Title

Local ancestry analysis reveals genomic convergence in extremophile fishes

Journal

Philosophical Transactions of the Royal Society B Biological Sciences, 374(1777)

ISSN

0962-8436

Authors

Brown, Anthony P
McGowan, Kerry L
Schwarzkopf, Enrique J
et al.

Publication Date

2019-07-22

DOI

10.1098/rstb.2018.0240

Peer reviewed

PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Brown AP, McGowan KL, Schwarzkopf EJ, Greenway R, Rodriguez LA, Tobler M, Kelley JL. 2019 Local ancestry analysis reveals genomic convergence in extremophile fishes. *Phil. Trans. R. Soc. B* **374**: 20180240.
http://dx.doi.org/10.1098/rstb.2018.0240

Accepted: 8 March 2019

One contribution of 16 to a theme issue 'Convergent evolution in the genomics era: new insights and directions'.

Subject Areas:
evolution, genomics

Keywords:
extremophile fishes, *Poecilia mexicana*, hydrogen sulfide, convergent evolution, genome sequencing

Authors for correspondence:
Michael Tobler
e-mail: tobler@ksu.edu
Joanna L. Kelley
e-mail: joanna.l.kelley@wsu.edu

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4467434.

THE ROYAL SOCIETY PUBLISHING

Local ancestry analysis reveals genomic convergence in extremophile fishes

Anthony P. Brown[1], Kerry L. McGowan[1], Enrique J. Schwarzkopf[1], Ryan Greenway[2], Lenin Arias Rodriguez[3], Michael Tobler[2] and Joanna L. Kelley[1]

[1]School of Biological Sciences, Washington State University, 100 Dairy Road, Pullman, WA 99164, USA
[2]Division of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS 66506, USA
[3]División Académica de Ciencias Biológicas, Universidad Juárez Autónoma de Tabasco (UJAT), CP 86150 Villahermosa, Tabasco, México

APB, 0000-0001-6100-2470; KLM, 0000-0001-6388-3873; RG, 0000-0002-7182-7932; MT, 0000-0002-0326-0890; JLK, 0000-0002-7731-605X

The molecular basis of convergent phenotypes is often unknown. However, convergence at a genomic level is predicted when there are large population sizes, gene flow among diverging lineages or strong genetic constraints. We used whole-genome resequencing to investigate genomic convergence in fishes (*Poecilia* spp.) that have repeatedly colonized hydrogen sulfide (H_2S)-rich environments in Mexico. We identified genomic similarities in both single nucleotide polymorphisms (SNPs) and structural variants (SVs) among independently derived sulfide spring populations, with approximately 1.2% of the genome being shared among sulfidic ecotypes. We compared these convergent genomic regions to candidate genes for H_2S adaptation identified from transcriptomic analyses and found that a significant proportion of these candidate genes (8%) were also in regions where sulfidic individuals had similar SNPs, while only 1.7% were in regions where sulfidic individuals had similar SVs. Those candidate genes included genes involved in sulfide detoxification, the electron transport chain (the main toxicity target of H_2S) and other processes putatively important for adaptation to sulfidic environments. Regional genomic similarity across independent populations exposed to the same source of selection is consistent with selection on standing variation or introgression of adaptive alleles across divergent lineages. However, combined with previous analyses, our data also support that adaptive changes in mitochondrially encoded subunits arose independently via selection on de novo mutations. Pressing questions remain on what conditions ultimately facilitate the independent rise of adaptive alleles at the same loci in separate populations, and thus, the degree to which evolution is repeatable or predictable.

This article is part of the theme issue 'Convergent evolution in the genomics era: new insights and directions'.

1. Introduction

Convergent evolution—the evolution of similar traits in response to shared sources of selection—is a common phenomenon of biological diversification [1,2]. Convergence is evident in molecular traits, like gene expression [3], as well as complex emergent phenotypes, like development [4], morphology [5] or behaviour [6,7]. By contrast, convergence at a genomic level is much less common [8,9], and the molecular basis of convergent phenotypes is often complex and non-convergent [10–12]. One reason for the relative paucity of genomic convergence is probably the functional redundancy built into genomes, where alternative modifications have equivalent impacts on trait expression and organismal performance [13–16]. However, convergence at a genomic level might be expected when selection acts on standing genetic variation [17,18], when adaptive alleles are introgressed across divergent lineages

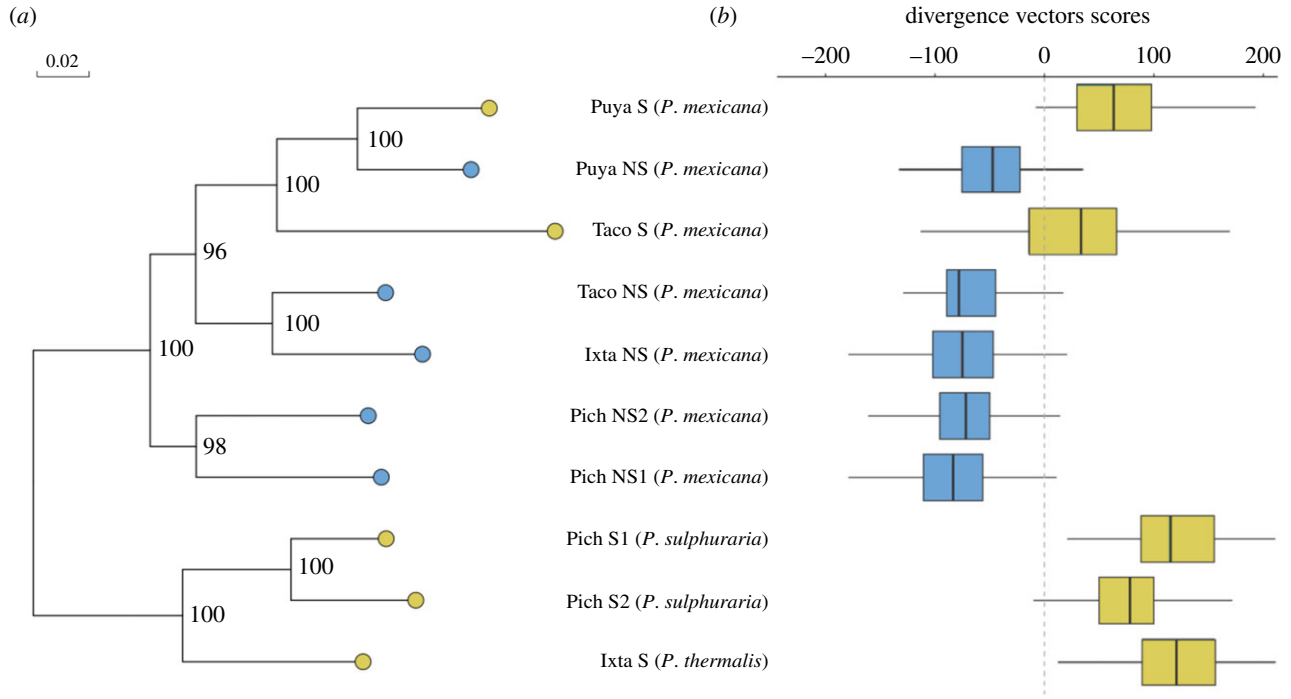

Figure 1. The tree in (*a*) represents the maximum-likelihood phylogeny of individuals from populations of the *P. mexicana* species complex; Pichucalo (Pich), Tacotalpa (Taco), Puyacatengo (Puya), Ixtapangajoya (Ixta), from southern Mexico. Populations in sulfide springs are indicated by yellow tips, populations in non-sulfidic environments are represented by blue tips. (*b*) An example of convergent phenotypic evolution in the same populations. Specifically, the figure summarizes variation in body shape summarized as divergence vectors scores [22]. Positive scores observed in sulfide spring populations are associated with large heads compared to negative scores observed in populations from non-sulfidic habitats.

[19,20], or when there are strong genetic constraints that limit the outcome of selection [21]. Here, we examined patterns of evolution in a unique system with characteristics that are conducive to the emergence of convergence at the genomic level.

In southern Mexico, small livebearing fishes of the *Poecilia mexicana* species complex have independently colonized multiple hydrogen sulfide (H_2S)-rich springs that occur in different rivers of the Río Grijalva basin [22]. H_2S is a strong source of selection. As a naturally occurring toxicant, it interferes with oxidative phosphorylation in mitochondria and inhibits cellular respiration [23]; hence, exposure to micromolar concentrations of H_2S is lethal to most metazoans [24,25]. *Poecilia* populations in sulfide springs are locally adapted and can tolerate H_2S concentrations that are orders of magnitudes above this toxicity threshold [26]. They have diverged phenotypically and genetically from ancestral populations in adjacent non-sulfidic habitats, even though spatial distances are small (typically less than 100 m) and physical barriers to fish movement are lacking [26,27]. Phenotypes across independently derived sulfide spring populations in different river drainages exhibit strong patterns of convergent evolution, such as enlarged heads and increased gill surface area in sulfidic populations, which probably improve oxygen acquisition in sulfidic environments (general pattern of convergence is shown in figure 1; [26]). Convergent phenotypes include gene expression and traits associated with physiological, morphological, behavioural and life-history adaptation (see [28] for a review). By contrast, a previous study has concluded that the underlying genetic mechanisms are largely unique to specific sulfide spring lineages (i.e. non-convergent), with selection acting on de novo mutations rather than standing genetic variation [29]. However, the population genetic approach employed by Pfenninger *et al.* [29] was not conducive to analysing patterns of ancestry of alternative alleles that may have uncovered genomic convergence. Genomic convergence might be expected in this system for several reasons: (i) *P. mexicana* is a widespread and abundant species in non-sulfidic environments of the Río Grijalva basin and other river basins in Mexico and Central America [30,31]. The large population sizes of this species provide a pool of rare, potentially adaptive alleles that selection can act upon during colonization of sulfide springs (i.e. selection on standing genetic variation); (ii) despite significant genetic differentiation between adjacent populations in sulfidic and non-sulfidic habitats, low rates of gene flow are present and may export adaptive alleles from the sulfidic to non-sulfidic populations, from which they can be introgressed into, or facilitate the colonization of, sulfide springs in other parts of the Río Grijalva basin (transporter hypothesis; [32]); and (iii) H_2S is a strong source of selection with clear biochemical consequences. Specifically, it binds to cytochrome *c* oxidase in the mitochondrial respiratory chain and is detoxified through a highly conserved physiological pathway associated with sulfide : quinone oxidoreductase (SQR; [25]). Hence, evolution of tolerance to H_2S may be constrained to specific genes associated with H_2S toxicity or detoxification.

In order to identify potential convergence at a genomic level, we sequenced 10 new genomes of *Poecilia* at high coverage, including specimens from all known sulfide springs and adjacent populations in non-sulfidic habitats in the Río Grijalva basin. We employed two complementary genome-wide scans to identify genomic regions where sulfide spring lineages were similar to each other but different from individuals from non-sulfidic habitats, one based on single nucleotide polymorphisms (SNPs) and another based on genomic structural variation. We assessed whether there were biological process gene ontology (GO) terms

over-represented in genes showing evidence for convergence in SNPs or structural variants (SVs). Finally, we tested whether candidate genes for H_2S adaptation identified from transcriptomic analyses were associated with specific types of genomic convergence.

2. Methods

(a) Sample collection and library preparation for whole-genome sequencing

Samples were collected from 10 sites in the Río Grijalva basin, including one individual each from five sulfidic and five non-sulfidic habitats (electronic supplementary material, figure S1 and table S1). Specimens were sacrificed in a buffered MS222 solution, and muscle tissue was dissected and preserved in 96% ethanol. DNA was extracted using the MagAttract High Molecular Weight DNA extraction kit, quantified using a Qubit fluorometer and then visualized on a 0.5% agarose gel to confirm the presence of high molecular weight DNA (greater than 10 kb). Illumina's TruSeq DNA PCR-Free LT Library Preparation Kit was then used to prepare libraries for sequencing. The DNA was sheared using a Covaris M220 with the 550 bp insert size settings. We followed the protocol in the user guide for the rest of the library preparation and used compatible indexed adapters for multiplexing. Quantitative polymerase chain reaction (qPCR) was used to determine the concentration of DNA successfully ligated to the Illumina adapters in each sample. We used the KAPA Library Quantification kit for Illumina Sequencing Platforms along with the StepOnePlus Real-Time PCR System. Sample dilutions of 1 : 10 000 and 1 : 20 000 were used. Each sample dilution, each standard and the non-template control were run in triplicate. Estimated concentrations from the qPCR results allowed us to create two equimolar pools of the libraries. The libraries were paired-end sequenced with a read length of 100 bp on an Illumina HiSeq 2500 at the Washington State University Genomics Core.

(b) Single-nucleotide polymorphism calling and filtering

Reads were trimmed using Trim Galore! [33] with the following trimming: (i) bases with a quality less than 20, (ii) 3′ ends of reads that had at least six bases matching the Illumina adapter sequence were trimmed, (iii) the first five bases of each read owing to inconsistent base sequence composition, and (iv) any reads that were shorter than 50 bp or that had a paired read that was shorter than 50 bp were discarded. Reads were mapped to the *P. mexicana* reference genome (NCBI accession: GCA_001443325.1; [34]) using the BWA-MEM algorithm from the Burrows–Wheeler aligner (BWA, v. 0.7.12-r1039; [35]). Bam files that were from the same individuals but different lanes were merged using the *MergeSamFiles* utility in Picard.

SNPs were initially called on a per-individual basis using the *EMIT_ALL_SITES* option in the *UnifiedGenotyper* module of the Genome Analysis Toolkit (GATK v. 3.5; [36–38]). The individual vcf files were then merged using the *vcfmerge* Perl module from Vcftools (v. 0.1.15; [39]). The combined vcf file was subsequently filtered using Vcftools to only retain biallelic sites (*–min-alleles 2* and *–max-alleles 2*) with no missing data (*–max-missing 1.0*). Genotypes with a quality score below 30 (*–minQ 30*) and genotypes supported by fewer than 10 reads *(–minDP 10)* were discarded prior to applying the missing data filter. Scripts for the analyses can be found at https://github.com/jokelley/Pmex-GenomicConvergence.

(c) Inferring a phylogeny from whole-genome data

We used the SNPhylo pipeline [40] to infer a maximum-likelihood phylogeny from our whole-genome data. SNPhylo first reduces SNP redundancy by filtering out SNPs in linkage disequilibrium (SNPs with a correlation coefficient greater than or equal to 0.1 were filtered, a total of 24 275 SNPs were retained), and then infers a maximum-likelihood tree from the remaining sites using *dnaml* from PHYLIP [41]. We used *Poecilia reticulata* (NCBI accession: SRP038017; [42]) as an outgroup, and raw reads were subjected to the same filtering and quality control steps described above. SNPhylo was then run with default settings, except that we specified the outgroup (*-o*), set the number of bootstrap replicates to 1000 (*-B 1000*) and set the number of chromosomes equal to the number of different contigs in our vcf (*-a*). The resulting phylogeny was visualized using Dendroscope [43].

(d) Analysis of local relationship patterns across the genome

To identify regions of the genome with shared ancestry among lineages, we used Saguaro [44]. Saguaro uses a combination of a hidden Markov model and a self-organizing map to build 'cacti' (matrices of pairwise genetic distance between samples) that describe local phylogenetic relationships among samples. Briefly, Saguaro initially builds one cactus (topology) to fit the entire genome, then uses regions that do not fit the initial cactus to hypothesize a new cactus. This process is then iterated a user-defined number of times, and a new cactus is hypothesized in each iteration. At the end of the last iteration, each SNP-containing region is assigned to the cactus that best fits the pattern of variation in that region.

To run Saguaro, we converted our vcf (without the *P. reticulata* outgroup) into hidden Markov model format using the *VCF2HMMFeature* command within Saguaro [44]. After converting the file, we ran Saguaro (*Saguaro* command) for 29 iterations (*-i* 29) resulting in 30 cacti that each best described the local ancestry for at least 100 kb of the genome. We used the Neighbor module from PHYLIP [41] to generate unrooted neighbour-joining trees from the cacti, which were visualized in Dendroscope [43].

(e) Analysis of structural variation across the genome

We also identified shared SVs among lineages in sulfidic environments. To identify SVs, we used SVMerge (v. 1.2r37) [45]. SVs included insertions, deletions, duplications and inversions greater than or equal to 100 bp in length relative to the *P. mexicana* reference genome. We integrated results from three separate SV callers: BreakDancerMax (v. 1.3.6) [46], Pindel (v. 0.2.4) [47] and SECluster [45], and then validated the calls via local assembly using Velvet [48]. Default parameter settings were used for each of these tools. Variants that passed the validation steps were retained in the final set. Variants that were present in at least nine out of 10 individuals were filtered, as these were probably SVs present in all individuals.

(f) Identifying convergence at a genomic level

We considered regions where either all five sulfidic individuals, all five sulfidic individuals and one non-sulfidic individual, or four sulfidic individuals either clustered together or shared SVs as showing a pattern of convergence. We used this classification based on a prior study that indicated that two of the sulfidic populations (Puyacatengo and Tacotalpa) exhibited few shared SNPs, suggesting unique pathways to adaptation [29]. Moreover,

other studies of convergent evolution have classified convergent regions in a similar manner [18].

To annotate regions of the genome with strong signals of clustering by ecotype in the SAGUARO analysis, we identified genes within each region using bedtools *intersect* (v. 2.25.0; [49]). To identify SVs that were shared by sulfidic individuals, we used MULTOVL (v. 1.3) [50]. We then identified genes that overlapped with the shared SVs, and SVs that overlapped with regions that clustered by ecotype in the SAGUARO analysis using bedtools [49].

(g) Annotation of convergent genomic regions

We used GORILLA [51] to identify enriched biological process GO terms in the genes that were assigned to each cactus showing strong ecotype clustering and separately identified enriched terms in genes contained in convergent SVs. We used all *P. mexicana* genes (NCBI accession: GCA_001443325.1; [34]) as the reference set. Annotations for the reference set were previously generated via a BLASTX search against the human SwissProt database [52]. GO terms were considered over-represented at a *p*-value of lesser than or equal to 0.001.

(h) Comparing candidate genes to genes that clustered by ecotype

To generate a list of candidate genes that may be important for adaptation to hydrogen sulfide-rich environments, we reanalysed RNA-sequencing data from a previous study [53]. We identified genes with convergent changes in gene expression between sulfidic and non-sulfidic populations in three of the drainages (Pichucalco, Tacotalpa, Puyacatengo) as candidate genes mediating responses to H_2S (see the electronic supplementary material). We compared these genes to the regions of the genome that exhibited evidence for convergence in either the SAGUARO or SV analysis. We conducted a Fisher's exact test to determine whether there were significantly more candidate genes in convergent regions than expected by chance.

3. Results

(a) Genome-wide and local relationship patterns

We sequenced 10 genomes from five sulfidic and five non-sulfidic populations of the *P. mexicana* species complex to high coverage (see the electronic supplementary material, table S1). After filtering, we identified 8 538 973 SNPs across the genome. Maximum-likelihood analyses of genome-wide SNP data using SNPHYLO resulted in a highly supported tree (figure 1), which is consistent with the previously inferred independent colonization of sulfide springs and distinct evolutionary trajectories for the different sulfide spring lineages [22,54,55].

Classifying patterns of genetic similarity into the 30 most common patterns (cacti/topologies) using SAGUARO [44] allowed us to investigate the prevalence of convergence across the genome (table 1; electronic supplementary material, figure S2). The topology of the most frequently assigned cactus (cactus 1) was consistent with the maximum-likelihood tree (figure 1) and best characterized approximately 81% of the genome. Six of the 30 cacti (cactus 6, 13, 16, 18, 24 and 25; table 1) exhibited a strong signal of clustering by ecotype (sulfidic versus non-sulfidic; figure 2), indicating convergent aspects of genome evolution. The most commonly assigned cactus in this set was no. 6 (corresponding to approx. 0.7% of the genome), which clustered the sulfide spring individuals of the Pichucalco, Ixtapangajoya and Puyacatengo drainages opposite to all non-sulfidic individuals plus the sulfide spring individual of the Tacotalpa drainage. Around 0.1% of the genome was assigned to cactus 18 that juxtaposed all individuals from sulfide springs together against all individuals from non-sulfidic habitats. There were 967 genes that were assigned to cacti that clustered by ecotype (electronic supplementary material, table S2).

Table 1. Local ancestry across the genome, as described by 30 topologies (cacti) hypothesized by SAGUARO [44]. (Highlighted rows are cacti with clustering by ecotype (sulfidic versus non-sulfidic) (see also the electronic supplementary material, table S2).)

cactus	length (bp)	% of length	no. regions	% of regions
1	599 102 149	80.66	8507	56.03
2	67 822 137	9.13	2776	18.28
3	21 852 583	2.94	369	2.43
4	10 258 179	1.38	504	3.32
5	6 505 330	0.88	314	2.07
6	5 253 321	0.71	305	2.01
7	4 584 980	0.62	299	1.97
8	4 534 583	0.61	160	1.05
9	3 840 449	0.52	246	1.62
10	3 801 264	0.51	225	1.48
11	3 051 456	0.41	230	1.51
12	1 640 440	0.22	134	0.88
13	1 538 517	0.21	139	0.92
14	1 122 496	0.15	137	0.9
15	1 079 620	0.15	31	0.2
16	967 373	0.13	132	0.87
17	814 436	0.11	74	0.49
18	639 869	0.09	59	0.39
19	613 559	0.08	84	0.55
20	529 482	0.07	66	0.43
21	481 642	0.06	79	0.52
22	475 128	0.06	23	0.15
23	474 919	0.06	68	0.45
24	470 041	0.06	54	0.36
25	351 943	0.05	32	0.21
26	251 569	0.03	32	0.21
27	225 882	0.03	27	0.18
28	164 949	0.02	37	0.24
29	137 840	0.02	22	0.14
30	131 260	0.02	19	0.13

(b) Structural variation in the genome

To characterize possible convergent changes in genome structure, we identified SVs (insertions, deletions, duplications and inversions greater than or equal to 100 bp) (electronic supplementary material, table S3). The SVs covered on average 31 555 221 $\pm$ 2 312 757 bp (mean $\pm$ s.d.), or 3.9 $\pm$ 0.29% of the genome (electronic supplementary material, table S3).

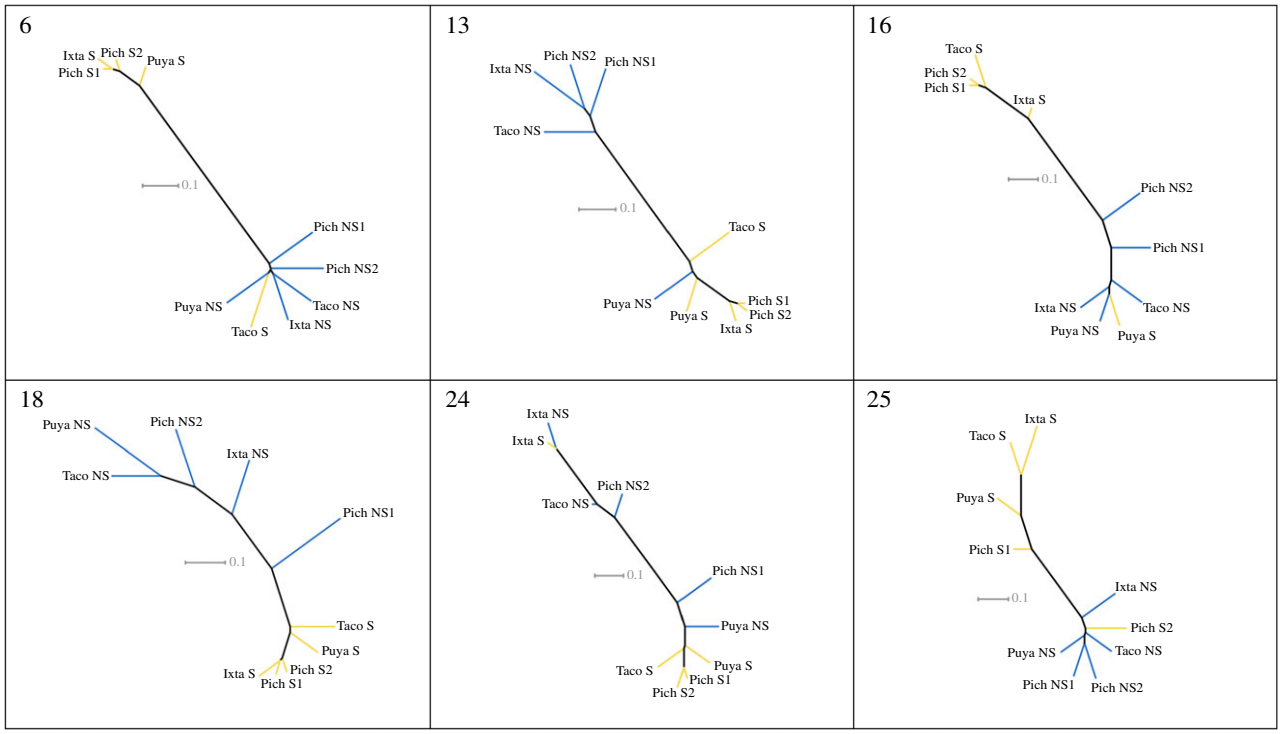

Figure 2. Six of the 30 cacti exhibited a strong signal of clustering by ecotype (sulfidic versus non-sulfidic). These six cacti cover approximately 1.2% of the genome.

The majority (55.7%) of all SVs were unique to an individual (electronic supplementary material, figure S3). A total of 125 SVs were shared by at least four sulfidic individuals with at most one non-sulfidic individual (electronic supplementary material, table S4). These SVs that were shared by sulfidic individuals covered a total of 433 303 bp (0.05% of the genome). The 125 convergent SVs overlapped with a total of 80 unique genes (electronic supplementary material, table S5).

(c) Gene ontology analysis of convergent genomic regions

We identified enriched biological process GO terms in the set of genes that were located in convergent genomic regions (cactus 6, 13, 16, 18, 24 and 25) using GORILLA (electronic supplementary material, table S6). When considering all genes assigned to any of the cacti with clustering by ecotype, there were 26 enriched terms, including *hydrogen sulfide metabolic process* (GO: 0070813), *oxalate transport* (GO: 0019532), *carboxylic acid transport* (GO: 0046942) and 15 terms related to transport of ions or other materials (electronic supplementary material, table S6). We also identified enriched Biological Process GO terms in the genes that contained convergent SVs. These 80 genes were enriched for two Biological Process GO terms: *hepatocyte growth factor receptor signalling pathway* (GO: 0048012) and *negative regulation of NIK/NF-κB signalling*. (GO: 1901223).

(d) Comparison of differentially expressed candidate genes and convergent regions of the genome

We assessed whether any candidate genes for H_2S adaptation identified from transcriptomic analyses (electronic supplementary material, table S7) were located in regions that best fit one of the six cacti that clustered individuals by ecotype. Out of the 299 candidate genes, 25 displayed clustering by ecotype in the SAGUARO analysis. This was significantly more than expected by chance (random expectation = 10.8, Fisher's exact test $p = 0.0001$). These 25 genes included two major genes involved in sulfide detoxification, *SQRDL* and *ETHE1* [56,57]. The region around and including *SQRDL* was assigned to the cactus splitting all sulfidic individuals from all non-sulfidic individuals (cactus 18, figure 3). Whereas, a region around *ETHE1* was best represented by the cactus separating the sulfidic individuals from the Pichucalco, Ixtapangajoya and Puyacatengo drainages from all the other individuals (cactus 6). Other differentially expressed genes that were assigned to some of these cacti included one component of COX (*COX15*; [23,58]), a positive regulator of COX (*HIG1*; [59]), a gene involved in sulfur processing (*MPST*; [60]), and one chloride channel gene (*CLCN2*; hydrogen sulfide is a known modulator of ion channels; [61]) (electronic supplementary material, table S2).

We also assessed whether any candidate genes overlapped SVs that were shared among sulfidic individuals. Of the 80 genes in convergent SVs, five were in the candidate gene set (electronic supplementary material, table S7), which was significantly more than expected by chance (random expectation = 0.9, Fisher's exact test $p = 0.002$). Two of the genes, *SLC25A5*, an anion transporter, and *GCLC*, a rate-limiting enzyme in the synthesis of glutathione [62], contained duplications in sulfidic individuals and were upregulated in sulfidic populations. Only one gene that was differentially expressed contained an SV shared by sulfidic individuals and clustered by ecotype in the SAGUARO analysis (*SLC25A5*, an anion transporter).

4. Discussion

Extremophile populations from the *P. mexicana* species complex have adapted to H_2S-rich environments in southern

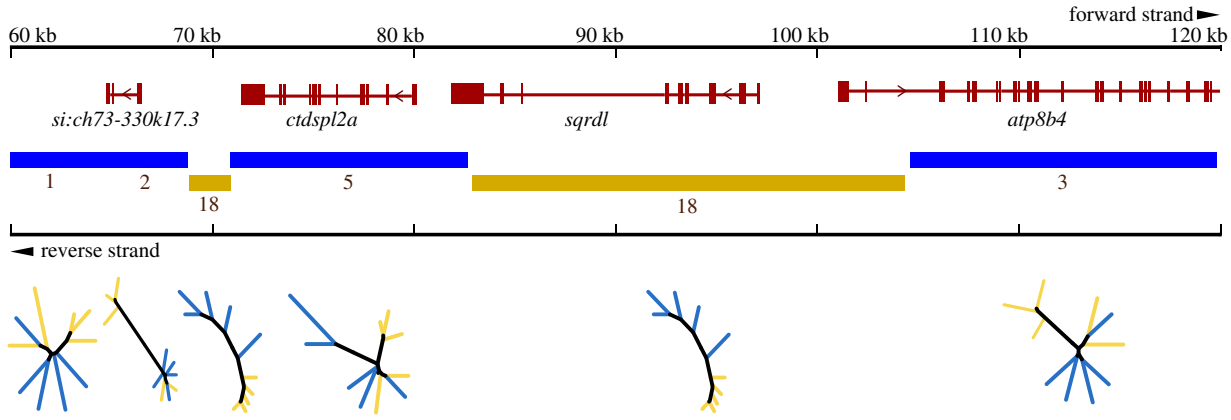

Figure 3. Local ancestry patterns around *sqrdl* (a key gene involved in sulfide detoxification). A portion of the genomic contig (NW_015096034.1) containing *sqrdl* is shown. Genes (with exons) in the region are shown, with the arrow representing the direction of transcription. Horizontal bars and numbers represent the cactus assigned to the corresponding region (blue indicates no ecotype clustering; yellow indicates clustering by ecotype). A cartoon of the cactus is included beneath each region (note: the genetic distances in these topologies are not to scale, see the electronic supplementary material, figure S2 for proper scaling). A region starting roughly 6.7 kb upstream of the first exon of *sqrdl* and ending roughly in the middle of the last exon of *sqrdl* was assigned to cactus 18, which split all sulfidic individuals from all non-sulfidic individuals.

Mexico and exhibit evolutionary convergence in traits associated with morphology [28], physiology [63], behaviour [64], life history [65] and transcription [53,66]. Here, we documented evidence for convergence at the genomic level. Using whole-genome sequencing data from individuals from five populations inhabiting sulfide springs and five populations in adjacent non-sulfidic habitats, we identified regions of the genome (approx. 1.2% in total) where individuals from sulfide springs in different river drainages were genetically similar. The extent of convergence documented here includes a greater proportion of the genome than that documented in prior studies with the same methodological approach, for example, derived freshwater sticklebacks (0.46%; [18]) and snowshoe hare species with similar winter coat colours (0.08%; [67]). Genomic regions with evidence for convergence included candidate genes involved in H_2S adaptation, including genes related to H_2S toxicity and detoxification. Additionally, we identified regions of the genome (approx. 0.05%) where sulfidic individuals shared convergent SVs, though relatively few of these contained previously identified candidate genes, implying that these SVs probably play a comparatively minor role in adaptation to H_2S-rich environments. Our data indicate a role for gene flow and or selection on standing genetic variation in shaping genomic convergence between independently derived sulfidic populations.

(a) Convergence occurs in regions with H_2S responsive candidate genes

Previous analyses of the molecular underpinnings of adaptation to sulfide springs have uncovered signatures of selection on, or differential expression of, genes associated with H_2S detoxification (e.g. *SQRDL* and *ETHE1*), the toxicity target in the mitochondrial respiratory chain (COX), other OXPHOS components, assembly proteins, and regulators (e.g. *COA4*, *HIG1*), as well as alternative physiological pathways that compensate for the inhibition of direct toxicity targets (anaerobic metabolism and oxidative stress responses; [29,53,66]). To quantify the proportion of candidate genes that showed evidence of genomic convergence, we compared genes that were consistently differentially expressed between sulfidic and non-sulfidic populations to regions of the genome with evidence for convergence (either by topology or shared SVs). Most prominently, *SQRDL*—the enzyme mediating the first step of H_2S detoxification [56]—is in a region of the genome where all sulfide spring individuals form a single cluster (figure 3), suggesting a shared origin of the derived alleles in all extremophile populations. Similarly, other genes related to H_2S processing, OXPHOS, anaerobic metabolism and responses to xenobiotics are present in regions where sulfide spring individuals cluster together.

Of the 25 differentially expressed candidate genes that were assigned to cacti that clustered by ecotype, 13 were in regions assigned to cactus no. 6 (figure 2), which represented the most frequent cactus with clustering of sulfide spring individuals. This cactus grouped the sulfide spring individual from the Tacotalpa drainage (El Azufre I) together with individuals from non-sulfidic habitats. This sulfide spring population has previously been hypothesized to achieve H_2S adaptation via different mechanisms. For example, the sulfide spring population in the Tacotalpa drainage has a COX that is susceptible to inhibition by H_2S similar to fish in non-sulfidic populations, whereas sulfidic individuals from the Puyacatengo and Pichucalco drainages evolved an H_2S-resistant COX that maintains function even in high H_2S conditions [54]. Together these data support that the Tacotalpa sulfidic population took a unique pathway to adaptation, even though Puyacatengo sulfidic population is more closely related to the Tacotalpa sulfidic population than the other sulfidic populations (figure 1; [22]).

Additional evidence for convergence in regions of the genome associated with H_2S adaptation comes from functional annotation that did not hinge on *a priori* predictions. The combined set of genes in all convergent SAGUARO topologies was enriched for *hydrogen sulfide metabolic processes*. Additional enriched GO terms in the combined set were related to ion transport, a relevant finding considering that H_2S is a known modulator of ion channels [61] and the ionic composition in sulfide springs differ vastly from that in nearby non-sulfidic habitats. Other analyses on differential expression [53] and sites under selection [66] in this system did not find enrichment of genes related to ion channels,

showing that our local ancestry-based analysis has the potential to illuminate new genomic regions that might be involved in adaptation to sulfidic environments. Finally, genes with shared SVs were enriched for GO terms associated with the *negative regulation of NIK/NF-κB signalling*, an interesting result considering that H_2S mediates the anti-apoptotic actions of the NF-κB transcription factor [68]. Overall, our analyses have indicated that genomic convergence includes a number of processes that have previously been implicated in mediating adaptation directly to H_2S or other environmental factors (e.g. hypoxia, elevated salinity) that are consistently correlated with the presence of H_2S in Mexico's sulfide springs (see [53] for an in-depth discussion).

(b) Distinguishing between different hypotheses for the origin of genomic convergence

Convergence at a genomic level can be the result of selection on standing genetic variation, adaptive introgression and genetic constraints that limit the outcomes of selection [18,21,69,70]. Considering our ancestry-based approach, convergent genomic regions identified here are probably not a consequence of genetic constraints, where selection on independently derived alleles with de novo mutations favours the evolution of shared genomic features in different lineages (simply because this mechanism does not predict monophyly of putatively adaptive alleles). Consequently, our results at first glance contradict the results of a previous study concluding that adaptation to sulfide springs (in the Puyacatengo and Tacotalpa drainages) was primarily driven by selection owing to de novo mutations, because there were few shared SNPs between sulfide spring populations, and putatively adaptive alleles were absent in the adjacent non-sulfidic populations [29]. However, previous conclusions may have been premature because (i) the ancestry relationships among alternative alleles were not previously considered (i.e. alternative major and fixed alleles in different sulfide spring lineages may still be of monophyletic origin, even though they may exhibit unique SNPs), (ii) available sample sizes were potentially too low to detect rare alleles in non-sulfidic populations, and (iii) the population level sampling was low, including only sulfide springs in the Puyacatengo and the Tacotalpa drainages, with the latter indeed exhibiting some unique mechanisms to H_2S adaptation (see above) that could be driven by selection on de novo mutation [29].

Based on our current data, we cannot distinguish whether convergent genomic regions were the product of selection on standing genetic variation or introgression. Both scenarios seem plausible considering the biology of our study system. *Poecilia mexicana*—the ancestral species of all sulfide spring populations—is one of the most common freshwater fish species in Mexico and Central America [30,31], providing a large reservoir of rare alleles that could be selected on upon colonization of extreme environments. In addition, the species can sometimes be found in marginal habitats like ditches and isolated ponds [71], where environmental conditions characteristic of sulfide springs (e.g. hypoxia and surges of H_2S) can occur intermittently, especially during periods of drought (e.g. see [72]). This scenario may actually maintain the frequency of alleles that are adaptive in sulfide springs at a reasonably high frequency even in populations residing in non-sulfidic habitats, potentially representing an exaptation for the initial colonization of permanently extreme habitats. Alternatively, adaptive alleles may have come to high frequency in a sulfide spring based on selection on de novo mutations and were then introgressed into other sulfide spring populations by gene flow. While there is strong selection against migrants into sulfide springs [27], there is evidence for low levels of migration between sulfidic and non-sulfidic populations [52,66,73], opening the possibility that gene flow from an original sulfide spring population to non-sulfidic populations generated standing genetic variation that ultimately allowed for the colonization of new sulfide springs (transporter hypothesis; [32]). This scenario is plausible because different river drainages with sulfide springs are interconnected, especially during the wet season, when flooding is pervasive [74]. In addition, sulfide springs in some drainages (e.g. Pichucalco) have been colonized much earlier than springs in others (e.g. Tacotalpa and Puyacatengo; [54,66]), perhaps suggesting that adaptive alleles originated in western drainages and subsequently moved eastwards. While a method to attempt to distinguish between these hypotheses has been developed [75], it requires population-scale data that are not currently available for our system, highlighting one limitation of our approach of using in-depth whole-genome data from single individuals per population.

It is important to highlight that our methodological bias to detect shared genomic regions that have arisen through selection on standing genetic variation or gene flow does not preclude a possible role for selection on independent de novo mutations playing a role in adaptation to sulfide springs and generating patterns of convergence at the genomic level. There is ample evidence for potentially adaptive genetic variation that is unique to specific sulfide spring lineages [29,66], and past research actually identified candidates of genomic convergence caused by genetic constraints [54]. H_2S directly inhibits COX [23], but sulfide spring populations in the Puyacatengo and Pichucalco drainages have evolved COX to maintain function in the presence of H_2S [54]. The evolution of an H_2S-resistant COX was hypothesized to be linked to a handful of amino acid substitutions in two mitochondrially encoded subunits of COX, which have clearly arisen independently in the Puyacatengo and Pichucalco drainages (electronic supplementary material, figure S4; [54]). In our analyses, this is reflected in mitochondrial genes occurring in a cactus that separates the highly divergent sulfide spring endemics of the Pichucalco and Ixtapangajoya drainages (i.e. *Poecilia sulphuraria* and *Poecilia thermalis*) from all the other populations that are nominally *P. mexicana*. While introgression of mitochondrial genomes across divergent lineages has been documented in a variety of systems [76–78], the absence of mitochondrial introgression in this instance provides evidence for constraints shaping genomic convergence in genes encoding the primary toxicity target of H_2S. Other instances of constraint causing convergence of a genomic level may be present in our study system, but currently available data are not suitable to identify such instances.

Convergence through gene flow and selection on standing genetic variation is comparatively easy to detect through comparative genomic analyses, and the approach employed in this study was not designed to detect selection on de novo mutations. Conversely, previous analyses that compared SNPs across multiple population pairs in the

absence of ancestry information were biased towards the detection of selection on de novo mutations. The difficulty of estimating the relative roles of gene flow, selection on standing genetic variation and selection on de novo mutations is not restricted to our study system, but it represents a general problem for evolutionary biology, both when we consider the importance of these mechanisms in the evolution of adaptive quantitative traits within study systems as well as adaptive evolution across study systems. Future studies will consequently require integrative analytical approaches that allow us to better understand how potential detection biases influence our interpretation of data. Thorough documentation of selection on de novo mutations—especially if they contribute to quantitative traits—is arguably difficult, because functional information is required to conclusively establish that independent de novo mutations actually have similar effects on organisms' phenotype and fitness. Consequently, we may generally underestimate the role of selection on de novo mutations in convergent genomic evolution.

5. Conclusion

We have documented genomic convergence in regions containing candidate genes associated with H_2S adaptation in the *P. mexicana* system. This represents a paradigm shift in our understanding of convergent evolution in independent lineages of sulfide spring fishes, in which adaptation has previously been assumed to be mediated primarily through selection on de novo mutations. Our findings also raise questions about whether introgression of adaptive alleles may be possible across broader taxonomic scales (rather than just within the *P. mexicana* species complex), because other species of the family Poeciliidae have also evolved high tolerance to H_2S and coexist with *P. mexicana* in some springs. Overall, our findings illustrate natural selection's ability to bring similar alleles to high frequency in disjunct populations with distinct genetic backgrounds. Evidence for the roles of introgression and recruitment of closely related alleles from standing genetic variation during adaptation is mounting [18,69,79], involving examples for gene flow across larger geographical distances [80] and among more divergent lineages [81] than investigated by our study. Moreover, it was recently shown that particular DNA motifs can change DNA structure and increase local mutation probabilities which could lead to repeated de novo mutations [82].

Future research will have to eluciate the relative roles different mechanisms play in shaping convergence at a genomic level. Adaptive introgression and selection on standing genetic variation may be more common than previously appreciated, but convergent evolution in these cases is limited to the repeated assembly of related alleles into different genomic backgrounds. The epitome of convergent evolution arguably is the indepenent origin of adaptive mutations at the same locus (perhaps even the same codon), ultimately leading to consistent functional outcomes [21,83]. Identifying the circumstances under which adaptive alleles at the same loci originate independently and rise to high frequency in different populations is perhaps among the most pressing tasks in our endeavour to determine to what degree genomic evolution is repeatable and predictable.

Ethics. Appropriate ethical approval and licences were obtained for sampling. Permits were kindly provided by the Mexican Federal Agencies SEMARNAT and CONAPESCA (DGOPA.09004.041111.3088, SGPA/DGVS/04315/11, PRMN/DGOPA-003/2014, PRMN/DGOPA-009/2015).

Data accessibility. The raw whole-genome sequencing reads are available at NCBI short-read archive BioProject no. PRJNA473350. The scripts for the analyses can be found at https://github.com/jokelley/Pmex-GenomicConvergence.

Authors' contributions. A.P.B., M.T. and J.L.K conceived this study. A.P.B., R.G., M.T., L.A.R. and J.L.K conducted fieldwork. A.P.B., K.L.M., E.J.S. and J.L.K. performed experiments and contributed to data analysis. A.P.B., R.G., M.T. and J.L.K. wrote the manuscript. All authors have approved the final version of the manuscript.

Competing interests. We have no competing interests.

Funding. A.P.B. was supported in part by the Abelson Fellowship, the Guy Brislawn Scholarship Award, the James R. King Graduate Fellowship, and the Carl H. Elling Endowment from the School of Biological Sciences at Washington State University. R.G. was supported in part by a National Science Foundation Graduate Research Fellowship. This work was supported by grants from the National Science Foundation (grant nos. IOS-1121832, IOS-1463720, IOS-1557795 and IOS-1557860) and US Army Research Office (grant nos. W911NF-15-1-0175, W911NF-16-1-0225) to M.T. and J.L.K.

Acknowledgements. We would like to thank Nick Barts, Julian Bennett-Ponsford, Zach Culumber, Garrett Hopper, Courtney Passow and Diana Rennison for assistance collecting field samples and the Centro de Investigación e Innovación para la Enseñanza y Aprendizaje (CIIEA) for providing support in the field.

References

1. Endler JA. 1986 *Natural selection in the wild*. Princeton, NJ: Princeton University Press.
2. Schluter D. 2000 *The ecology of adaptive radiation*. Oxford, UK: Oxford University Press.
3. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014 Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. *Proc. Natl Acad. Sci. USA* **111**, E4736–E4742. (doi:10.1073/pnas.1416574111)
4. Furness AI, Reznick DN, Springer MS, Meredith RW. 2015 Convergent evolution of alternative developmental trajectories associated with diapause in African and South American killifish. *Proc. R. Soc. B* **282**, 20142189. (doi:10.1098/rspb.2014.2189)
5. Mahler DL, Ingram T, Revell LJ, Losos JB. 2013 Exceptional convergence on the macroevolutionary landscape in island lizard radiations. *Science* **341**, 292–295. (doi:10.1126/science.1232392)
6. Kowalko JE *et al.* 2013 Loss of schooling behavior in cavefish through sight-dependent and sight-independent mechanisms. *Curr. Biol.* **23**, 1874–1883. (doi:10.1016/j.cub.2013.07.056)
7. Fischer E, Nowicki J, O'Connell L. 2019 Evolution of affiliation: patterns of convergence from genomes to behaviour.
8. Elmer KR, Meyer A. 2011 Adaptation in the age of genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**, 298–306.
9. Rosenblum EB, Parent CE, Brandt EE. 2014 The molecular basis of convergent evolution. *Annu. Rev. Ecol. Evol. Syst.* **45**, 203–226.
10. Gould BA, Stinchcombe JR. 2017 Population genomic scans suggest novel genes underlie convergent flowering time evolution in the introduced range of *Arabidopsis thaliana*. *Mol. Ecol.* **26**, 92–106. (doi:10.1111/mec.13643)
11. Mandic M, Ramon ML, Gerstein AC, Gracey AY, Richards JG. 2018 Variable gene transcription underlies phenotypic convergence of hypoxia tolerance in sculpins. *BMC Evol. Biol.* **18**, 163. (doi:10.1186/s12862-018-1275-1)

12. McGee MD, Neches RY, Seehausen O. 2016 Evaluating genomic divergence and parallelism in replicate ecomorphs from young and old cichlid adaptive radiations. *Mol. Ecol.* **25**, 260–268. (doi:10.1111/mec.13463)
13. Feldman CR *et al*. 2016 Is there more than one way to skin a newt? Convergent toxin resistance in snakes is not due to a common genetic mechanism. *Heredity (Edinb.)* **116**, 84–91. (doi:10.1038/hdy.2015.73)
14. Krakauer DC, Plotkin JB. 2002 Redundancy, antiredundancy, and the robustness of genomes. *Proc. Natl Acad. Sci. USA* **99**, 1405–1409. (doi:10.1073/pnas.032668599)
15. Tenaillon O, Rodriguez-Verdugo A, Gaut RL, McDonald P, Bennett AF, Long AD, Gaut BS. 2012 The molecular diversity of adaptive convergence. *Science* **335**, 457–461. (doi:10.1126/science.1212986)
16. Griffin PC, Hangartner SB, Fournier-Level A, Hoffmann AA. 2017 Genomic trajectories to desiccation resistance: convergence and divergence among replicate selected *Drosophila* lines. *Genetics* **205**, 871–890. (doi:10.1534/genetics.116.187104)
17. Colosimo PF *et al.* 2005 Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science* **307**, 1928–1933.
18. Jones FC *et al.* 2012 The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* **484**, 55–61.
19. Grant PR, Grant BR, Markert JA, Keller LF, Petren K. 2004 Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. *Evolution* **58**, 1588–1599.
20. Richards EJ, Martin CH. 2017 Adaptive introgression from distant Caribbean islands contributed to the diversification of a microendemic adaptive radiation of trophic specialists pupfishes. *PLoS Genet.* **13**, e1006919. (doi:10.1371/journal.pgen.1006919)
21. Feldman CR, Brodie ED, Brodie ED, Pfrender ME. 2012 Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc. Natl Acad. Sci. US A* **109**, 4556–4561. (doi:10.1073/pnas.1113468109)
22. Palacios M, Arias-Rodriguez L, Plath M, Eifert C, Lerp H, Lamboj A, Voelker G, Tobler M. 2013 The rediscovery of a long described species reveals additional complexity in speciation patterns of poeciliid fishes in sulfide springs. *PLoS ONE* **8**, e71069. (doi:10.1371/journal.pone.0071069)
23. Cooper CE, Brown GC. 2008 The inhibition of mitochondrial cytochrome oxidase by the gases carbon monoxide, nitric oxide, hydrogen cyanide and hydrogen sulfide: chemical mechanism and physiological significance. *J. Bioenerg. Biomembr.* **40**, 533–539. (doi:10.1007/s10863-008-9166-6)
24. Bagarinao T. 1992 Sulfide as an environmental factor and toxicant: tolerance and adaptations in aquatic organisms. *Aquat. Toxicol.* **24**, 21–62.
25. 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)
26. Tobler M *et al.* 2011 Evolution in extreme environments: replicated phenotypic differentiation in livebearing fish inhabiting sulfidic springs. *Evolution* **65**, 2213–2228.
27. 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)
28. 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)
29. Pfenninger M, Patel S, Arias-Rodriguez L, Feldmeyer B, Riesch R, Plath M. 2015 Unique evolutionary trajectories in repeated adaptation to hydrogen sulphide-toxic habitats of a neotropical fish (*Poecilia mexicana*). *Mol. Ecol.* **24**, 5446–5459. (doi:10.1111/mec.13397)
30. Alda F, Reina RG, Doadrio I, Bermingham E. 2013 Phylogeny and biogeography of the *Poecilia sphenops* species complex (Actinopterygii. Poeciliidae) in Central America. *Mol. Phylogenet. Evol.* **66**, 1011–1026.
31. Palacios M, Voelker G, Arias-Rodriguez L, 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. Phylogenet. Evol.* **103**, 230–244.
32. Schluter D, Conte GL. 2009 Genetics and ecological speciation. *Proc. Natl Acad. Sci. USA* **106**, 9955–9962. (doi:10.1073/pnas.0901264106)
33. Krueger F. 2014 Trim Galore! version 0.3.7. See https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/.
34. 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)
35. Li H, Durbin R. 2010 Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589–595. (doi:10.1093/bioinformatics/btp698)
36. McKenna A *et al.* 2010 The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303. (doi:10.1101/gr.107524.110)
37. DePristo MA *et al.* 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498. (doi:10.1038/ng.806)
38. Van der Auwera GA *et al.* 2013 From FastQ data to high confidence variant calls: the genome analysis toolkit best practices pipeline. *Curr. Protoc. Bioinform.* **43**, 11.10.11–33. (doi:10.1002/0471250953.bi1110s43)
39. Danecek P *et al.* 2011 The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158. (doi:10.1093/bioinformatics/btr330)
40. Lee TH, Guo H, Wang X, Kim C, Paterson AH. 2014 SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. *BMC Genomics* **15**, 162. (doi:10.1186/1471-2164-15-162)
41. Felsenstein JP. 2004 PHYLIP (phylogeny inference package), Version 3.6. See http://evolution.genetics.washington.edu/phylip.html.
42. Kunstner A, Hoffmann M, Fraser BA, Kottler VA, Sharma E, Weigel D, Dreyer C. 2016 The genome of the Trinidadian guppy, *Poecilia reticulata*, and variation in the Guanapo population. *PLoS ONE* **11**, e0169087. (doi:10.1371/journal.pone.0169087)
43. Huson DH, Richter DC, Rausch C, Dezulian T, Franz M, Rupp R. 2007 Dendroscope: an interactive viewer for large phylogenetic trees. *BMC Bioinf.* **8**, 460. (doi:10.1186/1471-2105-8-460)
44. Zamani N *et al.* 2013 Unsupervised genome-wide recognition of local relationship patterns. *BMC Genomics* **14**, 347. (doi:10.1186/1471-2164-14-347)
45. Wong K, Keane TM, Stalker J, Adams DJ. 2010 Enhanced structural variant and breakpoint detection using SVMerge by integration of multiple detection methods and local assembly. *Genome Biol.* **11**, R128. (doi:10.1186/gb-2010-11-12-r128)
46. Chen K *et al.* 2009 BreakDancer: an algorithm for high resolution mapping of genomic structural variation. *Nat. Methods* **6**, 677–681. (doi:10.1038/nmeth.1363)
47. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009 Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871. (doi:10.1093/bioinformatics/btp394)
48. Zerbino DR, Birney E. 2008 Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* **18**, 821–829. (doi:10.1101/gr.074492.107)
49. Quinlan AR, Hall IM. 2010 BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics (Oxford, England)* **26**, 841–842. (doi:10.1093/bioinformatics/btq033)
50. Aszódi A. 2012 MULTOVL: fast multiple overlaps of genomic regions. *Bioinformatics* **28**, 3318–3319. (doi:10.1093/bioinformatics/bts607)
51. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. 2009 GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. *BMC Bioinf.* **10**, 48–48. (doi:10.1186/1471-2105-10-48)
52. Passow CN *et al.* 2017 Complexities of gene expression patterns in natural populations of an extremophile fish (*Poecilia mexicana*, Poeciliidae). *Mol. Ecol.* **26**, 4211–4225. (doi:10.1111/mec.14198)
53. Kelley JL, Arias-Rodriguez L, Patacsil 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)
54. Pfenninger M *et al.* 2014 Parallel evolution of COX genes in H2S-tolerant fish as key adaptation to a toxic environment. *Nat. Commun.* **5**, 3873. (doi:10.1038/ncomms4873)
55. Barts N, Greenway R, Passow CN, Arias-Rodriguez L, Kelley JL, Tobler M. 2017 Molecular evolution and

expression of oxygen transport genes in livebearing fishes (Poeciliidae) from hydrogen sulfide rich springs. *Genome* **61**, 273–286. (doi:10.1139/gen-2017-0051)

56. Hildebrandt TM, Grieshaber MK. 2008 Three enzymatic activities catalyze the oxidation of sulfide to thiosulfate in mammalian and invertebrate mitochondria. *FEBS J.* **275**, 3352–3361. (doi:10.1111/j.1742-4658.2008.06482.x)
57. 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. *Biochim. Biophys. Acta* **1797**, 4. (doi:10.1016/j.bbabio.2010.04.004)
58. Petersen LC. 1977 The effect of inhibitors on the oxygen kinetics of cytochrome *c* oxidase. *Biochim. Biophys. Acta* **460**, 299–307.
59. Hayashi T *et al.* 2015 Higd1a is a positive regulator of cytochrome *c* oxidase. *Proc. Natl Acad. Sci. USA* **112**, 1553–1558. (doi:10.1073/pnas.1419767112)
60. Shibuya N, Nagahara N, Mikami Y, Kimura Y, Kimura H. 2009 Vascular endothelium expresses 3-mercaptopyruvate sulfurtransferase and produces hydrogen sulfide. *J. Biochem.* **146**, 623–626. (doi:10.1093/jb/mvp111)
61. Tang G, Wu L, Wang R. 2010 Interaction of hydrogen sulfide with ion channels. *Clin. Exp. Pharmacol. Physiol.* **37**, 753–763. (doi:10.1111/j.1440-1681.2010.05351.x)
62. Lu SC. 2009 Regulation of glutathione synthesis. *Mol. Aspects Med.* **30**, 42–59. (doi:10.1016/j.mam.2008.05.005)
63. Passow C, Arias-Rodriguez L, Tobler M. 2017 Convergent evolution of reduced energy demands in extremophile fish. *PLoS ONE* **12**, e0186935. (doi:10.1371/journal.pone.0186935)
64. Tobler M, Riesch RM, Tobler C, Plath M. 2009 Compensatory behaviour in response to sulphide-induced hypoxia affects time budgets, feeding efficiency, and predation risk. *Evol. Ecol. Res.* **11**, 935–948.
65. Riesch R, Plath M, Schlupp I, Tobler M, Brian Langerhans R. 2014 Colonisation of toxic environments drives predictable life-history evolution in livebearing fishes (Poeciliidae). *Ecol. Lett.* **17**, 65–71. (doi:10.1111/ele.12209)
66. Brown AP, Arias-Rodriguez L, Yee M.-C., Tobler M, Kelley JL. 2018 Concordant changes in gene expression and nucleotides underlie independent adaptation to hydrogen-sulfide-rich environments. *Genome Biol. Evol.* **10**, 2867–2881. (doi:10.1093/gbe/evy198)
67. Jones MR *et al.* 2018 Adaptive introgression underlies polymorphic seasonal camouflage in snowshoe hares. *Science* **360**, 1355. (doi:10.1126/science.aar5273)
68. Sen N, Paul BD, Gadalla MM, Mustafa AK, Sen T, Xu R, Kim S, Snyder SH. 2012 Hydrogen sulfide-linked sulfhydration of NF-κB mediates its antiapoptotic actions. *Mol. Cell* **45**, 13–24. (doi:10.1016/j.molcel.2011.10.021)
69. 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)
70. Lee KM, Coop G. 2019 Population genomics perspectives on convergent adaptation. *Phil. Trans. R. Soc. B* **374**, 20180236. (doi:10.1098/rstb.2018.0236)
71. Miller RR, Minckley W, Norris S. 2005 *Freshwater fishes of Mexico*. Chicago, IL: University of Chicago Press.
72. Affonso EG, Polez V, Corrêa CF, Mazon A, Araújo MR.R., Moraes G, Rantin F. 2005 Physiological responses to sulfide toxicity by the air-breathing catfish, *Hoplosternum littorale* (Siluriformes, Callichthyidae). *Comp Biochem. Physiol. C: Tox. Pharm.* **139**, 251–257. (doi:10.1016/j.cca.2004.11.007)
73. Brown AP, Greenway R, Morgan S, Quackenbush CR, Giordani L, Arias-Rodriguez L, Tobler M, Kelley JL. 2017 Genome-scale data reveal that endemic *Poecilia* populations from small sulphidic springs display no evidence of inbreeding. *Mol. Ecol.* **26**, 4920–4934. (doi:10.1111/mec.14249)
74. Miller RR. 1966 Geographical distribution of Central American freshwater fishes. *Copeia* **1966**, 773–802. (doi:10.2307/1441406)
75. Lee KM, Coop G. 2017 Distinguishing among modes of convergent adaptation using population genomic data. *Genetics* **207**, 1591–1619. (doi:10.1534/genetics.117.300417)
76. Lehman N, Eisenhawer A, Hansen K, Mech LD, Peterson RO, Peter JP.G, Wayne RK. 1991 Introgression of coyote mitochondrial DNA into sympatric North American gray wolf populations. *Evolution* **45**, 104–119. (doi:10.2307/2409486)
77. McGuire JA, Charles WL, Michelle SK, Hutchison DW, Lappin AK, Orange DI, Lemos-Espinal J, Riddle BR, Jaeger JR. 2007 Mitochondrial introgression and incomplete lineage sorting through space and time: phylogenetics of crotaphytid lizards. *Evolution* **61**, 2879–2897.
78. Dias C, Lima KDA, Araripe J, Aleixo A, Vallinoto M, Sampaio I, Schneider H, Rêgo PSD. 2018 Mitochondrial introgression obscures phylogenetic relationships among manakins of the genus Lepidothrix (Aves: Pipridae). *Mol. Phylogenet. Evol.* **126**, 314–320. (doi:10.1016/j.ympev.2018.04.017)
79. Meier JI, Marques DA, Mwaiko S, Wagner CE, Excoffier L, Seehausen O. 2017 Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nat. Commun.* **8**, 14363. (doi:10.1038/ncomms14363)
80. Martin CH, Crawford JE, Turner BJ, Simons LH. 2016 Diabolical survival in Death Valley: recent pupfish colonization, gene flow and genetic assimilation in the smallest species range on earth. *Proc. R. Soc. B* **283**, 20152334. (doi:10.1098/rspb.2015.2334)
81. Martin SH *et al.* 2013 Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Res.* **23**, 1817–1828. (doi:10.1101/gr.159426.113)
82. Xie KT *et al.* 2019 DNA fragility in the parallel evolution of pelvic reduction in stickleback fish. *Science* **363**, 81–84. (doi:10.1126/science.aan1425)
83. 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)