Genomic Analysis of the Only Blind Cichlid Reveals Extensive Inactivation in Eye and Pigment Formation Genes

Matthew L. Aardema[1,2], Melanie L.J. Stiassny[3], and S. Elizabeth Alter[3,4,5,6,*]

[1]Department of Biology, Montclair State University

[2]Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York

[3]Department of Ichthyology, American Museum of Natural History, New York, New York

[4]The Graduate Center, City University of New York

[5]Department of Biology, York College/The City University of New York

[6]Present address: Biology & Chemistry Department, California State University, Monterey Bay, Chapman Academic Science Center, Building 53, 100 Campus Center, Seaside, CA

*Corresponding author: E-mail: ealter@csumb.edu.

Accepted: 6 July 2020

Data deposition: This project has been deposited at the NCBI Short Read Archive under the accession BioProject PRJNA577474. BAM files with the mapped, sorted, and realigned reads have been deposited at the NCBI Short Read Archive under accessions SRR10811633 and SRR10811634. Alignments and chromatograms from PCR/Sanger sequencing confirmation are available in Dryad (https://doi.org/10.5061/dryad.3ffbg79db)..

Abstract

Trait loss represents an intriguing evolutionary problem, particularly when it occurs across independent lineages. Fishes in light-poor environments often evolve "troglomorphic" traits, including reduction or loss of both pigment and eyes. Here, we investigate the genomic basis of trait loss in a blind and depigmented African cichlid, *Lamprologus lethops*, and explore evolutionary forces (selection and drift) that may have contributed to these losses. This species, the only known blind cichlid, is endemic to the lower Congo River. Available evidence suggests that it inhabits deep, low-light habitats. Using genome sequencing, we show that genes related to eye formation and pigmentation, as well as other traits associated with troglomorphism, accumulated inactivating mutations rapidly after speciation. A number of the genes affected in *L. lethops* are also implicated in troglomorphic phenotypes in Mexican cavefish (*Astyanax mexicanus*) and other species. Analysis of heterozygosity patterns across the genome indicates that *L. lethops* underwent a significant population bottleneck roughly 1 Ma, after which effective population sizes remained low. Branch-length tests on a subset of genes with inactivating mutations show little evidence of directional selection; however, low overall heterozygosity may reduce statistical power to detect such signals. Overall, genome-wide patterns suggest that accelerated genetic drift from a severe bottleneck, perhaps aided by directional selection for the loss of physiologically expensive traits, caused inactivating mutations to fix rapidly in this species.

Key words: convergent evolution, cave fish, depigmentation, adaptation, vision genes, trait loss, troglomorphic.

Introduction

The independent and repeated loss of phenotypic traits across lineages offers the opportunity to study the modes and predictability of genomic evolution under particular selective regimes. Fishes living in lightless or low-light subterranean habitats number over 165 teleostean species (Niemiller and Soares 2015) and represent one of the most well-examined cases of such repeated trait loss. Many have evolved a distinctive set of morphological features, sometimes called "troglomorphic syndrome," that are generally considered to be adaptations to a lightless or light-poor, and limited resource environment. These features include loss of pigmentation and reduction or loss of eyes, as well as enhancements in other aspects of sensory anatomy. Cave fishes such as the Mexican cave tetra, *Astyanax mexicanus* (Teleostei, Characiformes) in particular, have emerged as model

organisms for studying these traits (e.g., Jeffery 2005; Protas et al. 2007, 2008). However, the loss of eyes, pigments, and other features is not limited to organisms that live in caves but is found across a variety of ecosystems, including deep sea, fossorial, parasitic, and some riverine habitats. In particular, this phenomenon has been identified in a diverse set of fish species that are endemic to the lower Congo River (Alter et al. 2015; Stiassny and Alter, forthcoming), including the cichlid *Lamprologus lethops* (Teleostei, Ovalentaria). These striking examples of phenotypic convergence across deep phylogenetic distances provide an excellent opportunity to investigate the extent to which genomic mechanisms generating these phenotypes are the same or similar across species, or entirely phylogenetically contingent.

Discovered and described in the 1970s (Roberts and Stewart 1976), *L. lethops* is the only known blind and depigmented species in the otherwise colorful and highly diverse family Cichlidae. Information about its habitat and ecology are extremely limited, and relatively few specimens have ever been recovered. However, available evidence suggests that populations of *L. lethops* live in highly turbulent waters at extreme depths in lower Congo River canyons (Stiassny and Alter, forthcoming). The hydrology of the lower Congo River is extraordinarily complex, and the river is characterized by some of the world's largest rapids, high-velocity in-stream flows, and a complex bathymetry with shallow rapids located adjacent to canyons with depths >220 m recorded. Notably, all *L. lethops* samples have been recovered dead or moribund at the surface from a short stretch of the lower Congo in the regions of Bulu and Luozi (fig. 1*a*). Among *L. lethops* specimens recovered, some have been found with accumulations of subcutaneous gas bubbles, and most have fenestrated gas bladders, indicating catastrophic decompression following rapid ascent from depth (Stiassny and Alter, forthcoming). A related congener, *L. tigripictilis*, and other species of the lower Congo are also found in these locations, but with none of these same traumas.

Phenotypic resemblance to cave fishes in this species extends to numerous anatomical attributes (figs. 1*a*, 2*a*, and *b* and supplementary figs. 1*a* and 2, Supplementary Material online), with the most obvious being a lack of any visible pigmentation in all specimens examined to date, and lack of image-forming eyes (cryptophthalmia). The evidence for the former includes absence of any observable melanin (supplementary fig. 2, Supplementary Material online) and evidence for the latter includes greatly foreshortened optic globes, a decreased number of neuronal layers in the retina, and the absence of extraocular muscles and choroid rete mirabile, which oxygenates the retina (fig. 2*b* and supplementary fig. 1*a*, Supplementary Material online; Schobert et al. 2012). Detailed anatomical analysis of the degeneration in the eye by Schobert et al. (2012) indicates that *L. lethops* is similar to the Mexican cavefish, *A. mexicanus*, in displaying some conservation of the architecture of the neuronal retina (albeit embedded deep in the tissues of the head); however, eyes in *L. lethops* appear to have an intact lens whereas the lens in *A. mexicanus* degenerates during development (e.g., Langecker et al. 1993). *Lamprologus lethops* also shows a number of other characters commonly found in cave-dwelling *A. mexicanus*, such as enhanced neurocranial lateral-osensory canals and pores (supplementary fig. 1*c*, Supplementary Material online), and enhanced fat deposits (fig. 2*b*). In addition, and in contrast with other cichlids which typically have a thin-walled gas bladder, *L. lethops* has an enlarged gas bladder with a highly thickened outer layer, or tunica externa (supplementary fig. 1*e* and *g*, Supplementary Material online). Such a heavily reinforced gas bladder supports the hypothesis that this species occupies a turbulent deep-water habitat, as it would allow an increase in the range of depths over which an individual could resist positive buoyancy if entrained in upwelling currents.

Determining the genetic underpinnings of this cryptophthalmic or troglomorphic phenotype can illuminate how evolution proceeds under strong environmental selection and can provide a basis for future comparative studies of similar phenotypes across a broad phylogenetic spectrum of other cryptophthalmic teleosts in the lower Congo (Alter et al. 2015). However, the lack of genome-level data for *L. lethops* has hampered efforts to better understand the mode and timing of evolution in this unique lineage (Stiassny and Alter, forthcoming). The evolutionary origins of the species remain opaque, and although its phylogenetic placement among the 90+ described lamprologine species is not completely certain, morphological and biogeographic evidence indicates that it is most closely related to the other lamprologine species endemic to the lower Congo, including *L. tigripictilis*, *L. markerti*, *L. werneri*, and *L. teugelsi*. Data from mitochondrial and traditional nuclear markers have proven insufficient to adequately resolve phylogenetic relationships within the lower Congo *Lamprologus* clade (unpublished data), suggesting these lineages likely radiated rapidly and recently. This observation, in addition to the young age of the current high-energy hydrological regime of the lower Congo system (estimated at 2–5 Myr; reviewed by Stiassny and Alter [forthcoming]), suggests that *L. lethops* split from its congeners relatively recently, making it an interesting case study in rapid phenotypic and genomic divergence.

In addition to the timing of evolution, genomic data can also inform our understanding of the evolutionary mechanisms underlying trait loss. One hypothesis posits that when the maintenance of a trait is no longer under purifying selection, deleterious mutations (including mutations that inactivate trait-associated genes) will accumulate through neutral processes (i.e., drift). An alternative (though not mutually exclusive) hypothesis is that inactivating mutations can be adaptive when a trait is no longer beneficial, especially if the trait is energetically costly to build or maintain. In such cases, inactivating mutations will be positively selected for and fix rapidly.

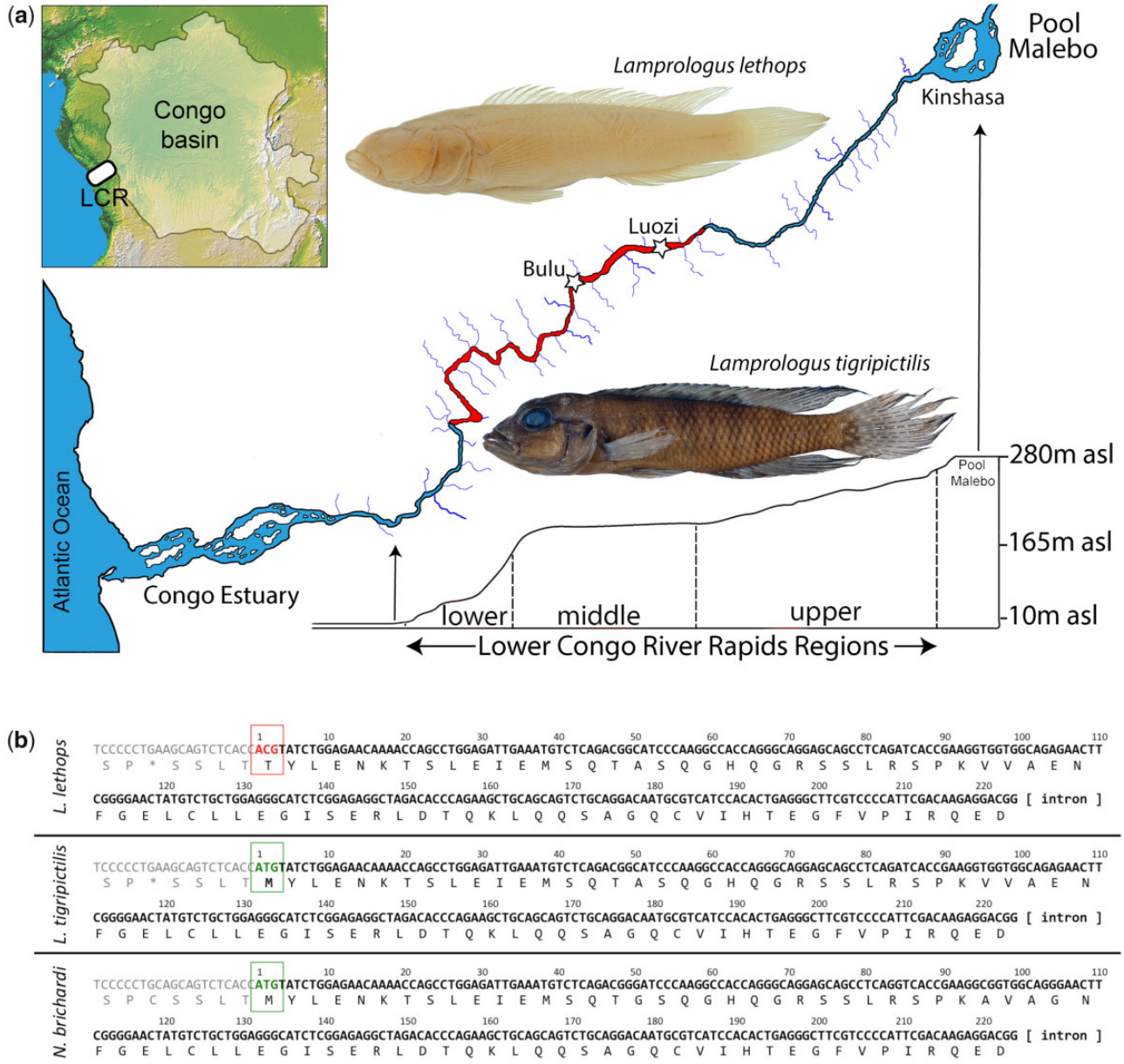

Fig. 1—Distribution of *Lamprologus lethops* and its loss of pigmentation. (*a*) Map of the lower Congo River with geographical location indicator inset on left. Known distribution of *L. lethops* (white stars) and *Lamprologus tigripictilis* (in red). Altitudinal profile of the river inset below. The two fish images show the complete loss of pigmentation in *L. lethops* (top fish image), compared with the closely related *L. tigripictilis* (bottom fish image). (*b*) Alignment of the nucleotides and amino acids for the first exon of the Oca2 gene in addition to 21 upstream nucleotides (in gray text). The single-nucleotide change in the start codon of the gene in *L. lethops* from a thymine to a cytosine is indicated (red box). Numbers indicate position in the gene.

Here, we investigate the genomic basis of the cryptophthalmic phenotype in *L. lethops* by 1) examining genes with variants of likely large effect relative to both sympatric and allopatric lamprologines and 2) identifying and assessing candidate genes for several traits of interest (including those related to vision, pigment, and circadian rhythm). We present both draft de novo and reference-guided genome assemblies for the focal species, *L. lethops*, as well as that of a related, sympatric congener with a noncryptophthalmic phenotype (*L. tigripictilis*). Examining genes with inactivating mutations specific to the *L. lethops* genome, we find that many are known to influence eye and pigment formation, circadian rhythms, metabolism, and UV damage repair. Some genetic variants, including those in UV damage repair genes and metabolic genes, provide strong clues regarding the ecology of this difficult-to-study species. These observations yield further insights into the molecular mechanisms involved in evolution in extreme environments and add to our understanding of the

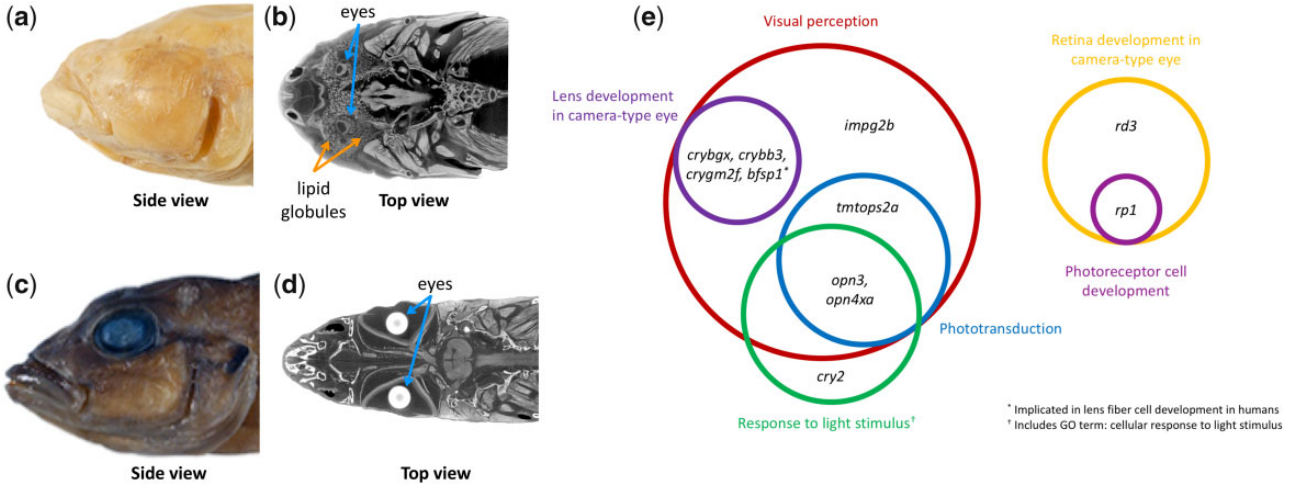

Fig. 2—Eye morphology and genes affecting vision. (*a*) External morphology and (*b*) transverse section through cranium and eyes (Computerized tomography scan after incubation in phosphotungstic acid) of *Lamprologus lethops* (the locations of the degenerate eyes and high concentrations of fat globules are indicated by the arrows), and (*c*) external morphology and (*d*) transverse section through cranium and eyes (Computerized tomography scan after incubation in phosphotungstic acid) of *Lamprologus tigripictilis* (the locations of the well-formed eyes are indicated by the arrows). (*e*) Venn diagrams summarizing likely genes found to contain inactivating mutations that may contribute to degeneration of the eye. Each circle represents a specific GO term(s). Genes listed in the overlapping areas indicate more than one GO eye-related term is associated with that gene. Gene names are given in table 1.

genetic underpinnings of both constructive and regressive traits in cryptophthalmic, hypogean, and troglomorphic species.

Results and Discussion

Inactivating Mutations in *L. lethops*

Following previous work in other systems characterized by degenerative phenotypes (e.g., Leys et al. 2005; Cai and Patel 2010; Wessinger and Rausher 2015), including *Astyanax* cavefish (Gross and Wilkens 2013), we hypothesized that the loss of pigmentation and substantial reduction in the eyes of *L. lethops* were likely facilitated in part by gene-inactivating mutations (also called "loss-of-function" mutations). Here, such mutations include premature stop codons, loss of the start codon, frameshift mutations, and changes to exon/intron splice sites. Such mutations are generally presumed to result in a gene with reduced or no function. To look for the presence of such mutations, we generated Illumina paired-end data for *L. lethops* and mapped it to the genome of *Neolamprologus brichardi*, the evolutionarily closest available annotated reference (Brawand et al. 2014). For comparison, we also generated similar data for *L. tigripictilis*. We annotated *lethops*-specific genetic variants for their predicted impact on known cichlid genes using functional characterization based on the zebrafish, *Danio rerio* (supplementary table 1, Supplementary Material online), and determined specific variants that likely resulted in gene inactivation. Characterization of inactivating mutations unique to *L. lethops* with such methods naturally depends on the accuracy of the reference genome assembly and annotation, as well as functional characterization in *D. rerio*. To further assess the accuracy of potential inactivating variants annotated in *L. lethops*, we also compared the focal exons containing these variants with orthologous sequences in six other cichlid species with annotated genomes available (not including *N. brichardi)*, as well as five other fish species in the subseries Ovalentaria (following Hughes et al. 2018).

Our analysis of likely inactivating mutations in genes potentially contributing to the cryptophthalmic phenotype of *L. lethops* resulted in 15 genes of interest. Of these, the majority (11) related to eye formation (fig. 2e, table 1, and supplementary table 2, Supplementary Material online), with additional genes implicated in pigmentation, circadian rhythm, metabolism, and UV damage repair. Frameshift mutations were the most common route for inactivation among these genes (table 1 and supplementary table 2, Supplementary Material online), followed by premature stop codons. Two genes that influence eye formation (cryptochrome circadian regulator 2 [cry2] and beaded filament structural protein 1 [bfsp1]) had both a frameshift mutation and a premature stop codon. In one eye formation gene (interphotoreceptor matrix proteoglycan 2b [impg2b]), the inactivating variant was heterozygous (segregating).

Convergent evolution of phenotypes can arise from the same mutations, different mutations in the same gene, mutations across different gene functional groups, or may reflect completely independent processes (Manceau et al. 2010; Pankey et al. 2014). We find that some of the same inactivated genes in *L. lethops* are implicated in trait evolution in other troglomorphic fishes including the characid *A. mexicanus* and the cyprinids *Sinocyclocheilus* spp. and

Table 1
Genes Found to Contain Inactivating Mutations

Predicted Phenotypic Effect	Gene Name (in *Danio rerio*)	Gene Symbol	Relevant GO Terms	Implication in Species with Similar Phenotypes
Loss of pigmentation	Promelanin-concentrating hormone[a]	pmch	Melanin-concentrating hormone activity	None identified
	Oculocutaneous albinism II	oca2	Melanin biosynthetic process, melanocyte differentiation, pigmentation	Am: deletion (Protas et al. 2006, 2007; Klaassen et al. 2018)
Loss of vision, eye reduction (**Circadian Rhythms**)	Melanopsin-B-like (opn4xa)	opn4xa	Cellular response to light stimulus, phototransduction, visual perception	None identified
	Retinal degeneration 3, GUCY2D regulator	rd3	Retina development in camera-type eye[a]	None (found to be expressed at similar levels in Am and Sa)
	Retinitis pigmentosa 1	rp1	Photoreceptor cell development, retina development in camera-type eye	Nonfunctional in subterranean mammals (Emerling and Springer 2014)
	Cryptochrome circadian regulator 2	cry2	Response to light stimulus	None identified
	Beaded filament structural protein 1 (filensin)	bfsp1	Lens fiber cell development[b]	Upregulated in Sa (Meng et al. 2013)
	Crystallin beta gamma X	crybgx	Lens development in camera-type eye, visual perception	Downregulated in Am (Hinaux et al. 2013)
	Crystallin, beta B3	crybb3	Lens development in camera-type eye, visual perception	Premature stop codon in naked mole rat (Kim et al. 2011)
	Crystallin, gamma M2f	crygm2f	Lens development in camera-type eye, visual perception	Sa: upregulation (Meng et al. 2013)
	Opsin-3	opn3	Cellular response to light stimulus, phototransduction, response to stimulus, visual perception	None identified
	Teleost multiple tissue opsin 2a	tmtops2a	Phototransduction, response to stimulus, visual perception	Pa: premature stop codon (Cavallari et al. 2011)
	Interphotoreceptor matrix proteoglycan 2b	impg2b	Visual perception	Am: downregulated (Stahl and Gross 2017); nonfunctional in subterranean mammals (Emerling and Springer 2014)
Metabolism	Spexin	spx	Negative regulation of appetite	None identified
UV damage repair	Damage-specific DNA binding protein 2	ddb2	Cellular response to DNA damage stimulus, DNA repair, response to UV	Am: constitutively expressed (Foulkes et al. 2016)

Note.—Gene names and symbols come from the genome annotation of *Danio rerio*. The relevant GO terms are also indicated (from the ZFIN database; Howe et al. 2012). Finally, if the gene has been implicated in similar phenotypic changes in other vertebrate species (predominately the cavefish, *Astyanax mexicanus*), these references are given. Bolded genes are also also associated with circadian rhythms. Sa = *Sinocyclocheilus anophthalmus*; Am = *Astyanax mexicanus*; Pa = *Phreatichthys andruzzii*.

[a]Also implicated in appetite control.

[b]In humans.

Phreatichthys andruzzii, though the specific mutations differ (table 1 and supplementary table 2, Supplementary Material online).

Other genes that show loss of function in *L. lethops* appear fully intact in *A. mexicanus*, but previous work shows the expression of those genes are modified in the latter species.

In still other cases, we found that although the specific genes differed, the same broad gene families (e.g., crystallins, opsins) were affected. Given the large phylogenetic distance between these species (ca., 230–280 Myr; Near et al. 2012), convergence at the gene/gene family levels suggests that the mechanisms of evolution underlying troglomorphism and/or cryptophthalmia may be predictable at the gene level for some traits. When inactivated genes are shared across these distant clades (e.g., *oca2*), it is likely that such genes have low pleiotropic constraints and/or influence developmentally independent traits (e.g., Womack et al. 2018).

Loss of Pigmentation

Although >150 genes are associated with color variation in vertebrates (Hubbard et al. 2010), two genes have been primarily implicated in pigmentation loss or reduced melanism in many cave fishes: *oca2* and *mc1r* (Protas et al. 2006, 2007; Gross et al. 2009; Stahl and Gross 2015). Deletions in the *oca2* gene in *A. mexicanus* have been linked to the amelanistic cave fish phenotype (Protas et al. 2006, 2007) and have been shown via gene editing to result in amelanism in surface populations (Klaassen et al. 2018). Available evidence suggests that *oca2*, a melanosome-specific 12-transmembrane domain protein, plays a role in the maturation and transport of a key enzyme involved in melanin production, tyrosinase (Sitaram et al. 2009). Data from a number of vertebrate species, including humans, mice, dogs, reptiles, and other fish species, implicate mutations in *oca2* in the loss of pigment, ranging from a reduction in pigment (hypomelanism) to apparent complete loss of pigmentation (Brilliant et al. 1991; Rinchik et al. 1993; Fukamachi et al. 2004; Saenko et al. 2015; Caduff et al. 2017; Manga 2018; Kratochwil et al. 2019).

In the *oca2* gene of *L. lethops*, a single-nucleotide change at the second position of the first codon disrupts the start codon (fig. 1). We confirmed that this mutation is homozygous across five additional individuals of *L. lethops* using polymerase chain reaction (PCR) and Sanger sequencing (see supplementary Methods, Supplementary Material online). In addition, we searched for possible noncanonical, in-frame initiation sites upstream and downstream of the start codon; although methionine codons occur downstream, they are not embedded within the Kozak consensus sequence (Kozak 1984) and therefore unlikely to represent alternative start sites for translation. Disruption of the melanin pathway via loss of functional *oca2* transcripts is sufficient to explain the apparent lack of melanin pigmentation in *L. lethops*. We also observe the accumulation of additional nonsynonymous changes in the *oca2* gene of *L. lethops* and a relatively higher ratio of nonsynonymous to synonymous changes in *L. lethops* (0.0128) as compared with *L. tigripictilis* (0.0102) (excluding the start codon). This further suggests that selection is no longer acting to maintain the function of this gene. This finding adds to a growing number of examples, including a leucistic morph in another cichlid species, *Melanochromis auratus* (Kratochwil et al. 2019), in which *oca2* is implicated in pigment reduction or loss in fishes. Mechanisms of loss of function vary in *oca2*, from deletions of part or all of exons 21 and 24 in the terminal end of the protein in *A. mexicanus* (Protas et al. 2006), to loss of the second exon in amelanistic *Melanochromis auratus* (Kratochwil et al. 2019); strikingly, *L. lethops* appears to be the first case in which the start codon has been lost. Several hypotheses have been put forth to explain why *oca2* is such a frequent evolutionary target, including in humans. These include its large size, developmental independence, and the possibility that freeing the substrates of the *oca2* protein (e.g., L-tyrosine) could increase availability of the substrate for the catecholamine pathway involved in feeding (Bilandžija et al. 2013) and sleep (Bilandžija et al. 2018) (however, advantages of the latter could be lessened if downregulation of *oca2* is already occurring). In contrast, no differences in amino acid sequences of *mc1r* were observed between *L. lethops* and *L. tigripictilis*, indicating that variation in the amino acid sequence of *mc1r* (which causes reduced melanism in cavefish) does not affect pigmentation in *L. lethops*.

Vision, Eye Reduction, and Circadian Rhythms

As expected in a species lacking external, image-forming eyes, many genes related to photoreception and eye function appear inactivated or functionally altered (via frameshift mutations) in *L. lethops* (fig. 2e, table 1, and supplementary table 2, Supplementary Material online). These include retinal degeneration 3 (*rd3*), which encodes a protein that, when truncated, causes retinal degeneration in mice (Friedman et al. 2006). Frameshift mutations in *L. lethops* are also observed in a number of genes encoding for structural proteins of the lens, including a number of crystallins (*cry2*, *crygm3*, *crygm2f*, *crybg1a*, *crybgx*, and *crybb3*) and filensin. These observations are consistent with studies of other troglomorphic or hypogean animals including *A. mexicanus* and the naked mole rat (*Heterocephalus glaber*) in which some crystallin genes are underexpressed or show loss-of-function mutations (Kim et al. 2011; Hinaux et al. 2013). Although most of the crystallins genes differ from those identified in *L. lethops*, several are the same: *crygm3* is not expressed at all in *Astyanax* (Gross et al. 2013), *crybgx* is expressed at a very reduced level in early development (Hinaux et al. 2013), and *crybb3* has a premature stop codon in the naked mole rat (Kim et al. 2011).

Several opsin family genes, including both visual and nonvisual opsins, are also found to have splice site variants, frameshift mutations, or premature stop codons in *L. lethops* (*opn4xa* (= *opn4x-1*), *opn3*, and *TMT-opsinb*). *Opn4x* (melanopsin) is expressed in retinal ganglion cells where it mediates circadian photoentrainment and sleep (Matos-Cruz et al. 2011). In a comparative functional analysis, Cavallari et al.

(2011) found premature stop codons in another melanopsin paralog, *Opn4m2*, as well as *TMT-opsin*, in the Somalian cavefish, and provide evidence that these two genes are responsible for disruption of the circadian rhythm in that species. A subsequent study of zebrafish and the distantly related beloniform medaka (*Oryzias latipes*) indicates that *TMT-opsin* is highly conserved across teleostean fishes and confers light sensitivity to inter- and motor neurons (Fischer et al. 2013). A premature stop codon and a frameshift mutation in cryptochrome-2 (*cry2*) suggests possible impacts on circadian rhythm in *L. lethops*.

A number of photoreception genes are functionally affected or disabled in *L. lethops* (e.g., *Imp2/Impg2b, RP1L1*; we confirmed the latter via PCR/Sanger sequencing; see Materials and Methods). Mutations in the ortholog of the *Imp2/Impg2b* gene cause retinitis pigmentosa (Bandah-Rozenfeld et al. 2010) and vitelliform macular dystrophy (Meunier et al. 2014) in humans, whereas mutations in *RP1L1* cause occult macular dystrophy in humans and progressive photoreceptor degeneration in mice (Yamashita et al. 2009). In zebrafish, suppression of *RP1L1* along with another gene, *c2orf71*, causes reduction of eye size and loss of rhodopsin in photoreceptors (Liu et al. 2017).

Metabolic Genes Suggest Increased Appetite

A premature stop codon in the neuropeptide spexin (*spx*) in *L. lethops* (confirmed via PCR and Sanger sequencing, see Materials and Methods) may correspond with the extensive lipid deposits observed in our anatomical analysis as compared with *L. tigripictilis* (fig. 2*b* and *d*), and other cichlids. The encoded spexin protein, which was discovered originally using bioinformatics methods and later confirmed through protein purification and functional studies across vertebrates, functions in appetite suppression. Zebrafish *spx*−/− mutants created using TALENs showed hyperphagia (increased appetite), overexpression of orexigenic *AgRP*, and increased serum concentration of glucose, triacylglycerol, cholesterol, and high-density lipoprotein cholesterol (Zheng et al. 2017). Similarly, in goldfish, injection of *spx* suppressed appetite and feeding (Wong et al. 2013). Consistent with these observations in teleosts, studies have linked low spexin levels to obesity and insulin resistance in both rodents and humans (Walewski et al. 2014; Kumar et al. 2016; Kolodziejski et al. 2018). Interestingly, a different mechanism appears to underlie a similar phenotype in *A. mexicanus*, as coding mutations in the melanocortin 4 receptor were linked to hyperphagia in this species (Aspiras et al. 2015).

An additional inactivating mutation (a premature stop codon) is observed in promelanin-concentrating hormone (*pmch*), a neuropeptide implicated in melanin concentration (Kawauchi et al. 1983), energy regulation and appetite (Berman et al. 2009), and stress response (Green and Baker 1991), in diverse teleostean species including zebrafish, rainbow trout, flounder, eel, tilapia, and chum salmon (Kawauchi 2006). This suggests that the peptide encoded by *pmch* regulates food intake in many fishes (Matsuda et al. 2006; Berman et al. 2009; Matsuda 2009).

Inactivation of UV Repair Mechanisms

Repair of UV-damaged DNA is a critical function in most fishes, but in fishes living in a low-light or light-free environments, the need for such repair processes may be diminished. DNA repair mechanisms have been shown to be light dependent in most teleosts. The nucleotide excision repair pathway repairs UV-light-induced pyrimidine photodimers, and the initial step of detecting lesions is performed by a complex of damage surveillance proteins of the DDB1–DDB2 complex (Scrima et al. 2008). *Lamprologus lethops* shows an inactivating mutation in *DDB2*, a gene that is well characterized in humans, as mutations in this gene result in a form of xeroderma pigmentosa (reviewed by Tang and Chu [2002]). In mice, *DDB2*−/− and +/− individuals were hypersensitive to skin cancers induced by UV, whereas enhanced expression of *DDB2* delayed onset of such carcinomas (Alekseev et al. 2005). These studies suggest that *DDB2* plays an important role in the efficiency of UV-induced DNA repair. Although other molecular mechanisms for UV-induced DNA repair exist and may be functional in *L. lethops*, the loss of *DDB2* is likely to decrease the nucleotide excision repair capacity in this species. Interestingly, studies of *A. mexicanus* indicate that cave morphs have functional *DDB2* but have increased basal expression levels compared with surface forms, even in the absence of light (Beale et al. 2016). We also observe an inactivating mutation in a photolyase gene, an ancient and highly conserved family of genes which carry out repair of UV-damaged DNA and are related to circadian rhythm in numerous vertebrate taxa (Tamai et al. 2004; Zhao et al. 2018).

Evidence for Neutral Evolution and Direct Selection as Drivers of Trait Loss

For many cryptophthalmic or troglomorphic phenotypes, it remains unclear whether the loss of a trait arises due to a relaxation of selection and the accumulation of inactivating mutations via genetic drift, or whether the loss of a trait may in some cases arise through processes of directional selection. In the latter case, it is possible that when an inactivating mutation arises it will fix relatively rapidly due to positive selection, with a corresponding selective sweep accompanying it. This positive selection may be driven by a reduced energy expenditure that accompanies the reduction or elimination of certain morphological features (e.g., eyes) that have no function in a particular environment.

These and other nonmutually exclusive evolutionary processes have been proposed to contribute to trait loss, particularly of pigment and eye loss, in cave organisms (Culver et al.

1995; Culver and Wilkens 2000). First, relaxation of purifying selection on these traits can lead to the fixation of inactivating mutations through neutral processes (drift) (Barr 1964; Poulson and White 1969; Wilkens 1971, 1988). Second, the high energetic costs of these traits may result in positive selection on inactivating mutations in dark environments, rapidly sweeping inactivating alleles to fixation (Jeffery 2005; Borowsky and Cohen 2013; Moran et al. 2015). Finally, trait loss may be the result of trait integration, in which case inactivating mutations are linked to other traits that may be under directional selection (e.g., Yamamoto et al. 2009). Studies of diverse troglomorphic taxa have found varying support for these different hypotheses (Culver and Wilkens 2000; Leys et al. 2005; Hinaux et al. 2013; Klaus et al. 2013), with several recent studies highlighting drift/stochasticity as a major contributing factor (Niemiller et al. 2013; Stern and Crandall 2018). We hypothesize that in species that colonize a hypogean or troglomorphic environment through a single founder event with little subsequent gene flow with surface populations, drift would be expected to play a larger role than it would for cave species that maintain some level of past or present gene flow with surface forms, such as *A. mexicanus*.

To compare the two nonmutually exclusive hypotheses of drift versus selection in our data, we first sought to assess the nature of selective pressures that have acted on inactivated genes in the *L. lethops* lineage. Our analyses of divergence times between *L. lethops* and *L. tigripictilis* suggest they last shared a common ancestor around 1.35 (±0.06) Ma (fig. 3 and supplementary fig. 3, Supplementary Material online). Assessments of demographic changes within *L. lethops* and *L. tigripictilis* show a substantial difference in reconstructed population histories for the two species (fig. 3*b* and *c*). Although both experienced a population reduction starting ~1 Ma, coincident with a period of aridity across Africa and possible reduction in Congo River discharge (Dupont et al. 2001; Peter 2004; Bonnefille 2010), effective population size in *L. tigripictilis* appears to have grown substantially following this reduction, whereas *L. lethops* remained at a very low effective size. Correspondingly, genome-wide levels of heterozygosity are significantly lower in *L. lethops* compared with *L. tigripictilis* (paired *t*-test; $t = -19.123$, df = 777, *P* value <0.0001, fig. 3*d*). These findings suggest that genetic drift likely contributed to the fixation of inactivating mutations in *L. lethops*.

Interestingly, we do not see a significant increase in evolutionary divergence among the 15 genes found to have likely inactivating variants in *L. lethops*, with the ratios of nonsynonymous divergence (d*N*) to synonymous divergence (d*S*) being highly similar between this species and *L. tigripictilis* when compared with the more distantly related *N. brichardi* (fig. 3*e* and supplementary table 3, Supplementary Material online). Of the 15 genes, we were able to confidently align 13 orthologs from the ten cichlid species for which genomes were available at the time of analysis. Within these 13 genes, we assessed whether they had evolved at an elevated (increased) rate (based on d*N*/d*S* comparisons) along the *L. lethops* branch relative to the other species. Such an observation could suggest directional selection acting upon this gene. For all genes analyzed, the d*N*/d*S* ratio for the *L. lethops* branch was not found to be significantly different from that of other cichlids (supplementary table 4, Supplementary Material online). However, three of these genes had d*S* values of 0, an observation that strongly suggests either a selective sweep or a recent bottleneck in *L. lethops*.

Taken together, our results suggest that genetic drift and a corresponding relaxation of purifying selection, rather than strong directional selection, may best explain morphological degeneration in *L. lethops*. However, extremely low heterozygosity in this species reduces the statistical power to detect selection. An additional observation that supports the drift hypothesis is that in one of our focal genes (*impg2b*), the inactivating variant is segregating (not fixed) in our *L. lethops* sample. Future studies using a larger sample size of individuals and genes will be needed to fully test the contribution of selective pressures in producing and maintaining troglomorphic or cryptophthalmic phenotypes.

Learning about the Ecology of Cryptic Organisms through Their Genomes

In this study, we have focused on genes with inactivating variants of presumed large effect in *L. lethops* and for which functional evidence of their physiological role is available. It is possible that some of the inactivating mutations observed could be artifacts of the reference data used rather than true biological variants. Likewise, there are likely to be additional inactivating mutations contributing to the unique phenotypes of *L. lethops* that were not characterized with the methods used here, particularly as eye loss is primarily affected through altered gene expression during development. Nonetheless, our observations shed light on the genetic basis of the unusual traits in this species, and additional genetic influences may be discovered as reference annotations improve.

The falling costs of whole-genome sequencing now allow broader molecular evolutionary comparisons across the tree of life than have been feasible previously. For taxa like *L. lethops*, which inhabit ecosystems that are difficult or impossible to sample, genomic information can bring glimpses into the ecology of these species that would not have been possible otherwise (see also Wang et al. 2019). For example, although we have hypothesized that *L. lethops* lives at extreme depths because all specimens are recovered dead or moribund with subcutaneous embolism and burst gas bladders, our inability to sample its habitat directly has made

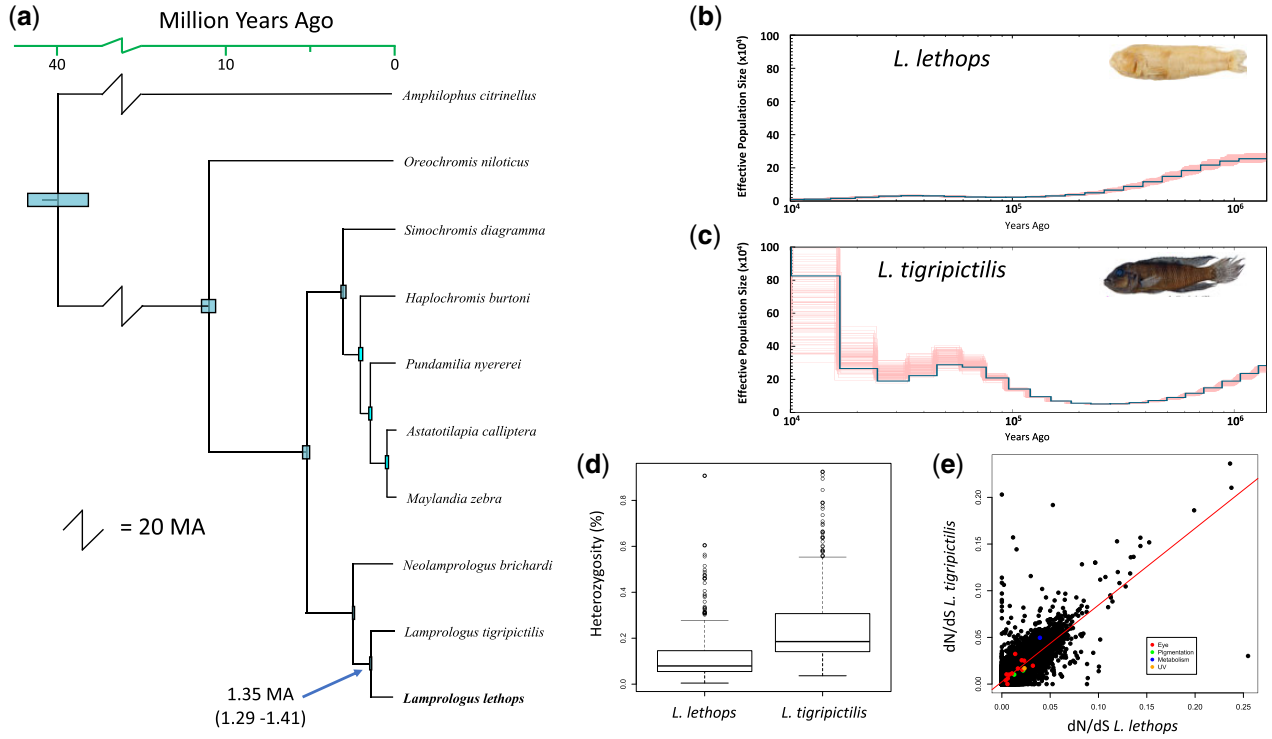

Fig. 3—Divergence and demographic changes. (*a*) Maximum-likelihood phylogenetic analysis of ten cichlid species, using conserved coding sequences (as determined using BUSCO). All relationships had 100% bootstrap support. The scale indicating divergence times (in millions of years) is based on the estimated age. PSMC analysis of effective population size changes in (*b*) *Lamprologus lethops* and (*c*) *Lamprologus tigripictilis* over the last 1.3 Myr, indicating their different demographic histories after species divergence. (*d*) A comparison if scaffold-wide chromosome levels (for all scaffolds larger than 10 kb in the reference genome, *Neolamprologus brichardi*) for *L. lethops* (mean: 0.157%, SD: 0.210%), and *L. tigripictilis* (mean: 0.274%, SD: 0.265). These differences are statistically significant (paired *t*-test; $t = -19.123$, df $= 777$, *P* value <0.0001). (*e*) A comparison of nonsynonymous versus synonymous evolutionary changes (as d*N*/d*S*) between *L. lethops* and *L. tigripictilis*. Genes were determined from the annotation of the reference genome, *N. brichardi*. No gene had a ratio >1 in either species. Colored points represent genes of interest; those with a likely inactivating mutation effecting pigmentation (green points), eye formation (red points), metabolism (blue point), or UV damage repair (orange point). Generally, more of these genes had a higher d*N*/d*S* ratio in *L. tigripictilis* than in *L. lethops*.

testing this hypothesis all but impossible. However, the disabled *DDB2* and photolyase genes in the genome of *L. lethops* strongly suggest that this species does not spend significant time exposed to UV light. Moreover, the loss of a functional spexin peptide suggests that *L. lethops* lives in a nutrient-poor environment in which access to food is highly sporadic. Because of advances in comparative genomics and an improved understanding of genome-to-phenome processes across the tree of life, we are poised to take advantage of genomic data to infer aspects of natural history from difficult-to-study species using genomic data alone. In the future, it may be possible to infer aspects of life history and ecology such as longevity, fertility, parasite susceptibility, and diet using a comparative genomic framework. Within this framework, *L. lethops* provides an intriguing example of how genomic data have the potential to contribute to our understanding of the habitat and life history of nonmodel organisms that may be difficult or impossible to study in the field.

Conclusion

Our draft genome analysis of *L. lethops* yielded new insights into the genomic basis and evolutionary history of this highly unusual cichlid species, allowing comparisons to other diverse fishes that share strikingly similar phenotypes (Alter et al. 2015; Stiassny and Alter, forthcoming). Future studies that extend this comparative framework across an even broader swath of the teleostean tree of life, from the population level (e.g., *Astyanax*) to families and orders, will allow deeper insights into the genomic mechanisms leading to phenotypic convergence, as well as a more sophisticated understanding of the selective and demographic processes that drive the evolution of troglomorphic traits. We have shown that multiple inactivating mutations have accumulated in *L. lethops*, which likely contribute to its unique, degenerative phenotype. Genomic analyses revealed inactivating mutations in genes related to pigmentation, retinal and lens development, and metabolism, providing insights into the evolutionary mechanisms that allow this species to occupy a deep, dark, and

isolated habitat. These inactivating mutations arose and spread relatively rapidly after this species diverged, possibly as a consequence of directional selection. However, we also see strong evidence of a major and long-lasting population bottleneck in *L. lethops*, which likely resulted in faster fixation of inactivating mutations that contribute to the unusual phenotype observed in this species. Genomic analyses of additional individuals will be needed to further explore this question.

Materials and Methods

Samples and DNA Preparation

Specimens of the cryptophthalmic cichlid, *L. lethops*, and a related sympatric species, *L. tigripictilis*, were obtained from the Bulu-Luozi region, Kongo Central Province, Democratic Republic of Congo. Muscle tissues were preserved in 95% ethanol until tissue subsampling. We used ~20 mg of tissue each from one individual of *L. lethops* (AMNH 263957) and *L. tigripictilis* (AMNH 263989) for DNA extraction. We extracted DNA from these specimens using the Gentra Puregene kit (Qiagen), according to manufacturers' instructions with the exception of an additional 30-min incubation with RNase A (ThermoFisher) at 37 °C. DNA quality and concentration were measured on a Bioanalyzer (Agilent Technologies).

We prepared a 1-µg aliquot of genomic DNA from each sample using the Illumina TruSeq PCR-free DNA HT sample preparation kit with 550-bp insert size. Intact genomic DNA was sheared using the Covaris sonicator (adaptive focused acoustics), followed by end-repair and bead-based size selection of fragmented molecules and ligation of Illumina barcodes. After ligation, we performed a final library QC, which included a measurement of the average size of library fragments using a FragmentAnalyzer, estimation of the total concentration by PicoGreen, and a measurement of the yield and efficiency of the adapter ligation process with a quantitative PCR assay (Kapa) using primers specific to the adapter sequence.

Sequencing and Read Processing

The two whole-genome sequencing libraries were pooled and sequenced according to the Illumina protocol over two lanes of an Illumina HiSeq 2500, generating 125-bp, paired-end sequence reads. This yielded ~300 million raw, paired-end reads per sample. We removed read duplicates using custom scripts (NYGC), then used Cutadapt 1.8.1 (Martin 2011) to trim adapters and low-quality bases ($Q = 30$). Next, we performed Enterobacteria phage phiX decontamination by mapping the reads against the phiX174 reference genome (GenBank accession number: NC_001422.1) with GEM mapper (Marco-Sola et al. 2012). Finally, we performed read error correction using Lighter 1.0.7 (Song et al. 2014).

Reference Mapping and Degenerative Gene Evolution

To assess genomic variation in *L. lethops* that could contribute to its cryptophthalmic phenotype, we examined *lethops*-specific genetic variation in relation to other cichlids. To do this, we first mapped the trimmed and filtered read sets to a reference genome of the Lake Tanganyikan lamprologine, *N. brichardi* (NeoBri1.0, GCF_000239395.1, Brawand et al. 2014) using BWA-MEM (v. 0.7.15) with default settings (Li 2013). Although most duplicated sequences were removed in the initial processing of the data, additional duplicates were identified and marked by using MarkDuplicates from Picard (v. 1.77; http://broadinstitute.github.io/picard/; last accessed May 31, 2019). This was followed by indel realignment using IndelRealigner from the Genome Analysis Toolkit (GATK v. 3.8; McKenna et al. 2010). The resulting BAM files with the mapped, sorted, and realigned reads are available in the NCBI Short Read Archive under the BioProject accession number PRJNA577474 and accession numbers SRR10811633 and SRR10811634.

Variant calling was done using GATK's HaplotypeCaller (specific flags envoked: –emitRefConfidence GVCF, –variant_index_type LINEAR, –variant_index_parameter 128000 -rf BadCigar). The resulting VCFs of the two species were then combined and variants genotyped with GATK's GenotypeGVCFs. Next, we separated single-nucleotide variants (SNVs) and insertion/deletions (INDELs) into two separate VCF files. "Mixed" variants (SNVs and INDELs combined) were included in the INDEL variant file.

We removed SNVs that had a quality by depth (QD) <10.0, mapping quality (MQ) <40.0, Fisher strand bias (FS) >60.0, a strand odds ratio (SOR) >3.0, mapping quality rank sum (MQRankSum) <-5.0, and read position rank sum (ReadPosRankSum) <-6.0. All filtering options were based on the developer's recommended cutoffs, with adjustments for QD, ReadPosRankSum, and MQRankSum based on the observed distributions for these parameters (supplementary fig. 4, Supplementary Material online). We removed INDELs and mixed INDELs/SNVs with QD < 10.0, FS > 200.0, SOR $>$ 10.0, and ReadPosRankSum > 20.0. Again, these thresholds were chosen based on the developer's recommendations with adjustments made based on the observed distributions (supplementary fig. 5, Supplementary Material online). After filtering, SNVs, INDELs, and mixed variants were recombined into a single gVCF.

Next, we used the program SnpEff (v. 4.3m, Cingolani et al. 2012) with a custom database for the gene annotations of *N. brichardi* to annotate our filtered variants. With SnpEff, we used the "classic" option, with the addition of the "no-intergenic," "no-intron," and "no-utr" options. All other parameters were default. After annotation, we filtered the variants to leave only those that were divergent from the reference in *L. lethops* but fixed and matching the reference

in *L. tigripictilis*. The former were assumed to be derived, and the latter ancestral. We then further sorted our list of *lethops*-specific variants to those that were likely to be gene-inactivating mutations. These were variants of "high" impact as determined with SnpEff and included frameshift mutations, loss of the start codon, gain of stop codon, loss of stop codon, or disruption of an exonic splice site.

For each of these genes, we examined the relevant region by eye using the Integrative Genomics Viewer (Robinson et al. 2011; Thorvaldsdóttir et al. 2013) to confirm that the inactivating variants were present in *L. lethops* but not in *L. tigripictilis* (see supplementary table 2, Supplementary Material online, for the reference location, depth of coverage, read orientation, and reference transcript ID for each high impact variant analyzed here). Read pileups showing 80 nucleotides around each inactivating variant for both *L. lethops* and *L. tigripictilis* are available on Dryad (https://doi.org/10.5061/dryad.3ffbg79db). These pileups were generated using the "tview" function in SAMtools v. 1.9 (Li et al. 2009). We also compared the exon containing the inactivating variant (based on the annotation of *N. brichardi*) with annotated exons of six other cichlid species and five other closely related fishes for which genomes were available (see supplementary Methods, Supplementary Material online). Finally, for three of the genes (oca2, spexin, and RP1L1), we independently verified loss-of-function mutations with PCR and Sanger sequencing in additional specimens of *L. lethops* (see supplementary Methods, Supplementary Material online).

In conjunction with our *L. lethops*-specific inactivating variants analysis, we compared the annotated gene set of *N. brichardi* with the gene set of the zebrafish, *D. rerio*. To do this, we used a local protein blast (BlastP) with an *e*-value of 1e-10. The best hit (as defined by percent similarity and match length) was considered a possible ortholog in further analyses. Genes with potential inactivating variants unique to *L. lethops* were assessed for gene ontology (GO) using the *D. rerio* annotations and our correlation of these with *N. brichardi* genes. We were interested in five phenotypes: eye degeneration, loss of pigmentation, potential loss of circadian rhythm, potential loss of UV damage repair mechanisms, and potential shifts in metabolism (specific GO Biological Terms used in this search are given in supplementary table 1, Supplementary Material online). Genes that had an inactivating variant in *L. lethops*, had a likely ortholog in *D. rerio*, and that matched one of more of the GO terms searched were then further considered.

De Novo Genome Assembly, Phylogenetic Analysis, and Divergence Time Estimates

To assess the degree to which demographic changes and/or selection may have influenced the emergence of cryptophthalmic traits in *L. lethops*, we needed to determine when this species evolved and whether it has undergone any changes in population size. To do this, we assembled de novo genomes for *L. lethops* and *L. tigripictilis* using the trimmed and filtered Illumina short reads described above. Assembly was performed with the program ABySS (v. 2.0.2) with default settings and a kmer (*k*) value of 96 (supplementary table 5, Supplementary Material online). After assembly, we used the program BUSCO (v. 3.0.1, Simão et al. 2015) to assess the completeness of these two de novo assembled genomes in comparison to the Actinopterygian (ray-finned fishes) data set (supplementary table 5, Supplementary Material online). We also performed this same analysis on all available cichlid genomes (at time of analysis) for the purposes of phylogenetic comparison and to estimate a divergence time for *L. lethops* (supplementary fig. 6, Supplementary Material online).

BUSCO genes that were determined to be complete and single copy in all ten cichlid genomes were consolidated into a single FASTA file (one FASTA per gene), then aligned based on amino acid similarity using the program TranslatorX (Abascal et al. 2010). We removed poorly aligned regions with Gblocks (Castresana 2000; Talavera and Castresana 2007). We also used a custom Perl script to flag any potentially misaligned sequences based on high rates of consecutively divergent sites in any sample relative to the others. Flagged sequences were examined by eye for proper alignment, then either realigned or else removed from further analyses.

Next, we produced individual gene trees for each of our aligned BUSCO gene regions (2,468 regions). To do this, we first selected the most appropriate model of divergence for the region based on corrected AIC score using jModelTest 2 (v. 2.1.10, Darriba et al. 2012). Then with this model, we performed a phylogenetic analysis for the genetic region using PhyML (v. 3.1, Guindon et al. 2010). Finally, we selected 200 regions first based on bipartition support (relative to the species tree) and root-to-tip variance (preferentially selecting regions with lower variance). These values for each gene were calculated using the program SortaDate (Smith et al. 2018).

To estimate approximate species' splitting times, we concatenated these 200 genetic regions, then partitioned this data set into the first, second, and third codon positions. We used the Bayesian Evolutionary Analysis Utility tool (BEAUti v. 2.3.2) to create the XML file for implementation in BEAST. We ran two Markov chain Monte Carlo chains of 10^8 iterations in the program BEAST (v. 2.5.2, Drummond and Rambaut 2007), with an estimated, strict molecular clock. We used the Generalized Time Reversible substitution model (Tavaré 1986), with 1% of invariant sites and four gamma categories, and a calibrated Yule model (Heled and Drummond 2012). To help calibrate the analysis and determine when *L. lethops* arose, we used a prior of 2.36 Ma with a log normal distribution and an *S* parameter of 0.006 for the origin of the tribe Lamprologini (Irisarri et al. 2018). In our analysis, the tribe Lamprologini is represented by *L. lethops*, *L. tigripictilis*, and *N. brichardi*.

We used a burn-in period of the initial 10% of states, and parameters were logged every 1,000 iterations. LogCombiner v.1.8.1 was used to merge two separate runs. Log files were checked using Tracer (v. 1.7.1; Rambaut et al. 2014) to ensure that an effective sampling size (ESS) >200 was achieved for each parameter. Divergence times were estimated based on the 95% highest posterior density interval.

Lamprologus Demographic History

Population reductions, particularly in the case of severe bottlenecks, can result in rapid fixation of substitutions due to accelerated genetic drift. It is probable that the riverine reach between Bulu and Luozi was colonized by a relatively small number of individuals representing the progenitor of *L. lethops*. Moreover, given the environmental and physiological challenges presented by their putative isolated habitat, we hypothesize that this population remained at a small size relative to other cichlid species. For these reasons, it is possible that the emergence of inactivating mutations leading to degenerative trait evolution may have spread and been maintained in this species through neutral demographic processes alone. To examine the demographic history of *L. lethops*, we used the pairwise sequentially Markovian coalescent (PSMC) model, as implemented in the PSMC software package (v. 0.6.5-r67, Li and Durbin 2009). The PSMC uses patterns of heterozygosity across the genome to infer demographic changes. To establish heterozygous regions in *L. lethops* and *L. tigripictilis*, we first masked all INDELs in our filtered variant gVCF for our combined species. Then, using the "FastaAlternateReferenceMaker" tool in GATK, we produced genomes of each species, incorporating our high-quality single-nucleotide polymorphisms. IUPAC symbols were used to indicate heterozygous sites. We used this approach to examining heterozygosity rather than mapping our reads back to our de novo reference genomes because it allowed us to more accurately make relative comparisons between *L. lethops* and *L. tigripictilis*.

Time parameters were set to $4 + 25 \times 2 + 4 + 6$ (Li and Durbin 2009). We bootstrapped the analyses with 100 iterations to determine the robustness of our demographic estimates. Although the average generation time for these cichlids in the wild is unknown, following Won et al. (2005) we assumed one generation every 2 years. We also assumed a mutation rate (μ) of 3.5×10^{-9} (95% CI: 1.6×10^{-9} to 4.6×10^{-9} per bp per generation [Malinsky et al. 2018]). We calculated genome-wide levels of heterozygosity using a custom Perl script. Heterozygosity was determined for genome scaffolds that in the *N. brichardi* reference are >10 kb in length.

Selection on Inactivated Genes of Interest

To examine selection, we first produced maximum-likelihood estimates for the ratio of nonsynonymous to synonymous divergence (d*N*/d*S*) for each gene using PAML's codeml program (PAML v. 4.8, Yang 2007). These ratios were determined in relation to *N. brichardi*. For our 15 genes found to have *lethops*-specific inactivating mutations, we compared these ratios in *L. lethops* and *L. tigripictilis* using a pairwise *t*-test as implemented in R (v. 3.5.1, R Core Team 2018).

We used branch-specific models with PAML's codeml program (PAML v. 4.8, Yang 2007) to examine d*N*/d*S* ratios in *L. lethops* compared with the other cichlid species in each of our genes of interest. We first aligned the orthologous sequences from each of ten cichlid species. Orthologous sequences were found through a combination of gene-name search (when species' genomes were annotated) and nucleotide blasts (using "BlastN," BLAST v. 2.7.1). We aligned orthologous regions based on amino acid similarity with MUSCLE (Edgar 2004) as implemented in SeaView (v. 4.6.3, Galtier et al. 1996; Gouy et al. 2010), and then checked the sequences by eye for correct alignment. Any region that had a gap in any species was removed, along with any adjacent, segregating amino acids. Segregating synonymous nucleotides adjacent to a gap were not removed. Lastly, we removed the codon(s) that overlapped with the *L. lethops*-specific inactivating mutation. All sequences retained 100 or more codons after this sequence cleanup.

With our aligned sequences, we compared two models of evolution in our data. The null model had a fixed rate of synonymous and nonsynonymous evolution for all species (one d*N*/d*S* ratio for all branches, codeml model = 0), whereas the alternative model allowed the *L. lethops* branch to have a different rate of evolutionary change (one d*N*/d*S* for *L. lethops* and one d*N*/d*S* ratio for all other branches, codeml model = 2). Likelihood ratio tests were used to determine if the likelihoods of the models were statistically different.

Supplementary Material

Supplementary data are available at *Genome Biology and Evolution* online.

Acknowledgments

We gratefully acknowledge assistance from Oliver Lucanus, Tobit Liyandja, and Raoul Monsembula in obtaining specimens. We are also grateful to Margaret MacNeil and Kimberly Bernotas for assistance with imaging and light microscopy analysis of fish tissues. Funding was provided by National Science Foundation/DEB Award 1655694 to S.E.A. and M.L.J.S. and PSC-CUNY Award 41. M.L.A. was supported in part by a Gerstner Fellowship in Bioinformatics and Computational Biology to the American Museum of Natural History.

Author Contributions

M.L.J.S. and S.E.A. conceived the study. M.L.A., M.L.J.S., and S.E.A. designed the study. S.E.A. performed laboratory work.

M.L.A. conducted the genome assembly and bioinformatics analyses. M.L.J.S. conducted the morphological comparisons. All authors prepared, edited, and approved the manuscript.

Literature Cited

Abascal F, Zardoya R, Telford MJ. 2010. TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. Nucleic Acids Res. 38(Suppl 2):W7–W13.

Alekseev S, Kool H, Rebel H, Fousteri M, Moser J, Backendorf C, de Gruijl FR, Vrieling H, Mullenders LH. 2005. Enhanced DDB2 expression protects mice from carcinogenic effects of chronic UV-B irradiation. Cancer research. 65(22):10298–10306.

Alter SE, Brown B, Stiassny ML. 2015. Molecular phylogenetics reveals convergent evolution in lower Congo River spiny eels. BMC Evol Biol. 15(1):1–2.

Aspiras AC, Rohner N, Martineau B, Borowsky RL, Tabin CJ. 2015. Melanocortin 4 receptor mutations contribute to the adaptation of cavefish to nutrient-poor conditions. Proc Natl Acad Sci U S A. 112(31):9668–9673.

Bandah-Rozenfeld D, et al. 2010. Mutations in IMPG2, encoding interphotoreceptor matrix proteoglycan 2, cause autosomal-recessive retinitis pigmentosa. Am J Hum Genet. 87(2):199–208.

Barr HJ. 1964. The epistemology of causality from the point of view of evolutionary biology. Philos Sci. 31(3):286–288.

Beale AD, Whitmore D, Moran D. 2016. Life in a dark biosphere: a review of circadian physiology in "arrhythmic" environments. J Comp Physiol B 186(8):947–968.

Berman JR, Skariah G, Maro GS, Mignot E, Mourrain P. 2009. Characterization of two melanin-concentrating hormone genes in zebrafish reveals evolutionary and physiological links with the mammalian MCH system. J Comp Neurol. 517(5):695–710.

Bilandžija H, Abraham L, Ma L, Renner KJ, Jeffery WR. 2018. Behavioural changes controlled by catecholaminergic systems explain recurrent loss of pigmentation in cavefish. Proc R Soc B 285(1878):20180243.

Bilandžija H, Ma L, Parkhurst A, Jeffery WR. 2013. A potential benefit of albinism in Astyanax cavefish: downregulation of the oca2 gene increases tyrosine and catecholamine levels as an alternative to melanin synthesis. PLoS One 8(11):e80823.

Bonnefille R. 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical Africa. Global Planet Change 72(4):390–411.

Borowsky R, Cohen D. 2013. Genomic consequences of ecological speciation in Astyanax cavefish. PLoS One 8(11):e79903.

Brawand D, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513(7518):375–381.

Brilliant MH, Gondo Y, Eicher EM. 1991. Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning. Science 252(5005):566–569.

Caduff M, Bauer A, Jagannathan V, Leeb T. 2017. OCA2 splice site variant in German Spitz dogs with oculocutaneous albinism. PLoS One 12(10):e0185944.

Cai X, Patel S. 2010. Degeneration of an intracellular ion channel in the primate lineage by relaxation of selective constraints. Mol Biol Evol. 27(10):2352–2359.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol. 17(4):540–552.

Cavallari N, et al. 2011. A blind circadian clock in cavefish reveals that opsins mediate peripheral clock photoreception. PLoS Biol. 9(9):e1001142.

Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly 6(2):80–92.

Culver DC, Kane TC, Fong DW. 1995. Adaptation and natural selection in caves: the evolution of Gammarus minus. Cambridge: Harvard University Press.

Culver DC, Wilkens H. 2000. Critical review of the relevant theories of the evolution of subterranean animals. In: Ecosystems of the world. Vol. 30. p. 381–398. Amsterdam: Elsevier.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat Methods. 9(8):772.

Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol. 7(1):214.

Dupont LM, Donner B, Schneider R, Wefer G. 2001. Mid-Pleistocene environmental change in tropical Africa began as early as 1.05 Ma. Geology 29(3):195–198.

Emerling CA, Springer MS. 2014. Eyes underground: regression of visual protein networks in subterranean mammals. Molecular Phylogenetics and Evolution. 78:260–270.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32(5):1792–1797.

Fischer RM, et al. 2013. Co-expression of VAL-and TMT-opsins uncovers ancient photosensory interneurons and motorneurons in the vertebrate brain. PLoS Biol. 11(6):e1001585.

Foulkes NS, Whitmore D, Vallone D, Bertolucci C. 2016. Studying the evolution of the vertebrate circadian clock: the power of fish as comparative models. In Advances in Genetics Vol. 95, pp. 1–30. Academic Press.

Friedman JS, et al. 2006. Premature truncation of a novel protein, RD3, exhibiting subnuclear localization is associated with retinal degeneration. Am J Hum Genet. 79(6):1059–1070.

Fukamachi S, et al. 2004. Conserved function of medaka pink-eyed dilution in melanin synthesis and its divergent transcriptional regulation in gonads among vertebrates. Genetics 168(3):1519–1527.

Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. Bioinformatics 12(6):543–548.

Gouy M, Guindon S, Gascuel O. 2010. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 27(2):221–224.

Green JA, Baker BI. 1991. The influence of repeated stress on the release of melanin-concentrating hormone in the rainbow trout. J Endocrinology. 128(2):261–266.

Gross JB, Borowsky R, Tabin CJ. 2009. A novel role for Mc1r in the parallel evolution of depigmentation in independent populations of the cavefish *Astyanax mexicanus*. PLoS Genet. 5(1):e1000326.

Gross JB, Furterer A, Carlson BM, Stahl BA. 2013. An integrated transcriptome-wide analysis of cave and surface dwelling *Astyanax mexicanus*. PLoS One 8(2):e55659.

Gross JB, Wilkens H. 2013. Albinism in phylogenetically and geographically distinct populations of Astyanax cavefish arises through the same loss-of-function Oca2 allele. Heredity 111(2):122–130.

Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst Biol. 59(3):307–321.

Heled J, Drummond AJ. 2012. Calibrated tree priors for relaxed phylogenetics and divergence time estimation. Systematic biology. 61(1):138-49.

Hinaux H, et al. 2013. De novo sequencing of *Astyanax mexicanus* surface fish and Pachón cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. PLoS One 8(1):e53553.

Howe DG, et al. 2012. ZFIN, the Zebrafish Model Organism Database: increased support for mutants and transgenics. Nucleic Acids Res. 41(D1):D854–D860.

Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ. 2010. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 26(5):231–239.
Hughes LC, et al. 2018. Comprehensive phylogeny of ray-finned fishes (*Actinopterygii*) based on transcriptomic and genomic data. Proc Natl Acad Sci U S A. 115(24):6249–6254.
Irisarri I, et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the radiation of Lake Tanganyika cichlid fishes. Nat Commun. 9(1):1–12.
Jeffery WR. 2005. Adaptive evolution of eye degeneration in the Mexican blind cavefish. J Hered. 96(3):185–196.
Kawauchi H, Kawazoe I, Tsubokawa M, Kishida M et al. 1983. Characterization of melanin-concentrating hormone in chum salmon pituitaries. Nature. 305:321–323.
Kawauchi H. 2006. Functions of melanin-concentrating hormone in fish. J Exp Zool Part Zool. 305A:321–323.
Kim EB, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the naked mole rat. Nature 479(7372):223–227.
Klaassen H, Wang Y, Adamski K, Rohner N, Kowalko JE. 2018. CRISPR mutagenesis confirms the role of oca2 in melanin pigmentation in *Astyanax mexicanus*. Dev Biol. 441(2):313–318.
Klaus S, et al. 2013. Rapid evolution of troglomorphic characters suggests selection rather than neutral mutation as a driver of eye reduction in cave crabs. Biol Lett. 9(2):20121098.
Kolodziejski PA, et al. 2018. Serum levels of spexin and kisspeptin negatively correlate with obesity and insulin resistance in women. Physiol Res. 67:45–56.
Kozak M. 1984. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. Nature 308(5956):241–246.
Kratochwil CF, Urban S, Meyer A. 2019. Genome of the Malawi golden cichlid fish (*Melanochromis auratus*) reveals exon loss of oca2 in an amelanistic morph. Pigment Cell Melanoma Res. 32:719–723.
Kumar S, et al. 2016. Decreased circulating levels of spexin in obese children. J Clin Endocrin Metab. 101(7):2931–2936.
Langecker TG, Schmale H, Wilkens H. 1993. Transcription of the opsin gene in degenerate eyes of cave-dwelling *Astyanax fasciatus* (Teleostei, Characidae) and of its conspecific epigean ancestor during early ontogeny. Cell Tissue Res. 273(1):183–192.
Leys R, Cooper SJB, Strecker U, Wilkens H. 2005. Regressive evolution of an eye pigment gene in independently evolved eyeless subterranean diving beetles. Biol Lett. 1(4):496–499.
Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 [q-bio]. http://arxiv.org/abs/1303.3997 (accessed August 12, 2019).
Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14):1754–1760.
Li H, et al. 2009. The Sequence Alignment/Map (SAM) format and SAMtools. Bioinformatics 25(16):2078–2079.
Liu YP, et al. 2017. Putative digenic inheritance of heterozygous RP1L1 and C2orf71 null mutations in syndromic retinal dystrophy. Ophthal Genet. 38(2):127–132.
Malinsky M, et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations interconnected by gene flow. Nat Ecol Evol. 2(12):1940–1955.
Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. 2010. Convergence in pigmentation at multiple levels: mutations, genes and function. Phil Transactions Roy Soc B: Biol Sci. 365(1552):2439-50.
Manga P. 2018. Molecular biology of albinism. In: Kromberg J, Manga P, editors. Albinism in Africa. Cambridge (MA): Academic Press. p. 99–119.
Marco-Sola S, Sammeth M, Guigó R, Ribeca P. 2012. The GEM mapper: fast, accurate and versatile alignment by filtration. Nat Methods. 9(12):1185–1188.
Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 17(1):10–12.
Matos-Cruz V, et al. 2011. Unexpected diversity and photoperiod dependence of the zebrafish melanopsin system. PLoS One 6(9):e25111.
Matsuda K. 2009. Recent advances in the regulation of feeding behavior by neuropeptides in fish. Ann N Y Acad Sci. 1163(1):241–250.
Matsuda K, et al. 2006. Central administration of melanin-concentrating hormone (MCH) suppresses food intake, but not locomotor activity, in the goldfish, *Carassius auratus*. Neurosci Lett. 399(3):259–263.
McKenna A, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20(9):1297–1303.
Meng F, et al. 2013. Evolution of the eye transcriptome under constant darkness in Sinocyclocheilus cavefish. Mol Biol Evol. 30(7):1527–1543.
Meunier I, et al. 2014. Frequency and clinical pattern of vitelliform macular dystrophy caused by mutations of interphotoreceptor Matrix IMPG1 and IMPG2 genes. Ophthalmology 121(12):2406–2414.
Moran D, Softley R, Warrant EJ. 2015. The energetic cost of vision and the evolution of eyeless Mexican cavefish. Sci Adv. 1(8):e1500363.
Near TJ, et al. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. Proc Natl Acad Sci U S A. 109(34):13698–13703.
Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ. 2013. Evidence for repeated loss of selective constraint in rhodopsin of amblyopsid cavefishes (Teleostei: Amblyopsidae). Evolution 67(3):732–748.
Niemiller ML, Soares D. 2015. Cave environments. In: Riesch R, Tobler M, Plath, M, editors. Extremophile fishes: ecology, evolution, and physiology of teleosts in extreme environments. Cham (Switzerland): Springer International Publishing. p. 161–191.
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. 111(44):E4736–42.
Peter BD. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene. Earth Planet Sci Lett. 220:3–24.
Poulson TL, White WB. 1969. The cave environment. Science 165(3897):971–981.
Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. 2007. Regressive evolution in the Mexican cave tetra, *Astyanax mexicanus*. Curr Biol. 17(5):452–454.
Protas M, et al. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat Genet. 38(1):107–111.
Protas M, et al. 2008. Multi-trait evolution in a cave fish, *Astyanax mexicanus*. Evol Dev. 10(2):196–209.
R Core Team. 2018. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing. Available from: https://www.R-project.org/
Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6. Available from: http://beast.bio.ed.ac.uk/Tracer. Accessed December 1, 2019.
Rinchik EM, et al. 1993. A gene for the mouse pink-eyed dilution locus and for human type II oculocutaneous albinism. Nature 361(6407):72–76.
Roberts TR, Stewart DJ. 1976. An ecological and systematic survey of fishes in the rapids of the lower Zaire or Congo River. Bull Museum Comp Zool. 147(6):239-317.
Robinson JT, et al. 2011. Integrative genomics viewer. Nature biotechnology. 29(1):24–26.
Saenko SV, et al. 2015. Amelanism in the corn snake is associated with the insertion of an LTR-retrotransposon in the OCA2 gene. Sci Rep. 5(1):17118.
Schobert CS, et al. 2013. Comparative ocular anatomy in a blind African cichlid fish, Lamprologus lethops. Vet Ophthalmol. 16(5):359–364.
Scrima A, et al. 2008. Structural basis of UV DNA-damage recognition by the DDB1–DDB2 complex. Cell 135(7):1213–1223.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 31(19):3210–3212.
Sitaram A, et al. 2009. Localization to mature melanosomes by virtue of cytoplasmic dileucine motifs is required for human OCA2 function. Molecular biology of the cell. 20(5):1464–1477.
Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: a practical approach to divergence-time estimation in the genomic era. PLoS One 13(5):e0197433.
Song L, Florea L, Langmead B. 2014. Lighter: fast and memory-efficient sequencing error correction without counting. Genome Biol. 15(11):509.
Stahl BA, Gross JB. 2015. Alterations in Mc1r gene expression are associated with regressive pigmentation in Astyanax cavefish. Development Genes and Evol. 225(6):367–375.
Stahl BA, Gross JB. 2017. A comparative transcriptomic analysis of development in two Astyanax cavefish populations. Journal of Experimental Zoology Part B: Molecular and Developmental Evolution. 328(6):515–532.
Stern DB, Crandall KA. 2018. The evolution of gene expression underlying vision loss in cave animals. Mol Biol Evol. 35(8):2005–2014.
Stiassny ML, Alter SE. In press. Life in the fast lane: diversity, ecology, and speciation of cichlids in the lower Congo River. The Behavior, Ecology and Evolution of Cichlid Fishes. Springer Science and Business Media, Berlin.
Talavera G, Castresana J. 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol. 56(4):564–577.
Tamai TK, Vardhanabhuti V, Foulkes NS, Whitmore D. 2004. Early embryonic light detection improves survival. Current Biology. 14(3):R104–105.
Tang J, Chu G. 2002. Xeroderma pigmentosum complementation group E and UV-damaged DNA-binding protein. DNA Repair 1(8):601–616.
Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences. Lectures Math Life Sci. 17:57–86.
Thorvaldsdœttir H, Robinson JT, Mesirov JP. 2013. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 14(2):178–192.
Walewski JL, et al. 2014. Spexin is a novel human peptide that reduces adipocyte uptake of long chain fatty acids and causes weight loss in rodents with diet-induced obesity. Obesity 22(7):1643–1652.
Wang K, et al. 2019. Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. Nat Ecol Evol. 3(5):823–833.
Wessinger CA, Rausher MD. 2015. Ecological transition predictably associated with gene degeneration. Mol Biol Evol. 32(2):347–354.
Wilkens H. 1971. Genetic interpretation of regressive evolutionary processes: studies on hybrid eyes of two Astyanax cave populations (Characidae, Pisces). Evolution 25(3):530–544.
Wilkens H. 1988. Evolution and genetics of epigean and cave *Astyanax fasciatus* (Characidae, Pisces). In: Hecht MK, Wallace B, editors. Evolutionary biology. Vol. 23. Boston (MA): Springer. p. 271–367.
Womack MC, Fiero TS, Hoke KL. 2018. Trait independence primes convergent trait loss. Evolution 72(3):679–687.
Won YJ, Sivasundar A, Wang Y, Hey J. 2005. On the origin of Lake Malawi cichlid species: a population genetic analysis of divergence. Proceedings of the National Academy of Sciences. 102(suppl 1):6581–6586.
Wong MK, et al. 2013. Goldfish spexin: solution structure and novel function as a satiety factor in feeding control. Am J Physiology-Endocrinology and Metabolism. 305(3):E348–366.
Yamamoto Y, Byerly MS, Jackman WR, Jeffery WR. 2009. Pleiotropic functions of embryonic sonic hedgehog expression link jaw and taste bud amplification with eye loss during cavefish evolution. Dev Biol. 330(1):200–211.
Yamashita T, et al. 2009. Essential and synergistic roles of RP1 and RP1L1 in rod photoreceptor axoneme and retinitis pigmentosa. J Neurosci. 29(31):9748–9760.
Yang Z. 2007. PAML4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24(8):1586–1591.
Zhao H, et al. 2018. Modulation of DNA repair systems in blind cavefish during evolution in constant darkness. Curr Biol. 28(20):3229–3243.
Zheng B, et al. 2017. Spexin suppress food intake in zebrafish: evidence from gene knockout study. Sci Rep. 7(1):14643.

Associate editor: Marta Barluenga