) mRNA library preparation and sequencing

Samples were prepared and sequenced as part of a previous study by Kelley *et al.* [27]. In brief, total RNA was extracted from gill tissues ( $N = 17$  sulfidic fish,  $N = 18$  non-sulfidic fish; electronic supplementary material, table S1), poly-A mRNA was isolated, and libraries were sequenced using an Illumina HiSeq 2000 using paired-end 101 bp reads.

#### (e) mRNA differential gene expression

The quality of all raw reads was assessed using FastQC. Trim Galore! (v. 0.4.2) was used to remove the first 11 bp from the 5' ends of reads (*--clip\_R1 11 --clip\_R2 11*) and perform adapter trimming (*--stringency 6*) but not quality trimming (*--quality 0*) [48]. A second round of Trim Galore! was used to quality trim the ends of reads with a Phred score below 24 (*--quality 24*) and remove reads less than 50 bp long (*--length 50*). HISAT2 *hisat2-build* (v. 2.1.0) was used to index the *P. mexicana* reference genome with the mitochondrial genome appended (GenBank accession: KC992995.1) [4]. Trimmed reads were mapped using the *--downstream-transcriptome-assembly* flag for compatibility with StringTie (v. 2.0.3), *-fr* specifying the mate pair orientation, and read group information included for sample, read group identifier and platform [49]. SAMtools *view* (v. 1.9) was used to convert the resulting SAM files to BAM files, *sort* to sort the BAM files by coordinate, and *merge* to combine samples with more than one technical replicate [50]. StringTie (v. 2.0.3) with the *--eB* flag was used to generate read coverage tables in GTF format for each sample [51]. An edited version of the *prepDE.py* script from StringTie (see code repository) was used to generate gene count matrices, resulting in counts for nuclear genes only. Genes were filtered by a minimum counts-per-million (cpm) of at least 5 in the smallest library and to have non-zero counts in at least five samples. Using *edgeR* (v. 4.0.12), libraries were normalized using the trimmed mean of *M*-values (TMM, the default) and quasi-likelihood *F*-tests were used to compare differential gene expression in sulfidic samples compared with non-sulfidic samples, with species and drainage included as covariates [47,52].

#### (f) Nascent transcription near regions with stable gene expression

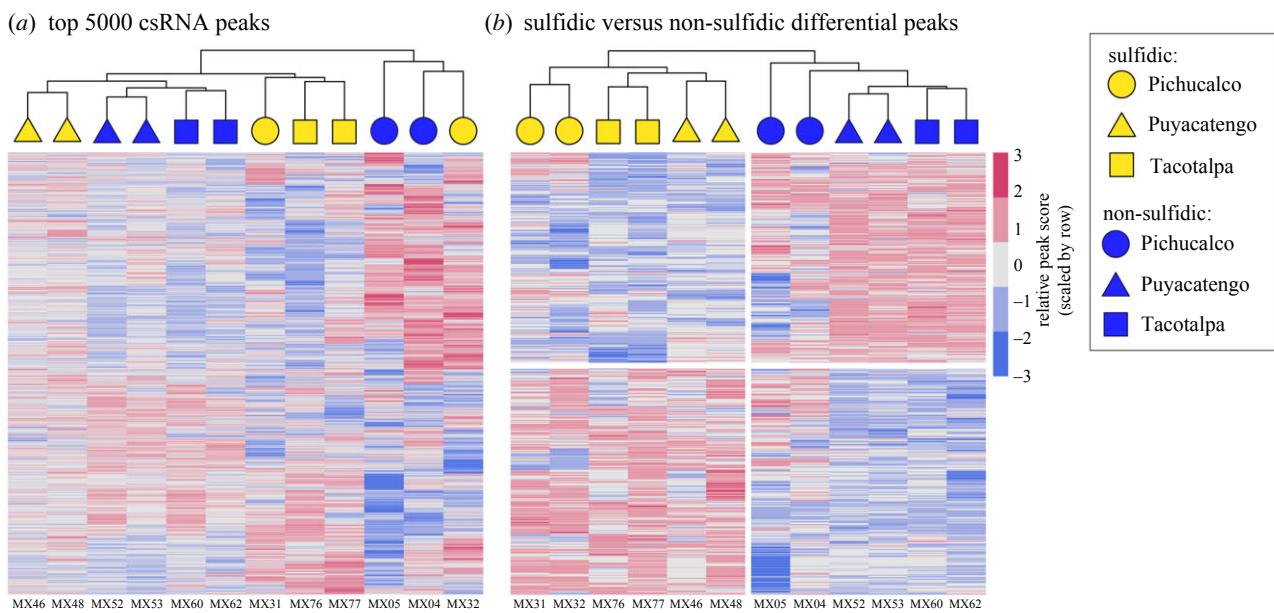
We investigated if patterns in differential transcript initiation between ecotypes matched patterns in nearby gene expression. Peaks with differential transcript initiation from csRNA-seq (FDR  $<0.05$ ) were intersected with differentially expressed genes from mRNA-seq (FDR  $<0.05$ ) based on their gene annotation to identify genomic regions that showed both changes in nascent transcription and stable gene expression between ecotypes. We then examined directionality to see if csRNA and mRNA datasets were both upregulated in sulfidic samples compared with non-sulfidic samples, both downregulated, or divergent. Those that were upregulated in both datasets or downregulated in both datasets are hereafter referred to as the 'parallel upregulated subset' and 'parallel downregulated subset', respectively.

#### (g) Transcription factor binding site enrichment in differentially initiated promoter regions

We identified TF binding sites in differentially initiated csRNA-seq peaks in the promoter-TSS regions of differentially expressed genes, focusing specifically on the upregulated and downregulated subsets defined above. We used *homerTools findMotifsGenome.pl* to extract FASTA sequences spanning a 200 bp region from 150 bp upstream through 50 bp downstream of the peak centre, which should approximate the site of transcription initiation. We chose this narrow window in order to conservatively identify only the primary TF binding sites involved in initiating gene expression. We used CiiIDER [53] to identify significantly enriched TF binding motifs in these regions using the JASPAR CORE 2022 non-redundant vertebrate transcription factors database [54]. We used a maximum deficit of 0.15 to allow for non-exact motif matching, and used a stringent *p*-value cutoff of 0.01 for classifying significantly enriched binding sites. A set of promoter-TSS regions that were not differentially initiated (FDR  $>0.05$  and  $\log_2$  fold-change  $<0.5$ ) and were not associated with a differentially expressed gene between ecotypes was used as the background for these enrichment analyses ( $n = 7443$ ).

#### (h) Identification and analysis of putative enhancer regions

csRNA-sequencing can capture enhancer RNAs, small unstable transcripts that result from active enhancer regions [55,56]. Enhancer RNA abundance often correlates with the expression of an enhancer's target gene(s) [57,58]; we therefore correlated differential initiation of putative enhancer csRNA-seq peaks with the expression of nearby genes to identify putative enhancers exhibiting differential activity between ecotypes and to infer genes that may be targeted by these enhancers. csRNA-seq peaks putatively corresponding to enhancer regions were defined as peaks annotated as distal from nearby genes (more than 500 bp from the promoter region of the nearest gene)



**Figure 2.** Hierarchical clustering analyses of (a) the top 5000 highest-abundance csRNA-seq peaks and (b) significantly differentially initiated peaks between sulfidic and non-sulfidic ecotypes. Each row represents a peak, each column a sample. The top dendrogram groups samples by similar patterns of transcript initiation. Heatmap colours represent relative peak score values, scaled by row to accentuate variation across samples. Shapes along the top dendrogram represent different drainages, with colour representing ecotype. Sample ID is listed below each column.

and comprising unstable transcripts (i.e. transcripts present in csRNA but not mRNA libraries). This list of peaks was then intersected with the results of differential initiation analysis to identify putative enhancers with significantly increased or decreased activity in the sulfidic ecotype.

We then identified all significantly differentially expressed genes with a TSS located less than 500 kbp from the centre of an ecotype-responsive putative enhancer peak by first expanding all ecotype-responsive putative enhancer peaks by 500 kbp in both directions using the *slop* tool within bedtools (v. 2.30.0) and then intersecting these expanded regions with a bedfile of gene TSS positions using bedtools *intersect*. The resulting pairs of putative enhancers and potential target genes were filtered to retain those with the same direction of differential initiation/expression in the sulfidic ecotype (i.e. upregulated csRNA-seq peak and upregulated gene expression, and *vice versa*), and the  $\log_2$  fold-change values of csRNA-seq and mRNA-seq were correlated using linear regressions in R. To focus downstream analyses only on putative target genes with high correlations to putative enhancer regions, pairs of putative enhancers and target genes were filtered to retain only those falling within one standardized residual of the regression line.

For the resulting putative enhancer peaks that were highly correlated with the expression of their inferred target gene(s), fasta sequences were extracted using the *getfasta* tool from bedtools. A background set of 1000 peaks were randomly selected from all putative enhancer peaks that were not differentially initiated between ecotypes (i.e. distal csRNA-seq peaks comprising unstable transcripts and with an FDR  $> 0.05$  in differential initiation analyses). CiiIDER was again used to identify significantly enriched TF binding motifs in these regions using the JASPAR CORE 2022 non-redundant vertebrate transcription factors database. We used a maximum deficit of 0.15 to allow for non-exact motif matching, and used a stringent *p*-value cutoff of 0.01 for defining significantly enriched binding sites.

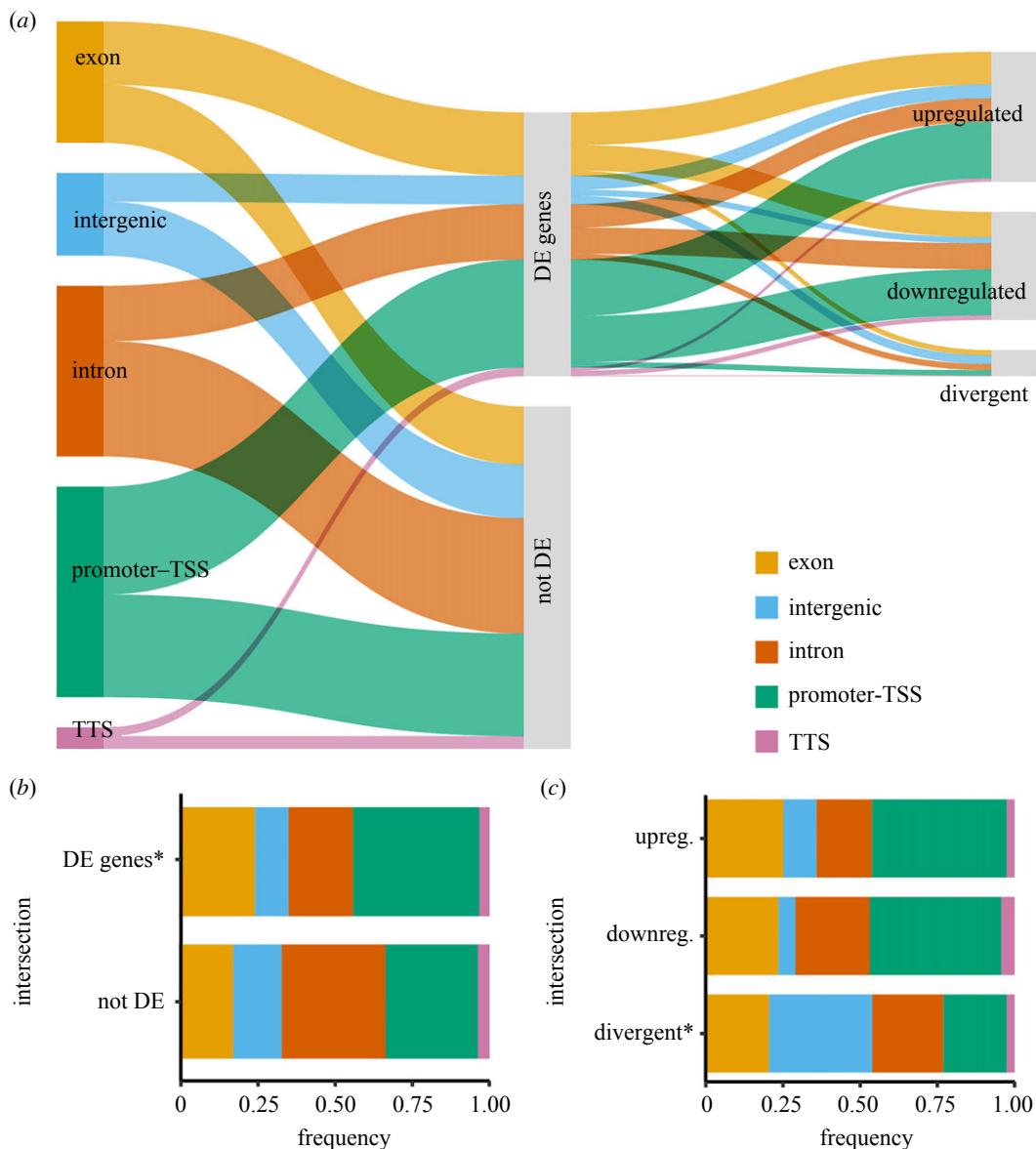
### 3. Results

#### (a) Transcriptional differences across ecotypes

Capturing actively initiating RNAs with csRNA-seq in sulfidic and non-sulfidic populations (see electronic supplementary material, tables S2 and S3) identified an average of  $27\,021 \pm 2096$  regions of active transcription initiation (peaks) per sample (see electronic supplementary material, table S4 for per-population metrics). Merging all directly overlapped peaks resulted in a total of 66 537 non-redundant peaks across all samples.

Hierarchical clustering of the top 5000 highest-scoring csRNA peaks grouped all samples by population except for one individual from the Pichucalco drainage (figure 2a). In a PCA generated from the top 5000 highest-scoring csRNA peaks, *P. sulphuraria* and *P. mexicana* populations separated along the second axis of variation, while the first axis of variation was driven largely by samples from the non-sulfidic Pichucalco population (electronic supplementary material, figure S1). Clustering of significantly differentially initiated peaks grouped samples within both ecotypes by population, indicating that population-specific variation in transcription initiation remains present even for peaks with broad differences in initiation between ecotypes (figure 2b).

We identified 1061 peaks with significantly different csRNA-seq peak density (i.e. differentially initiated peaks) between the two ecotypes (FDR  $< 0.05$ ; electronic supplementary material, table S5). Of those, 550 peaks were upregulated and 511 peaks were downregulated in sulfidic samples compared with non-sulfidic samples. Peaks of transcription initiation located in or near genes involved in H<sub>2</sub>S detoxification and response were roughly twice as likely to be differentially active between ecotypes (odds ratio = 1.997, 95% CI = 0.974–3.688) as transcripts that were not related to H<sub>2</sub>S (two-sided Fisher's exact test, *p* < 0.05; electronic supplementary material, table S6). Three peaks upregulated in sulfidic populations mapped to upstream regions of the genes encoding sulfide:quinone oxidoreductase (*sqrdl*) and persulfide dioxygenase (*ethe1*), both of which are directly involved



**Figure 3.** (a) The intersection of differentially initiated (DI) peaks ( $FDR < 0.05$ ) from csRNA-seq with differentially expressed (DE) genes from mRNA-seq ( $FDR < 0.05$ ). Of the DI peaks that also showed DE, most showed parallel shifts in the level of transcription (i.e. upregulated in both datasets, downregulated in both datasets) instead of divergent shifts. (b) Relative frequency of genomic locations of DI peaks associated with a DE gene (top bar) and DI peaks not associated with a DE gene (bottom bar). (c) Genomic locations of DI peaks with DE genes that were both upregulated (top bar), downregulated (middle bar), or divergent (bottom bar). TSS = transcription start site, TTS = transcription termination site. Asterisks indicate significantly different proportions in that group of peaks compared with all peaks combined ( $G$ -test goodness of fit tests,  $p < 0.05$ ).

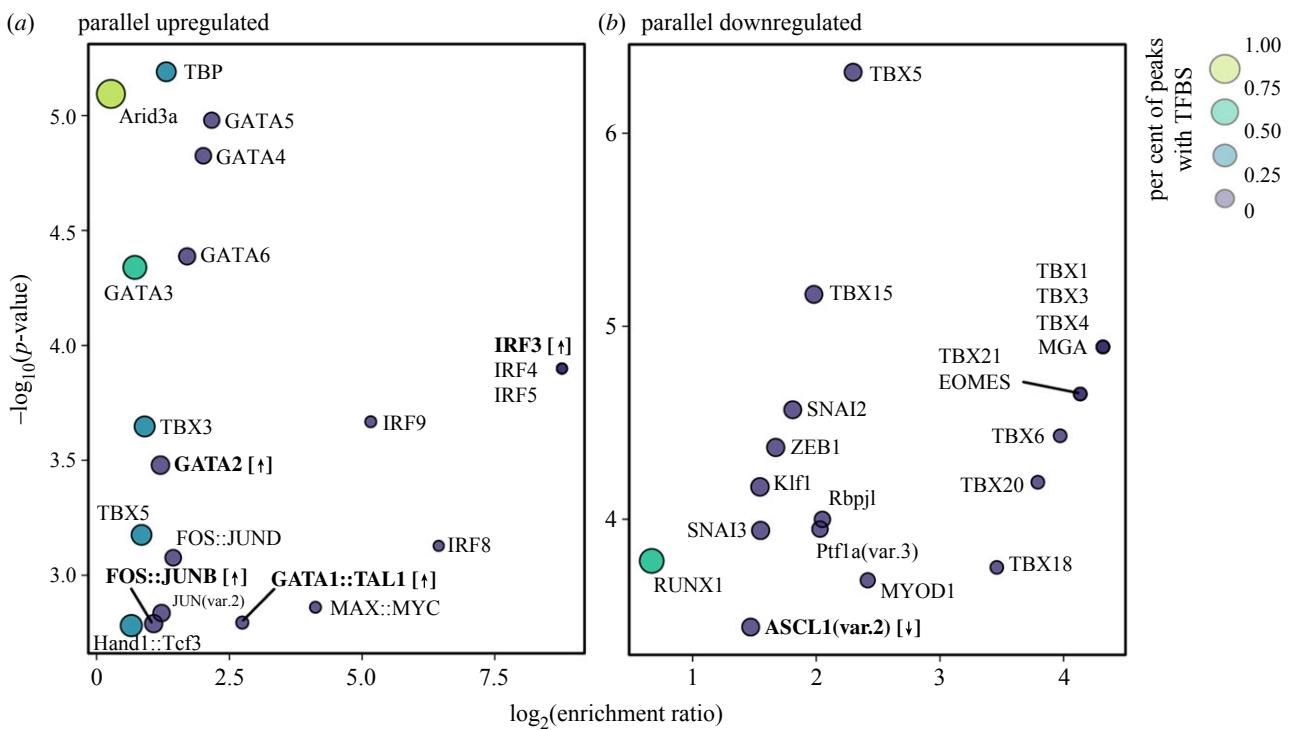
in  $H_2S$  detoxification via the SQR pathway in the mitochondria (electronic supplementary material, table S6). Differential transcription initiation also occurred near the genes encoding a glutathione S-transferase (*gstz1*), chloride intracellular channel (*clic2*), and solute carrier (*slc16a6b*), among others (electronic supplementary material, table S6).

### (b) Differential gene expression across ecotypes

We compared gill mRNA-seq data from Kelley *et al.* [27] with ecotype-level differences in nascent transcription to examine the relationship between coding and non-coding transcription shifts in the sulfidic ecotype. We sequenced a total of 531 million mRNA reads, with an average of  $15 \pm 0.9$  million reads per sample (see electronic supplementary material, tables S7 and S8 for read counts and mapping statistics per population). Gene expression clustered samples by ecotype along the first and second axes of variation in the PCA, with *P. sulphuraria* from the Pichucalco drainage clustering separately from the sulfidic *P. mexicana* from the Puyacatengo and Tacotalpa drainages (electronic supplementary material, figure S2). Analysis of differential expression between sulfidic and non-sulfidic ecotypes identified 2852 differentially expressed genes, 1395 of which were upregulated in the sulfidic ecotype ( $FDR < 0.05$ ; electronic supplementary material, table S9).

### (c) Ecotype-level changes in nascent transcription paralleled those in gene expression

Most nascent transcription occurred near coding regions, particularly in the promoter-TSS regions (32.7%) and introns (28.1%; figure 3a). Of the 910 differentially initiated peaks that were annotated with a nearby gene, 44% were associated with genes found to be DE in analyses of mRNA-seq (figure 3a, 'DE genes').



**Figure 4.** The top 20 most significantly enriched transcription factor binding sites (TFBSs) in (a) significantly upregulated csRNA peaks ( $\text{csRNA-seq } \log_2 \text{fold-change} > 0$ ,  $\text{FDR} < 0.05$ ) in the promoter-TSS of significantly upregulated genes ( $\text{mRNA-seq } \log_2 \text{fold-change} > 0$ ,  $\text{FDR} < 0.05$ ) and (b) significantly downregulated csRNA peaks ( $\text{csRNA-seq } \log_2 \text{fold-change} < 0$ ,  $\text{FDR} < 0.05$ ) in the promoter-TSS of significantly downregulated genes ( $\text{mRNA-seq } \log_2 \text{fold-change} < 0$ ,  $\text{FDR} < 0.05$ ). The  $\log_2$  enrichment ratio is represented on the  $x$ -axis, with points to the right in each panel having a higher degree of overrepresentation in the target peak set compared with the background. The  $-\log_{10}(p\text{-value})$  of the enrichment test is represented on the  $y$ -axis, with points towards the top of each panel having lower  $p$ -values. Point size and colour are scaled by the proportion of target peaks possessing a given TFBS. Bold labels indicate TFBSs for which the gene encoding that TF is significantly upregulated (up arrows) or downregulated (down arrow) in the sulfidic ecotype. The ‘::’ separates co-binding TFs. For a list of all enriched TFBSs, see electronic supplementary material, table S9.

The genomic annotations of csRNA peaks differed depending on whether the nearest gene was differentially expressed (figure 3b); csRNA peaks near differentially expressed genes were predominately located within promoter-TSS regions (figure 3b, DE genes), whereas peaks associated with non-differentially expressed genes were located primarily in intronic regions (figure 3b, not DE).

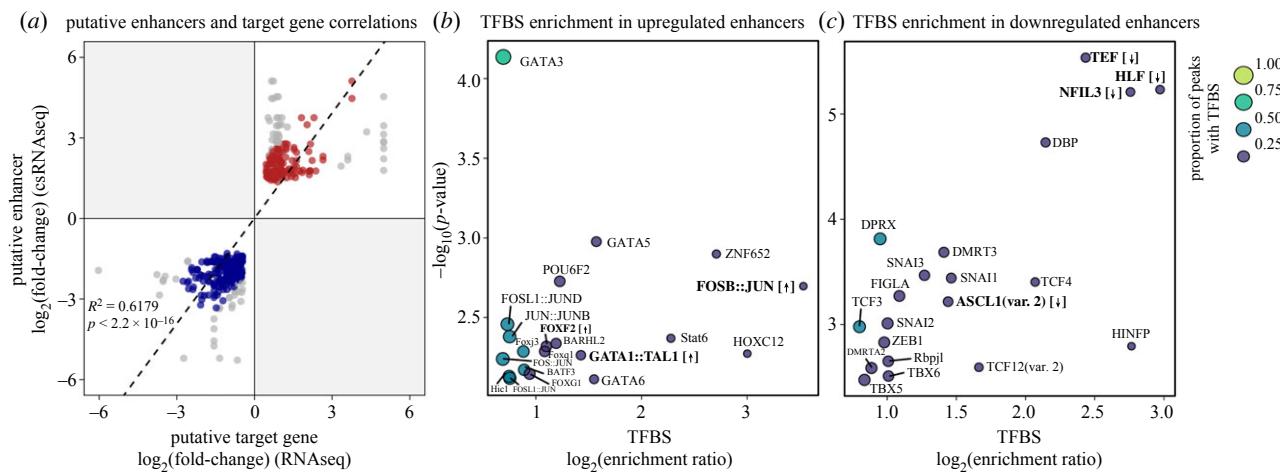
Of the 396 differentially initiated csRNA peaks associated with differentially expressed genes, 90% showed parallel shifts in the level of transcription (i.e. either both upregulated or both downregulated; figure 3a). We refer to these groups of peaks and associated genes as the ‘parallel upregulated subset’ and ‘parallel downregulated subset’ for simplicity. Peaks in these two parallel subsets were primarily located in promoter-TSS regions (figure 3c, ‘upregulated’ and ‘downregulated’), whereas peaks that showed an opposite change in magnitude to the nearby gene were predominately located in intergenic regions (figure 3c, ‘divergent’).

### (c) Transcription factor binding site enrichment reveals overrepresented motifs in sulfidic populations

To investigate TFs with potential roles in driving ecotype-specific gene expression, we performed TF binding site enrichment analysis in all promoter-TSS peaks belonging to the parallel upregulated and parallel downregulated subsets defined above (85 peaks in the parallel upregulated subset, 69 in the parallel downregulated subset). We identified 65 TF binding sites that were significantly enriched in the parallel upregulated promoter-TSS peaks and 56 enriched in the parallel downregulated promoter-TSS peak set, with 10 TF binding sites found to be enriched in upregulated and downregulated promoter-TSS peaks ( $p < 0.01$ , figure 4; electronic supplementary material, table S10). Of the 65 TF binding sites enriched in the parallel upregulated peak set, nine were further supported by significant upregulation of a gene encoding the TF (GATA2, SP8, IRF3, STAT::STAT2, GATA1::TAL1, KLF11, FOSB::JUNB, SHOX and Mafb; figure 4; electronic supplementary material, table S11). Only three TFs with enriched binding sites in the parallel downregulated peak set were further supported by the downregulation of the TF-encoding gene (KLF5, ASCL1 and TCF7; figure 4; electronic supplementary material, table S11).

### (d) Putative enhancer regions associated with sulfide-responsive genes

To identify putative enhancer regions, we first filtered all csRNA-seq peaks based on annotations consistent with expectations of enhancer RNAs (distal from the promoters of annotated genes and comprising unstable transcripts). This resulted in 33 799 regions, 448 of which were found to be significantly differentially initiated between sulfidic and non-sulfidic ecotypes ( $\text{FDR} < 0.05$ ); hereafter, these 448 peaks are referred to as differentially active putative enhancer regions.



**Figure 5.** (a) Correlations between the  $\log_2$  fold-change values of differentially active putative enhancer regions (y-axis) and potential differentially expressed target genes (x-axis). Red and blue points indicate  $\text{H}_2\text{S}$ -upregulated and  $\text{H}_2\text{S}$ -downregulated enhancer–gene pairs that fall within one standardized residual of the regression line, respectively. Grey points fall outside of one standardized residual of the regression line. (b, c) The top 20 most significantly enriched transcription factor binding sites (TFBSs) in (b) significantly upregulated putative enhancers ( $\log_2$  fold-change  $> 0$ ,  $\text{FDR} < 0.05$ ) with high correlation to a nearby upregulated gene (red points in (a);  $\log_2$  fold-change  $> 0$ ,  $\text{FDR} < 0.05$ ), and (c) significantly downregulated putative enhancers ( $\log_2$  fold-change  $< 0$ ,  $\text{FDR} < 0.05$ ) with high correlation to a nearby downregulated gene (blue points in (a);  $\log_2$  fold-change  $< 0$ ,  $\text{FDR} < 0.05$ ). The  $\log_2$  enrichment ratio is represented on the x-axis, with points to the right in each panel having a higher degree of overrepresentation in the target peak set compared with the background. The  $-\log_{10}(p\text{-value})$  of the enrichment test is represented on the y-axis, with points towards the top of each panel having lower  $p$ -values. Point size and colour are scaled by the proportion of target peaks possessing a given TFBS. Bold labels indicate TFBSs for which the gene encoding that TF is significantly upregulated (up arrows) or downregulated (down arrows) in the sulfidic ecotype. The ‘::’ separates co-binding TFs. For a list of all enriched TFBSs, see electronic supplementary material, table S12.

Of the 2852 differentially expressed genes identified between ecotypes, 535 (18.8%) were located within 500 kbp of one or more differentially active putative enhancer regions. The majority of these genes (56%) were differentially expressed in the same direction as the nearby enhancer in the sulfidic ecotype. For this set of putative enhancer–gene pairs with the same direction of upregulation/activity in sulfidic fish, we correlated csRNA-seq and mRNA-seq  $\log_2$  fold-change values and retained enhancer–gene pairs falling within one standardized residual of the linear regression line. This resulted in 213 enhancer–gene pairs, with 86 upregulated in the sulfidic ecotype (red points, figure 5a) and 127 downregulated (blue points, figure 5a; electronic supplementary material, table S12). Neither upregulated nor downregulated enhancer–gene pairs were enriched for candidate  $\text{H}_2\text{S}$  detoxification genes (two-sided Fisher’s exact test,  $p = 0.2307$ ), although two upregulated candidate genes, *slc26a5* and *slc26a11*, were inferred to be targeted by upregulated putative enhancer regions.

We identified 32 and 31 TF binding sites that were significantly enriched in the parallel  $\text{H}_2\text{S}$ -upregulated and  $\text{H}_2\text{S}$ -downregulated enhancer–gene pairs, respectively (figure 5b,c; electronic supplementary material, table S13). Six genes coding for TFs enriched in the upregulated enhancers were themselves upregulated in the sulfidic ecotype based on gene expression data, while genes encoding three of the TFs enriched in the downregulated enhancer set exhibited downregulated gene expression (figure 5b,c; electronic supplementary material, table S14). In upregulated putative enhancer regions, the *FOSB::JUN* TF binding site was the most enriched relative to the background, and increased activity of this TF complex is further supported by the upregulation of the gene encoding *fosh* in the sulfidic ecotype (figure 5b). In downregulated putative enhancer regions, *DBP*, *NFIL3*, *TEF* and *HLF* TF binding sites were the most significant and highly enriched TFs compared with the background, with the latter three being further supported by downregulation of the TF-encoding gene (figure 5c).

## 4. Discussion

Examining regulatory mechanisms underlying adaptation to toxic levels of  $\text{H}_2\text{S}$  in extremophile fishes allows us to gain a new functional understanding of some of the most highly conserved metabolic and detoxification pathways in metazoans. To achieve this, we quantified active transcription occurring outside of coding regions in sulfidic and non-sulfidic populations. We found convergent shifts in nascent transcription in sulfidic populations of *P. mexicana* (including *P. sulphuraria*) compared with non-sulfidic populations. Our results show that transcriptional differences between these ecotypes extend beyond that of mature, stable mRNA. As is true in other vertebrate systems, nascent transcripts likely reflect regulatory activity that affects gene expression [12,59]. Several differential transcription initiation sites mapped close to or within genes coding for enzymes involved in  $\text{H}_2\text{S}$  detoxification (persulfide dioxygenase, sulfide:quinone oxidoreductase) and metabolism (glutathione S-transferase zeta 1). There was also differential transcription initiation in the promoter region of a solute carrier gene, *slc16a6*, that is inhibited by sulfur-based compounds [60]. Within ecotypes, we found that nascent transcription patterns cluster samples by drainage, suggesting population-specific variation in transcription initiation and regulatory strategies that were not captured by ecotype comparisons alone. Drainage-specific physiological differences have been documented in the *P. mexicana* species complex previously [4,22],

and examining regulatory differences that may underlie this physiology warrants future study with larger within-drainage sample sizes.

Most differential transcription initiation shifted in parallel with nearby gene expression, such that peaks with higher csRNA-seq signal in sulfidic populations were often near to genes with upregulated expression, and *vice versa*. We identified sets of enriched TF binding motifs in these differentially initiated csRNA-seq peaks occurring in the promoter-TSS of genes with parallel shifts in expression. Binding sites for several members of the Interferon regulatory factors (IRF) family of TFs were found to have the greatest magnitude of enrichment in the parallel upregulated promoter-TSS peaks. This family of TFs is typically involved in immune and pro-inflammatory responses to pathogens [61]. H<sub>2</sub>S has been shown previously to reduce activation of IRF3 and associated inflammatory responses to viral infection [62]. However, IRF3 was among the most enriched binding sites in parallel upregulated promoter-TSS peaks, and the gene encoding this TF was also found to be upregulated in the sulfidic ecotype. Members of the GATA family of TFs were overrepresented in upregulated promoter-TSS csRNA-seq peaks, including GATA1–3, involved in the haematopoietic system [63–65], and GATA4–6, involved in heart development [66]. While this may be due in part to nucleated blood present in the gill tissues at the time of sampling, the GATA TFs may have important roles in sulfide adaptation given that H<sub>2</sub>S binds haemoglobin, decreasing its effectiveness [67], and that erythropoiesis is regulated in response to hypoxia in fishes [68]. Two genes annotated in the *P. mexicana* genome as GATA2 and a gene encoding TAL1, a TF that co-binds along with GATA1, are upregulated in the sulfidic ecotype, providing further evidence for the increased activity of these TFs in response to H<sub>2</sub>S.

Among the TF binding sites found to be enriched in H<sub>2</sub>S-downregulated promoter-TSS peaks, RUNX1 was present in a high proportion of peaks despite a relatively low enrichment ratio compared with the background sequences (figure 4b). RUNX1 interacts with hypoxia-inducible factor 1-alpha (HIF-1 $\alpha$ ), an important high-level regulator of cellular responses to hypoxia [69,70]. Additionally, these analyses identified enrichment of binding sites for ASCL1 (figure 4b), and the gene encoding ASCL1 was found to be H<sub>2</sub>S-downregulated in gene expression analyses. In amniotes, ASCL1 plays a critical role in the formation of pulmonary neuroendocrine cells that sense and initiate responses to hypoxia [71,72]. In other contexts (i.e. mammalian neuroblastomas), regulatory activity of ASCL1 has been shown to decrease in response to hypoxia [73]. While the exact roles and relevance of RUNX1 and ASCL1 activity in fish gills is not fully understood, their roles in sensing and responding to hypoxia in other systems suggests that they may play an important role in adaptation to sulfide springs with combined stress associated with H<sub>2</sub>S toxicity and hypoxia, warranting further investigation in future studies.

Our findings suggest that differential enhancer activity plays an important role in H<sub>2</sub>S adaptation by contributing to ecotype-specific gene regulation. Analysis of enriched TF binding sites in these putative enhancers revealed a suite of regulatory molecules that may stimulate shifts in gene expression in the sulfidic ecotype, many of which were not inferred to target the promoters of differentially expressed genes. Of note, the transcription factor FOXF2 was enriched in binding sites of H<sub>2</sub>S-upregulated putative enhancers targeting H<sub>2</sub>S-upregulated genes, and the *foxf2* gene was also upregulated in the sulfidic ecotype (figure 5a). In mammals, FOXF2 is highly expressed in the lung and is a major regulator of lung-specific gene expression [74,75]. Regulatory activity of FOXF2 has been shown to increase in response to cellular stress and hypoxic conditions [76]. Our inferences of increased FOXF2 regulatory activity in sulfidic populations of *P. mexicana* may suggest a similar role of FOXF2 in the response to hypoxia in fish gill, although subsequent study is needed to understand the precise role and relevance of this TF in H<sub>2</sub>S adaptation. In H<sub>2</sub>S-downregulated putative enhancers targeting H<sub>2</sub>S-downregulated genes, the transcription factors TEF, HLF, DBP and NFIL3 were among the most significantly enriched for binding sites (figure 5b). These four TFs belong to the PAR-domain basic leucine zipper TF family and are regulators of cellular responses to xenobiotics and metabolic detoxification [77,78]; accordingly, mice in which TEF, HLF and DBP are knocked out exhibit reduced expression and activity of downstream detoxification genes regulated by these TFs [77]. TEF, HLF and DBP have also been implicated in regulating pro-apoptotic responses to oxidative stress [79]. Evidence of decreased activity of these TFs in the sulfidic ecotype may indicate that aberrant activation of some detoxification pathways may be detrimental in H<sub>2</sub>S-adapted populations, similar to findings from studies of pollution-tolerant killifish populations [80]. However, subsequent experimental studies are needed to fully understand the relevance of these TFs in the sulfidic ecotype.

While the differential binding of TFs in promoter and enhancer regions likely contributes to many of the transcriptional differences separating the populations, relatively few genes encoding TFs with enriched binding sites were differentially expressed between ecotypes. It is probable that other mechanisms, such as differential splicing [81], phosphorylation (reviewed in [82]), sub-cellular localization [83], and dimerization (reviewed in [84]), have additional impacts on changes in TF function that were not examined here. Thus, the lack of differential gene expression for many TFs found to be enriched in promoter and/or putative enhancer peaks does not rule out their involvement or differential activity between ecotypes.

An important consideration of this study is that the *P. mexicana* reference genome assembly is highly fragmented (scaffold n50 = 275.3 kb, contig n50 = 39.8 kb), which limits our inference of genes targeted by putative enhancers to genes located on the same scaffold. This could result in the true target gene(s) of some putative enhancers going unidentified if not currently assembled on the same scaffold, and we therefore caution that our inferences of the genes targeted by differentially active putative enhancers remain preliminary at this stage. Should assembly contiguity improve in the future, the reanalysis of these data may reveal additional putative enhancer–gene pairs associated with H<sub>2</sub>S adaptation in this system. Further, the incorporation of other functional genomic evidence for enhancer activity using ATAC-seq and/or ChIP-seq would provide valuable clarity on both the differential activity and finer-scale TF binding of putative enhancers identified in this study.

In summary, our analyses revealed that transcriptional differences between the sulfidic and non-sulfidic ecotypes of the *P. mexicana* species complex differ beyond that of just gene expression and also include non-coding RNAs. Many genes differentially expressed between ecotypes also exhibited parallel shifts in nascent transcription in promoter, exonic and intronic regions. Additionally, analyses of putative enhancer regions suggest that differential enhancer activity contributes to ecotype-specific shifts in gene expression and provides a new and important set of genomic loci associated with H<sub>2</sub>S adaptation in this system.

Ecotype-specific differences in TF activity inferred from TF binding site enrichment were apparent in both promoter and putative enhancer regions, implicating a large number of high-level regulatory molecules potentially driving responses to hypoxia and cellular stress in the sulfidic ecotype, and a subset of these TFs were further supported by the differential expression of the TF-encoding gene itself. Ultimately, our analyses showed clear ecotype-level differences in non-coding transcription and TF binding, both of which help explain changes in gene expression observed in the sulfidic extremophile populations of the *P. mexicana* species complex compared with non-sulfidic populations.

**Ethics.** All experiments were approved by the Institutional Animal Care and Use Committee of Kansas State University (ACUP no. 3473). Permits were provided by the Mexican Federal Agencies SEMARNAT and CONAPESCA (DGOPA.09004.04111.3088, SGPA/DGVS/04315/11, PRMN/DGOPA-003/2014, PRMN/DGOPA-009/2015).

**Data accessibility.** Data are available from NCBI's GenBank. Accessions: PRJNA290391 (mRNA), PRJNA743555 (csRNA). Code for all analyses is available on Dryad: <http://dx.doi.org/10.5061/dryad.83bk3jb0q> [85].

Supplementary material is available online [86].

**Declaration of AI use.** We have not used AI-assisted technologies in creating this article.

**Authors' contributions.** B.W.P.: data curation, formal analysis, methodology, writing—original draft, writing—review and editing; K.L.M.: data curation, formal analysis, methodology, writing—original draft; L.A.-R.: investigation, resources, writing—review and editing; S.H.D.: methodology, resources, writing—review and editing; M.T.: conceptualization, funding acquisition, project administration, writing—review and editing; J.L.K.: conceptualization, funding acquisition, investigation, methodology, project administration, supervision, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed herein.

**Conflict of interest declaration.** We declare we have no competing interests.

**Funding.** This work was funded by grants from the Vern Parish Fund from the American Livebearer Association to K.L.M., the National Science Foundation (IOS-1931657, IOS-2311366) and the Army Research Office (W911NF-15-1-0175, W911NF-16-1-0225) to M.T. and J.L.K., the L'Oréal Fellowship for Women in Science to J.L.K., and the Ralph E. Powe Junior Faculty Enhancement Award from Oak Ridge Associated Universities to M.T. This work was also supported by a National Institutes of Health award (R00GM135515) to S.H.D.

**Acknowledgements.** We would like to thank El Centro de Investigación e Innovación para la Enseñanza y Aprendizaje and Universidad Juárez Autónoma de Tabasco, México for their support and facilitating the collection of samples. This research used resources from the Center for Institutional Research Computing at Washington State University. T. Algayer, E. E. Armstrong, J. W. Busch, O. E. Cornejo, R. De-Kayne, W. W. Dowd, S. Hotaling, J. D. Jensen, K. Ryan, and L. Wilson provided advice on analyses and/or comments on the manuscript.

## References

1. Perry BW *et al.* 2019 Multi-species comparisons of snakes identify coordinated signalling networks underlying post-feeding intestinal regeneration. *Proc. R. Soc. B* **286**, 20190910. ([doi:10.1098/rspb.2019.0910](https://doi.org/10.1098/rspb.2019.0910))
2. Zhao X, Bergland AO, Behrman EL, Gregory BD, Petrov DA, Schmidt PS. 2016 Global transcriptional profiling of diapause and climatic adaptation in *Drosophila melanogaster*. *Mol. Biol. Evol.* **33**, 707–720. ([doi:10.1093/molbev/msv263](https://doi.org/10.1093/molbev/msv263))
3. Veilleux HD, Ryu T, Donelson JM, Ravasi T, Munday PL. 2018 Molecular response to extreme summer temperatures differs between two genetically differentiated populations of a coral reef fish. *Front. Mar. Sci.* **5**, 349. ([doi:10.3389/fmars.2018.00349](https://doi.org/10.3389/fmars.2018.00349))
4. Pfenninger M *et al.* 2014 Parallel evolution of *cox* genes in H<sub>2</sub>S-tolerant fish as key adaptation to a toxic environment. *Nat. Commun.* **5**, 3873. ([doi:10.1038/ncomms4873](https://doi.org/10.1038/ncomms4873))
5. Natarajan C, Hoffmann FG, Weber RE, Fago A, Witt CC, Storz JF. 2016 Predictable convergence in hemoglobin function has unpredictable molecular underpinnings. *Science* **354**, 336–339. ([doi:10.1126/science.aaf9070](https://doi.org/10.1126/science.aaf9070))
6. Jacob F. 1977 Evolution and tinkering. *Science* **196**, 1161–1166. ([doi:10.1126/science.860134](https://doi.org/10.1126/science.860134))
7. King M-C, Wilson AC. 1975 Evolution at two levels in humans and chimpanzees: their macromolecules are so alike that regulatory mutations may account for their biological differences. *Science* **188**, 107–116. ([doi:10.1126/science.1090005](https://doi.org/10.1126/science.1090005))
8. Britten RJ, Davidson EH. 1969 Gene regulation for higher cells: a theory: new facts regarding the organization of the genome provide clues to the nature of gene regulation. *Science* **165**, 349–357. ([doi:10.1126/science.165.3891.349](https://doi.org/10.1126/science.165.3891.349))
9. Scott GR, Elogio TS, Lui MA, Storz JF, Chevron ZA. 2015 Adaptive modifications of muscle phenotype in high-altitude deer mice are associated with evolved changes in gene regulation. *Mol. Biol. Evol.* **32**, 1962–1976. ([doi:10.1093/molbev/msv076](https://doi.org/10.1093/molbev/msv076))
10. Linnen CR, Kingsley EP, Jensen JD, Hoekstra HE. 2009 On the origin and spread of an adaptive allele in deer mice. *Science* **325**, 1095–1098. ([doi:10.1126/science.1175826](https://doi.org/10.1126/science.1175826))
11. Whitehead A, Crawford DL. 2006 Neutral and adaptive variation in gene expression. *Proc. Natl. Acad. Sci. USA* **103**, 5425–5430. ([doi:10.1073/pnas.0507648103](https://doi.org/10.1073/pnas.0507648103))
12. Chang T-C, Pertea M, Lee S, Salzberg SL, Mendell JT. 2015 Genome-wide annotation of microRNA primary transcript structures reveals novel regulatory mechanisms. *Genome Res.* **25**, 1401–1409. ([doi:10.1101/gr.193607.115](https://doi.org/10.1101/gr.193607.115))
13. Kapranov P *et al.* 2007 RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* **316**, 1484–1488. ([doi:10.1126/science.1138341](https://doi.org/10.1126/science.1138341))
14. Cheng J-H, Pan DZ-C, Tsai ZT-Y, Tsai H-K. 2015 Genome-wide analysis of enhancer RNA in gene regulation across 12 mouse tissues. *Scient. Rep.* **5**, 12648. ([doi:10.1038/srep12648](https://doi.org/10.1038/srep12648))
15. Statello L, Guo C-J, Chen L-L, Huarte M. 2021 Gene regulation by long non-coding RNAs and its biological functions. *Nat. Rev. Mol. Cell Biol.* **22**, 96–118. ([doi:10.1038/s41580-020-00315-9](https://doi.org/10.1038/s41580-020-00315-9))
16. Popp AP, Hettich J, Gebhardt JCM. 2021 Altering transcription factor binding reveals comprehensive transcriptional kinetics of a basic gene. *Nucleic Acids Res.* **49**, 6249–6266. ([doi:10.1093/nar/gkab443](https://doi.org/10.1093/nar/gkab443))
17. Boos F *et al.* 2019 Mitochondrial protein-induced stress triggers a global adaptive transcriptional programme. *Nat. Cell Biol.* **21**, 442–451. ([doi:10.1038/s41556-019-0294-5](https://doi.org/10.1038/s41556-019-0294-5))
18. Palacios M, Voelker G, Rodriguez LA, Mateos M, Tobler M. 2016 Phylogenetic analyses of the subgenus *Mollienesia* (Poecilia, Poeciliidae, Teleostei) reveal taxonomic inconsistencies, cryptic biodiversity, and spatio-temporal aspects of diversification in Middle America. *Mol. Phylogenetic. Evol.* **103**, 230–244. ([doi:10.1016/j.ympev.2016.07.025](https://doi.org/10.1016/j.ympev.2016.07.025))
19. Tobler M, Kelley JL, Plath M, Riesch R. 2018 Extreme environments and the origins of biodiversity: adaptation and speciation in sulphide spring fishes. *Mol. Ecol.* **27**, 843–859. ([doi:10.1111/mec.14497](https://doi.org/10.1111/mec.14497))
20. Plath M *et al.* 2013 Genetic differentiation and selection against migrants in evolutionarily replicated extreme environments. *Evolution* **67**, 2647–2661. ([doi:10.1111/evo.12133](https://doi.org/10.1111/evo.12133))
21. Greenway R, Brown AP, Camarillo H, Delich C, McGowan KL, Nelson J, Arias-Rodriguez L, Kelley JL, Tobler M. 2021 Convergent adaptation and ecological speciation result from unique genomic mechanisms in sympatric extremophile fishes. *bioRxiv*, 2021.06.28.450104. ([doi:10.1101/2021.06.28.450104](https://doi.org/10.1101/2021.06.28.450104))

22. Greenway R *et al.* 2020 Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. *Proc. Natl Acad. Sci. USA* **117**, 16 424–16 430. ([doi:10.1073/pnas.2004223117](https://doi.org/10.1073/pnas.2004223117))
23. Jiang J *et al.* 2016 Hydrogen sulfide—mechanisms of toxicity and development of an antidote. *Scient. Rep.* **6**, 20831. ([doi:10.1038/srep20831](https://doi.org/10.1038/srep20831))
24. Tobler M, Passow CN, Greenway R, Kelley JL, Shaw JH. 2016 The evolutionary ecology of animals inhabiting hydrogen sulfide-rich environments. *Annu. Rev. Ecol. Evol. Syst.* **47**, 239–262. ([doi:10.1146/annurev-ecolsys-121415-032418](https://doi.org/10.1146/annurev-ecolsys-121415-032418))
25. Vaquer-Sunyer R, Duarte CM. 2010 Sulfide exposure accelerates hypoxia-driven mortality. *Limnol. Oceanogr.* **55**, 1075–1082. ([doi:10.4319/lo.2010.55.3.1075](https://doi.org/10.4319/lo.2010.55.3.1075))
26. Lukas J, Auer F, Goldhammer T, Krause J, Romanczuk P, Klamser P, Arias-Rodriguez L, Bierbach D. 2021 Diurnal changes in hypoxia shape predator-prey interaction in a bird-fish system. *Front. Ecol. Evol.* **9**, 619193. ([doi:10.3389/fevo.2021.619193](https://doi.org/10.3389/fevo.2021.619193))
27. Kelley JL, Arias-Rodriguez L, Patascil Martin D, Yee MC, Bustamante CD, Tobler M. 2016 Mechanisms underlying adaptation to life in hydrogen sulfide-rich environments. *Mol. Biol. Evol.* **33**, 1419–1434. ([doi:10.1093/molbev/msw020](https://doi.org/10.1093/molbev/msw020))
28. Tobler M, Henpita C, Bassett B, Kelley JL, Shaw JH. 2014 H<sub>2</sub>S exposure elicits differential expression of candidate genes in fish adapted to sulfidic and non-sulfidic environments. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* **175**, 7–14. ([doi:10.1016/j.cbpa.2014.04.012](https://doi.org/10.1016/j.cbpa.2014.04.012))
29. Olson KR. 2018 H<sub>2</sub>S and polysulfide metabolism: conventional and unconventional pathways. *Biochem. Pharmacol.* **149**, 77–90. ([doi:10.1016/j.bcp.2017.12.010](https://doi.org/10.1016/j.bcp.2017.12.010))
30. Theissen U, Hoffmeister M, Grieshaber M, Martin W. 2003 Single eubacterial origin of eukaryotic sulfide: quinone oxidoreductase, a mitochondrial enzyme conserved from the early evolution of eukaryotes during anoxic and sulfidic times. *Mol. Biol. Evol.* **20**, 1564–1574. ([doi:10.1093/molbev/msg174](https://doi.org/10.1093/molbev/msg174))
31. Passow CN, Henpita C, Shaw JH, Quackenbush CR, Warren WC, Schartl M, Arias-Rodriguez L, Kelley JL, Tobler M. 2017 The roles of plasticity and evolutionary change in shaping gene expression variation in natural populations of extremophile fish. *Mol. Ecol.* **26**, 6384–6399. ([doi:10.1111/mec.14360](https://doi.org/10.1111/mec.14360))
32. Kelley JL, Desvignes T, McGowan KL, Perez M, Rodriguez LA, Brown AP, Culumber Z, Tobler M. 2021 microRNA expression variation as a potential molecular mechanism contributing to adaptation to hydrogen sulphide. *J. Evol. Biol.* **34**, 977–988. ([doi:10.1111/jeb.13727](https://doi.org/10.1111/jeb.13727))
33. Kelley JL, Tobler M, Beck D, Sadler-Riggleman I, Quackenbush CR, Arias Rodriguez L, Skinner MK. 2021 Epigenetic inheritance of DNA methylation changes in fish living in hydrogen sulfide-rich springs. *Proc. Natl Acad. Sci. USA* **118**, e2014929118. ([doi:10.1073/pnas.2014929118](https://doi.org/10.1073/pnas.2014929118))
34. Duttke SH, Chang MW, Heinz S, Benner C. 2019 Identification and dynamic quantification of regulatory elements using total RNA. *Genome Res.* **29**, 1836–1846. ([doi:10.1101/gr.253492.119](https://doi.org/10.1101/gr.253492.119))
35. Core LJ, Waterfall JJ, Lis JT. 2008 Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848. ([doi:10.1126/science.1162228](https://doi.org/10.1126/science.1162228))
36. Mahat DB *et al.* 2016 Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476. ([doi:10.1038/nprot.2016.086](https://doi.org/10.1038/nprot.2016.086))
37. Yan B, Tzertzinis G, Schildkraut I, Ettwiller L. 2022 Comprehensive determination of transcription start sites derived from all RNA polymerases using ReCappable-seq. *Genome Res.* **32**, 162–174. ([doi:10.1101/gr.275784.121](https://doi.org/10.1101/gr.275784.121))
38. Xu Z *et al.* 2009 Bidirectional promoters generate pervasive transcription in yeast. *Nature* **457**, 1033–1037. ([doi:10.1038/nature07728](https://doi.org/10.1038/nature07728))
39. Onodera CS, Underwood JG, Katzman S, Jacobs F, Greenberg D, Salama SR, Haussler D. 2012 Gene isoform specificity through enhancer-associated antisense transcription. *PLoS ONE* **7**, e43511. ([doi:10.1371/journal.pone.0043511](https://doi.org/10.1371/journal.pone.0043511))
40. Tani H, Mizutani R, Salam KA, Tano K, Ijiri K, Wakamatsu A, Isogai T, Suzuki Y, Akimitsu N. 2012 Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* **22**, 947–956. ([doi:10.1101/gr.130559.111](https://doi.org/10.1101/gr.130559.111))
41. Losos JB. 2011 Convergence, adaptation, and constraint. *Evolution* **65**, 1827–1840. ([doi:10.1111/j.1558-5646.2011.01289.x](https://doi.org/10.1111/j.1558-5646.2011.01289.x))
42. Alvarez del Villar J. 1948 Descripción de una nueva especie de *Mollienisia* capturada en Baños del Azufre, Tabasco (Pisces, Poeciliidae) [Description of a new species of *Mollienisia* captured in Baños del Azufre, Tabasco (Fishes, Poeciliidae)]. *An. Esc. Nac. Cienc. Biol.* **V**, 275–281. [In Spanish.]
43. Duttke SH *et al.* 2022 Decoding transcription regulatory mechanisms associated with *Coccidioides immitis* phase transition using total RNA. *mSystems* **7**, e01404-21. ([doi:10.1128/msystems.01404-21](https://doi.org/10.1128/msystems.01404-21))
44. Andrews S. 2010 *FastQC: a quality control tool for high throughput sequence data*. See <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
45. Warren WC *et al.* 2018 Clonal polymorphism and high heterozygosity in the celibate genome of the Amazon molly. *Nat. Ecol. Evol.* **2**, 669–679. ([doi:10.1038/s41559-018-0473-y](https://doi.org/10.1038/s41559-018-0473-y))
46. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013 STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21. ([doi:10.1093/bioinformatics/bts635](https://doi.org/10.1093/bioinformatics/bts635))
47. Robinson MD, McCarthy DJ, Smyth GK. 2010 edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140. ([doi:10.1093/bioinformatics/btp616](https://doi.org/10.1093/bioinformatics/btp616))
48. Krueger F. 2014 *Trim Galore!: a wrapper tool around Cutadapt and FastQC to consistently apply quality and adapter trimming to FastQ files*. See [https://www.bioinformatics.babraham.ac.uk/projects/trim\\_galore/](https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/).
49. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019 Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **37**, 907–915. ([doi:10.1038/s41587-019-0201-4](https://doi.org/10.1038/s41587-019-0201-4))
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009 The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. ([doi:10.1093/bioinformatics/btp352](https://doi.org/10.1093/bioinformatics/btp352))
51. Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016 Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650–1667. ([doi:10.1038/nprot.2016.095](https://doi.org/10.1038/nprot.2016.095))
52. McCarthy DJ, Chen Y, Smyth GK. 2012 Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic Acids Res.* **40**, 4288–4297. ([doi:10.1093/nar/gks042](https://doi.org/10.1093/nar/gks042))
53. Gearing LJ, Cumming HE, Chapman R, Finkel AM, Woodhouse IB, Luu K, Gould JA, Forster SC, Hertzog PJ. 2019 CiiDER: a tool for predicting and analysing transcription factor binding sites. *PLoS ONE* **14**, e0215495. ([doi:10.1371/journal.pone.0215495](https://doi.org/10.1371/journal.pone.0215495))
54. Sandelin A, Alkema W, Engström P, Wasserman WW, Lenhard B. 2004 JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res.* **32**, D91–D94. ([doi:10.1093/nar/gkh012](https://doi.org/10.1093/nar/gkh012))
55. Lim J-Y *et al.* 2021 *DNMT3A* haploinsufficiency causes dichotomous DNA methylation defects at enhancers in mature human immune cells. *J. Exp. Med.* **218**, e20202733. ([doi:10.1084/jem.20202733](https://doi.org/10.1084/jem.20202733))
56. Duttke SH *et al.* 2022 Glucocorticoid receptor-regulated enhancers play a central role in the gene regulatory networks underlying drug addiction. *Front. Neurosci.* **16**, 858427. ([doi:10.3389/fnins.2022.858427](https://doi.org/10.3389/fnins.2022.858427))

57. Arnold PR, Wells AD, Li XC. 2020 Diversity and emerging roles of enhancer RNA in regulation of gene expression and cell fate. *Front. Cell Dev. Biol.* **7**, 377. ([doi:10.3389/fcell.2019.00377](https://doi.org/10.3389/fcell.2019.00377))

58. Azofeifa JG, Allen MA, Hendrix JR, Rubin JD, Dowell RD. 2018 Enhancer RNA profiling predicts transcription factor activity. *Genome Res.* **28**, 334–344. ([doi:10.1101/gr.225755.117](https://doi.org/10.1101/gr.225755.117))

59. Kaikkonen MU, Lam MTY, Glass CK. 2011 Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovasc. Res.* **90**, 430–440. ([doi:10.1093/cvr/cvr097](https://doi.org/10.1093/cvr/cvr097))

60. Poole RC, Halestrap AP. 1993 Transport of lactate and other monocarboxylates across mammalian plasma membranes. *Am. J. Physiol. Cell Physiol.* **264**, C761–C782. ([doi:10.1152/ajpcell.1993.264.4.C761](https://doi.org/10.1152/ajpcell.1993.264.4.C761))

61. Jefferies CA. 2019 Regulating IRFs in IFN driven disease. *Front. Immunol.* **10**, 325. ([doi:10.3389/fimmu.2019.00325](https://doi.org/10.3389/fimmu.2019.00325))

62. Bazhanov N, Escaffre O, Freiberg AN, Garofalo RP, Casola A. 2017 Broad-range antiviral activity of hydrogen sulfide against highly pathogenic RNA viruses. *Scient. Rep.* **7**, 41029. ([doi:10.1038/srep41029](https://doi.org/10.1038/srep41029))

63. Doré LC, Chlon TM, Brown CD, White KP, Crispino JD. 2012 Chromatin occupancy analysis reveals genome-wide GATA factor switching during hematopoiesis. *Blood* **119**, 3724–3733. ([doi:10.1182/blood-2011-09-380634](https://doi.org/10.1182/blood-2011-09-380634))

64. Ku C-J, Hosoya T, Maillard I, Engel JD. 2012 GATA-3 regulates hematopoietic stem cell maintenance and cell-cycle entry. *Blood* **119**, 2242–2251. ([doi:10.1182/blood-2011-07-366070](https://doi.org/10.1182/blood-2011-07-366070))

65. Han GC *et al.* 2016 Genome-wide organization of GATA1 and TAL1 determined at high resolution. *Mol. Cell. Biol.* **36**, 157–172. ([doi:10.1128/MCB.00806-15](https://doi.org/10.1128/MCB.00806-15))

66. Holtzinger A, Evans T. 2007 Gata5 and Gata6 are functionally redundant in zebrafish for specification of cardiomyocytes. *Dev. Biol.* **312**, 613–622. ([doi:10.1016/j.ydbio.2007.09.018](https://doi.org/10.1016/j.ydbio.2007.09.018))

67. Klingerman CM, Trushin N, Prokopczyk B, Haouzi P. 2013 H<sub>2</sub>S concentrations in the arterial blood during H<sub>2</sub>S administration in relation to its toxicity and effects on breathing. *Am. J. Physiol. Integr. Comp. Physiol.* **305**, R630–R638. ([doi:10.1152/ajpregu.00218.2013](https://doi.org/10.1152/ajpregu.00218.2013))

68. Chou C-F, Tohari S, Brenner S, Venkatesh B. 2004 Erythropoietin gene from a teleost fish, *Fugu rubripes*. *Blood* **104**, 1498–1503. ([doi:10.1182/blood-2003-10-3404](https://doi.org/10.1182/blood-2003-10-3404))

69. Yokota A, Huo L, Lan F, Wu J, Huang G. 2020 The clinical, molecular, and mechanistic basis of RUNX1 mutations identified in hematological malignancies. *Mol. Cells* **43**, 145–152. ([doi:10.14348/molcells.2019.0252](https://doi.org/10.14348/molcells.2019.0252))

70. Riddell A, McBride M, Braun T, Nicklin SA, Cameron E, Loughrey CM, Martin TP. 2020 RUNX1: an emerging therapeutic target for cardiovascular disease. *Cardiovasc. Res.* **116**, 1410–1423. ([doi:10.1093/cvr/cvaa034](https://doi.org/10.1093/cvr/cvaa034))

71. Cutz E, Pan J, Yeger H, Domnik NJ, Fisher JT. 2013 Recent advances and controversies on the role of pulmonary neuroepithelial bodies as airway sensors. *Semin. Cell Dev. Biol.* **24**, 40–50. ([doi:10.1016/j.semcd.2012.09.003](https://doi.org/10.1016/j.semcd.2012.09.003))

72. Hockman D *et al.* 2017 Evolution of the hypoxia-sensitive cells involved in amniote respiratory reflexes. *eLife* **6**, e21231. ([doi:10.7554/eLife.21231](https://doi.org/10.7554/eLife.21231))

73. Kasim M, Benko E, Winkelmann A, Mrowka R, Staudacher JJ, Persson PB, Scholz H, Meier JC, Fähling M. 2014 Shutdown of achaete-scute homolog-1 expression by heterogeneous nuclear ribonucleoprotein (hnRNP)-A2/B1 in hypoxia. *J. Biol. Chem.* **289**, 26 973–26 988. ([doi:10.1074/jbc.M114.579391](https://doi.org/10.1074/jbc.M114.579391))

74. Wu Q, Li W, You C. 2021 The regulatory roles and mechanisms of the transcription factor FOXF2 in human diseases. *PeerJ* **9**, e10845. ([doi:10.7717/peerj.10845](https://doi.org/10.7717/peerj.10845))

75. Hellqvist M, Mahlapuu M, Samuelsson L, Enerbäck S, Carlsson P. 1996 Differential activation of lung-specific genes by two forkhead proteins, FREAC-1 and FREAC-2. *J. Biol. Chem.* **271**, 4482–4490. ([doi:10.1074/jbc.271.8.4482](https://doi.org/10.1074/jbc.271.8.4482))

76. Tharappel JC, Cholewa J, Espandiari P, Spear BT, Gairola CG, Glauert HP. 2010 Effects of cigarette smoke on the activation of oxidative stress-related transcription factors in female A/J mouse lung. *J. Toxicol. Environ. Health A* **73**, 1288–1297. ([doi:10.1080/15287394.2010.484708](https://doi.org/10.1080/15287394.2010.484708))

77. Gachon F, Olela FF, Schaad O, Descombes P, Schibler U. 2006 The circadian PAR-domain basic leucine zipper transcription factors DBP, TEF, and HLF modulate basal and inducible xenobiotic detoxification. *Cell Metab.* **4**, 25–36. ([doi:10.1016/j.cmet.2006.04.015](https://doi.org/10.1016/j.cmet.2006.04.015))

78. Gavriouchkina D, Fischer S, Ivacevic T, Stolte J, Benes V, Dekens MPS. 2010 Thyrotroph embryonic factor regulates light-induced transcription of repair genes in zebrafish embryonic cells. *PLoS One* **5**, e12542. ([doi:10.1371/journal.pone.0012542](https://doi.org/10.1371/journal.pone.0012542))

79. Ritchie A, Gutierrez O, Fernandez-Luna JL. 2009 PAR bZIP-bik is a novel transcriptional pathway that mediates oxidative stress-induced apoptosis in fibroblasts. *Cell Death Differ.* **16**, 838–846. ([doi:10.1038/cdd.2009.13](https://doi.org/10.1038/cdd.2009.13))

80. Reid NM *et al.* 2016 The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. *Science* **354**, 1305–1308. ([doi:10.1126/science.aah4993](https://doi.org/10.1126/science.aah4993))

81. Jacobs A, Elmer KR. 2021 Alternative splicing and gene expression play contrasting roles in the parallel phenotypic evolution of a salmonid fish. *Mol. Ecol.* **30**, 4955–4969. ([doi:10.1111/mec.15817](https://doi.org/10.1111/mec.15817))

82. Whitmarsh AJ, Davis RJ. 2000 Regulation of transcription factor function by phosphorylation. *Cell. Mol. Life Sci. C* **57**, 1172–1183. ([doi:10.1007/PL000000757](https://doi.org/10.1007/PL000000757))

83. Chilov D, Camenisch G, Kvietikova I, Ziegler U, Gassmann M, Wenger RH. 1999 Induction and nuclear translocation of hypoxia-inducible factor-1 (HIF-1): heterodimerization with ARNT is not necessary for nuclear accumulation of HIF-1 $\alpha$ . *J. Cell Sci.* **112**, 1203–1212. ([doi:10.1242/jcs.112.8.1203](https://doi.org/10.1242/jcs.112.8.1203))

84. Amoutzias GD, Robertson DL, Van de Peer Y, Oliver SG. 2008 Choose your partners: dimerization in eukaryotic transcription factors. *Trends Biochem. Sci.* **33**, 220–229. ([doi:10.1016/j.tibs.2008.02.002](https://doi.org/10.1016/j.tibs.2008.02.002))

85. Perry BW, McGowan KL, Arias-Rodriguez L, Duttke SH, Tobler M, Kelley JL. 2024 Data from: Nascent transcription reveals regulatory changes in extremophile fishes inhabiting hydrogen sulfide-rich environments. Dryad Digital Repository. ([doi:10.5061/dryad.83bk3jb0q](https://doi.org/10.5061/dryad.83bk3jb0q))

86. Perry BW, McGowan KL, Arias-Rodriguez L, Duttke SH, Tobler M, Kelley JL. 2024 Data from: Nascent transcription reveals regulatory changes in extremophile fishes inhabiting hydrogen sulfide-rich environments. Figshare. ([doi:10.6084/m9.figshare.c.7247128](https://doi.org/10.6084/m9.figshare.c.7247128))