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], S. Elizabeth Alter[3,4,5*]

[1] Department of Biology, Montclair State University, Montclair, NJ, 07043

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

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

[4] The Graduate Center of the City University of New York, 365 Fifth Avenue, New York, NY 10016

[5] Department of Biology, York College/The City University of New York, 94-20 Guy R. Brewer Blvd, Jamaica NY 11451

* Correspondence

Keywords: Convergent evolution, cave fish, depigmentation, adaptation, vision genes, trait loss, troglomorphic

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 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 Mya, 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.

47 The independent and repeated loss of phenotypic traits across lineages offers the
48 opportunity to study the modes and predictability of genomic evolution under particular selective
49 regimes. Fishes living in lightless or low-light subterranean habitats number over 165 teleostean
50 species (Niemiller and Soares 2015), and represent one of the most well-examined cases of such
51 repeated trait loss. Many have evolved a distinctive set of morphological features, sometimes
52 called "troglomorphic syndrome" that are generally considered to be adaptations to a lightless or
53 light-poor, and limited resource environment. These features include loss of pigmentation and
54 reduction or loss of eyes, as well as enhancements in other aspects of sensory anatomy. Cave fishes
55 such as the Mexican cave tetra, *Astyanax mexicanus*, (Teleostei, Characiformes) in particular, have
56 emerged as model organisms for studying these traits (e.g. Protas et al. 2007, 2008; Jeffery et al.
57 2005). However, the loss of eyes, pigments, and other features is not limited to organisms that live
58 in caves, but is found across a variety of ecosystems, including deep sea, fossorial, parasitic, and
59 some riverine habitats. In particular, this phenomenon has been identified in a diverse set of fish
60 species that are endemic to the lower Congo River (Stiassny and Alter in press; Alter et al. 2015),
61 including the cichlid *Lamprologus lethops* (Teleostei, Ovalentaria).
62 Discovered and described in the 1970s (Roberts and Stewart 1976), *L. lethops* is the only
63 known blind and depigmented species in the otherwise colorful and highly diverse family
64 Cichlidae. Information about its habitat and ecology are extremely limited, and relatively few
65 specimens have ever been recovered. However, available evidence suggests that populations of *L.*
66 *lethops* live in highly turbulent waters at extreme depths in lower Congo River canyons (Stiassny
67 and Alter, in press). The hydrology of the lower Congo River is extraordinarily complex, and the
68 river is characterized by some of the world's largest rapids, high-velocity in-stream flows, and a
69 complex bathymetry with shallow rapids located adjacent to canyons with depths >220m 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. 1a). 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, in press). A related congener, *L. tigripictilis*, as well as 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. 1a, 2a,b and Supplementary Figs. 1a, 2), with the most obvious being complete lack of pigmentation and lack of image-forming eyes (cryptophthalmia). The evidence for the former includes complete absence of melanophores (Supplementary Fig. 2) 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. 2b and Supplementary Fig. 1a; 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, *Astyanax 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). *L. lethops* also shows a number of other characters commonly found in cave-dwelling *A. mexicanus*, such as enhanced neurocranial laterosensory canals and pores (Supplementary Fig. 1c), and enhanced fat deposits (Fig. 2b). 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. 1e,g). Such a heavily reinforced gas bladder supports the hypothesis that this species occupies a turbulent deep-

93 water habitat, as it would allow an increase in the range of depths over which an individual could
94 resist positive buoyancy if entrained in upwelling currents.

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

111 In addition to the timing of evolution, genomic data can also inform our understanding of
112 the evolutionary mechanisms underlying trait loss. One hypothesis posits that when the
113 maintenance of a trait is no longer under purifying selection, deleterious mutations (including
114 mutations that inactivate trait-associated genes) will accumulate through neutral processes (i.e.
115 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.

Here we investigate the genomic basis of the cryptophthalmic phenotype in *L. lethops* by 1) examining genes with variants of 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 non-cryptophthalmic 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 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 & Patel 2010, Wessinger & Rausher 2015), including *Astyanax* cavefish (Gross & 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

139 codon, frameshift mutations, and changes to exon/intron splice sites. Such mutations are generally
140 presumed to result in a gene with reduced or no function. To look for the presence of such
141 mutations, we generated Illumina paired-end data for *L. lethops* and mapped it to the genome of
142 *Neolamprologus brichardi*, the evolutionarily closest available annotated reference (Brawand et
143 al. 2014). For comparison, we also generated similar data for *L. tigripictilis*. We annotated *lethops*-
144 specific genetic variants for their predicted impact on known cichlid genes using functional
145 characterization based on the zebrafish, *Danio rerio* (Supplementary Table 1), and determined
146 specific variants that likely resulted in gene inactivation.

147 Our analysis of likely inactivating mutations in genes potentially contributing to the
148 cryptophthalmic phenotype of *L. lethops* resulted in 21 genes of interest. Of these, the majority
149 (17) related to eye formation (Fig. 2e and Supplementary Table 2), with additional genes
150 implicated in pigmentation, circadian rhythm, metabolism and UV damage repair. Frameshift
151 mutations were the most common route for inactivation among these genes (Supplementary Table
152 2), followed by premature stop codons. Two genes that influence eye formation (cryptochrome
153 circadian regulator 2 [cry2] and beaded filament structural protein 1 [bfsp1]) had both a frameshift
154 mutation and a premature stop codon. In two other eye formation genes (Synaptojanin 1 [synj1]
155 and Interphotoreceptor matrix proteoglycan 2b [impg2b]), the inactivating variant was
156 heterozygous (segregating).

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

161 including the characid *A. mexicanus* and the cyprinids *Sinocyclocheilus* spp. and *Phreatichthys*
162 *andruzzii*, though the specific mutations differ (Supplementary Table 2).

163 Other genes that show loss of function in *L. lethops* appear fully intact in *A. mexicanus*, but
164 previous work shows the expression of the gene is modified in the latter species. In still other
165 cases, we found that while the specific genes differed, the same broad gene families (e.g.
166 crystallins, opsins) were affected. Given the large phylogenetic distance between these species (ca.
167 230-280 Myr; Near et al. 2013), convergence at the gene/gene family levels suggests that the
168 mechanisms of evolution underlying troglomorphism and/or cryptophthalmia may be predictable
169 at the gene level for some traits. When inactivated genes are shared across these distant clades (e.g.
170 *oca2*), it is likely that such genes have low pleiotropic constraints and/or influence
171 developmentally independent traits (e.g. Womack et al. 2018).

172

173 **Loss of pigmentation**

174 Although >150 genes are associated with color variation in vertebrates (Hubbard et al.
175 2010), two genes have been primarily implicated in pigmentation loss or reduced melanism in
176 many cave fishes: *oca2* and *mc1r* (Protas et al. 2006, 2007; Gross et al. 2009; Stahl and Gross
177 2015). Deletions in the *oca2* gene in *A. mexicanus* have been linked to albinism (Protas et al. 2006,
178 2007) and have been shown via gene editing to result in albinism in surface populations (Klaassen
179 et al. 2018). In the *oca2* gene of *L. lethops*, a single nucleotide change at the second position of
180 the first codon, disrupts the start codon (Fig. 1). We confirmed that this mutation is monomorphic
181 across five additional individuals of *L. lethops* using PCR and Sanger sequencing (see
182 Supplementary Methods). In addition, we searched for possible non-canonical, in-frame initiation
183 sites upstream and downstream of the start codon; while methionine codons occur downstream,

184 they are not embedded within the Kozak consensus sequence (Kozak 1984) and therefore unlikely
185 to represent alternative start sites for translation. Disruption of the melanin pathway via loss of
186 functional oca2 transcripts is sufficient to explain the total lack of melanin pigmentation in *L.*
187 *lethops*. We also observe the accumulation of additional non-synonymous changes in the *oca2*
188 gene of *L. lethops* and a relatively higher ratio of non-synonymous to synonymous changes in *L.*
189 *lethops* (0. 0.0128) as compared to *L. tigripictilis* (0.0102) (excluding the start codon). This further
190 suggests that selection is no longer acting to maintain the function of this gene. This finding adds
191 to a growing number of examples, including a leucistic morph in another cichlid species,
192 *Melanochromis auratus* (Kratochwil et al. 2019), in which *oca2* is implicated in pigment reduction
193 or loss in vertebrates. Mechanisms of loss of function vary in oca2, from deletions of part or all of
194 exons 21 and 24 in the terminal end of the protein in *A. mexicanus* (Protas et al. 2006) to loss of
195 the second exon in amelanistic *M. auratus* (Kratochwil et al. 2019); strikingly, *L. lethops* appears
196 to be the first case in which the start codon has been lost. Several hypotheses have been put forth
197 about why *oca2* is such a frequent evolutionary target, including in humans. These include its large
198 size, developmental independence, and intriguingly, the possibility that freeing the substrates of
199 the *oca2* protein (e.g. L-tyrosine) could increase availability of the substrate for the catecholamine
200 pathway involved in feeding (Bilandžija et al. 2013) and sleep (Bilandžija et al. 2018). In contrast,
201 no differences in amino acid sequences of *mc1r* were observed between *L. lethops* and *L.*
202 *tigripictilis*, indicating that protein variation in *mc1r* (which causes reduced melanism in cavefish)
203 does not affect pigmentation in *L. lethops.*
204
205 **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 and Supplementary Table 2). These include Arrestin 3b, which is expressed in S- and UV-light sensitive photoreceptors in zebrafish; and retinal degeneration 3 (*rd3*), which encodes a protein that, when truncated, causes retinal degeneration in mice (Friedman et al. 2006). Retinal homeobox gene 1 (*rx1*), also inactivated in *L. lethops*, is an eye field transcription factor that plays a key role in eye development including retinal cell proliferation in the outer nuclear layer of the retina. In zebrafish, morpholino knockdown of *rx1* caused microphthalmia (Nelson et al. 2009); however, expression of *rx1* was not found to differ between cave and surface morphs of *A. mexicanus* (Strickler et al. 2002). 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, crygmxl1, crybgx, crybb3, cryba1a*) 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 under-expressed or show loss-of-function mutations (Kim et al. 2011; Hinaux et al. 2014). While most of the crystallins genes differ from those identified in *L. lethops*, several are the same: *crygm3* and *cryba1a* are not expressed at all in *Astyanax* (Gross et al. 2013), *crybgx* is expressed at a very reduced level in early development (Hinaux et al. 2014), and *crybb3* has a premature stop codon in the naked mole rat (Kim et al. 2011).

Several opsin family genes, including both visual and non-visual opsins, are also found to have splice site variants, frameshift mutations or premature stop codons in *L. lethops* (*opn1lw1*, *opn4xa* (= *opn4x-1*), *opn3*, *TMT-opsinb*). *Opn4x* (melanopsin) is expressed in retinal ganglion cells where it mediates circadian photoentrainment and sleep (Matos-Cruz et al. 2011). In a

229 comparative functional analysis, Cavallari et al. (2011) found premature stop codons in another
230 melanopsin paralog, *Opn4m2*, as well as *TMT-opsin*, in the Somalian cavefish, and provide
231 evidence that these two genes are responsible for disruption of the circadian rhythm in that species.
232 A subsequent study of zebrafish and the distantly related beloniform medaka (*Oryzias latipes*)
233 indicates that TMT-opsin is highly conserved across teleostean fishes and confers light sensitivity
234 to inter- and motor neurons (Fischer et al. 2013). Likewise, *opn1lw1* is expressed in a day-night
235 rhythm in zebrafish kept in darkness (Li et al. 2008). A premature stop codon and a frameshift
236 mutation in cryptochrome-2 (*cry2*) as well as an exon splice site variant in *opn1lw1* suggests
237 possible impacts on circadian rhythm in *L. lethops*.
238 A number of photoreception genes are functionally affected or disabled in *L. lethops* (e.g.,
239 Imp2/Impg2b, RP1L1; we confirmed the latter via PCR/Sanger sequencing; see Methods).
240 Mutations in the ortholog of the Imp2/Impg2b gene cause retinitis pigmentosa (Bandah-Rosenfeld
241 et al 2010) and vitelliform macular dystrophy (Meunier et al. 2014) in humans, whereas mutations
242 in RP1L1 cause occult macular dystrophy in humans and progressive photoreceptor degeneration
243 in mice (Yamashita et al. 2009). In zebrafish, suppression of RP1L1 along with another gene,
244 c2orf71, cause reduction of eye size and loss of rhodopsin in photoreceptors (Liu et al. 2017).
245
246 **Metabolic genes suggest increased appetite**
247 A premature stop codon in the neuropeptide spexin (*spx*) in *L. lethops* (confirmed via PCR
248 and Sanger sequencing, see Methods) may correspond with the extensive lipid deposits observed
249 in our anatomical analysis as compared with *L. tigripictilis* (Fig. 2b,d), and other cichlids. The
250 encoded spexin protein, which was discovered originally using bioinformatics methods and later
251 confirmed through protein purification and functional studies across vertebrates, functions in

252 appetite suppression. Zebrafish *spx*-/- mutants created using TALENs showed hyperphagia
253 (increased appetite), overexpression of orexigenic AgRP, and increased serum concentration of
254 glucose, triacylglycerol, cholesterol, and high-density lipoprotein cholesterol (Zheng et al. 2017).
255 Similarly, in goldfish, injection of *spx* suppressed appetite and feeding (Wong et al. 2013).
256 Consistent with these observations in teleosts, studies have linked low spexin levels to obesity and
257 insulin resistance in both rodents and humans (Walewski et al. 2014; Kumar et al. 2016;
258 Kolodziejskii et al. 2018). Interestingly, a different mechanism appears to underlie a similar
259 phenotype in *A. mexicanus*, as coding mutations in the melanocortin 4 receptor were linked to
260 hyperphagia in this species (Aspiras et al. 2015).

261 An additional loss-of-function mutation (a premature stop codon) is observed in pro-
262 melanin concentrating hormone (*pmch*), a neuropeptide implicated in melanin concentration
263 (Baker 1991; Kawauchi et al. 1983), energy regulation and appetite (Berman et al. 2009), and
264 stress response (Green and Baker 1991), in diverse teleostean species including zebrafish, rainbow
265 trout, flounder, eel, tilapia and chum salmon (Kawauchi 2006). This suggests that the peptide
266 encoded by *pmch* regulates food intake in many fishes (Berman et al. 2009; Matsuda et al. 2006,
267 2007).

268

269 **Inactivation of UV repair mechanisms**

270 Repair of UV-damaged DNA is a critical function in most fishes, but in fishes living in a
271 low light or light-free environments, the need for such repair processes may be diminished. DNA
272 repair mechanisms have been shown to be light-dependent in most teleosts. The nucleotide
273 excision repair pathway repairs UV-light induced pyrimidine photodimers, and the initial step of
274 detecting lesions is performed by a complex of damage surveillance proteins of the DDB1-DDB2

complex (Scrima et al. 2008). *L. 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 in Tang and Chu, 2002). In mice, DDB2-/- and +/- individuals were hypersensitive to skin cancers induced by UV, while enhanced expression of DDB2 delayed onset of such carcinomas (Alekseev et al. 2005). These studies suggest DDB2 plays an important role in the efficiency of UV-induced DNA repair. While 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 to surface forms, even in the absence of light (Beale et al. 2016). We also observe a loss-of-function 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 selective processes. 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 the reduction in energy expenditure that accompanies the reduction or elimination of certain morphological features (e.g. eyes) that have no function in a particular environment.

298 Several non-mutually exclusive evolutionary processes have been proposed to result in trait
299 loss, particularly of pigment and eye loss in cave organisms (Culver et al. 1995; Culver and
300 Wilkens 2000). First, relaxation of purifying selection on traits can lead to the fixation of
301 inactivating mutations through neutral processes (drift) (Barr 1964; Poulson and White
302 1969; Wilkens 1971; Wilkens 1988). Second, the high energetic costs of the traits may result in
303 positive selection on inactivating mutations in dark environments, rapidly sweeping these alleles
304 to fixation (Jeffery 2005; Borowsky and Cohen 2013; Moran et al. 2015). Finally, trait loss may
305 be the result of trait integration, in which case inactivating mutations are linked to other traits that
306 may be under directional selection (e.g. Yamamoto et al. 2009). Studies of diverse troglomorphic
307 taxa have found varying support for these different hypotheses (Culver and Wilkens 2000; Leys
308 et al. 2005; Hinaux et al. 2013; Klaus et al. 2013), with several recent studies highlighting
309 drift/stochasticity as a major contributing factor (Niemiller et al. 2013; Stern and Crandall 2018).
310 We hypothesize that in species that colonize a hypogean or troglomorphic environment through a
311 single founder event with little subsequent gene flow with surface populations, drift would be
312 expected to play a larger role than it would for cave species that maintain some level of past or
313 present gene flow with surface forms, such as *A. mexicanus*.
314 To compare the two non-mutually exclusive hypotheses of drift versus selection in our
315 data, we first sought to assess the nature of selective pressures that have acted on inactivated genes
316 in the *L. lethops* lineage. Our analyses of divergence times between *L. lethops* and *L. tigripictilis*
317 suggest they last shared a common ancestor around 1.35 (± 0.06) million years ago (Fig. 3 and
318 Supplementary Fig. 3). Assessments of demographic changes within *L. lethops* and *L. tigripictilis*
319 show a substantial difference in reconstructed population histories for the two species (Fig. 3b,c).
320 While both experienced a population reduction starting approximately one million years ago,

coincident with a period of aridity across Africa and possible reduction in Congo River discharge (Dupont et al. 2001; Peter 2004; Bonnefille 2010), effective size in *L. tigripictilis* appears to have grown substantially following this reduction, whereas *L. lethops* remained at a very low effective population size. Correspondingly, genome-wide levels of heterozygosity are significantly lower in *L. lethops* compared to *L. tigripictilis* (paired T-test; $t = -19.123$, $df = 777$, p-value < 0.0001, Figure 3d). These findings suggest that genetic drift likely contributed to the fixation of loss-of-function mutations in *L. lethops*.

Interestingly, we do not see a significant increase in evolutionary divergence among the 21 genes found to have likely inactivating variants in *L. lethops*, with the ratios of non-synonymous divergence (dN) to synonymous divergence (dS) being highly similar between this species and *L. tigripictilis* when compared to the more distantly related *N. brichardi* (Fig. 3e and Supplementary Table 3). Of the 21 genes, we were able to confidently align 16 orthologs for the 10 cichlid species for which genomes were available. Within these 16 genes we assessed whether they had evolved at an elevated (increased) rate (based on dN/dS comparisons) along the *L. lethops* branch relative to the other species. Such an observation could suggest directional selection acting upon this gene. From these analyses, one gene, a long-wave sensitive opsin (*opn1lw1*), had a dN/dS ratio for *L. lethops* that was significantly different from that of the other cichlids ($2\Delta l = 3.894$, $DF = 1$, $P = 0.048$, Supplementary Table 4). For this gene, the average dN/dS ratio across the other cichlids was 0.3317, whereas for *L. lethops* there were no synonymous changes. A lack of synonymous changes could arise from a rapid genetic sweep driven by selection. For the other genes analyzed, the dN/dS ratio for the *L. lethops* branch was not found to be significantly different from that of other cichlids (Supplementary Table 4). However, many of these genes also had dS values of 0, an observation that strongly suggests either a selective sweep or a recent bottleneck in *L. lethops*.

Taken together, these 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 two of our focal genes (synj1 and 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 variants of large effect in *L. lethops* and for which functional evidence of their physiological role is available. While these genes almost certainly do not represent the full suite of genetic changes responsible for the unusual phenotype observed in *L. lethops*, particularly as eye loss is primarily affected through altered gene expression during development, these observations nonetheless shed light on the genetic basis of the unusual traits in this species.

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, while 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 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 from their genomes 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 can contribute to our understanding of the habitat and life history of non-model 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 its striking phenotype (Alter et al. 2015; Stiassny and Alter in press). 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 pigment, 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.

Methods

Samples and DNA preparation

Specimens of the cryptophthalmic cichlid, *Lamprologus 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 approximately 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 1µg aliquot of genomic DNA from each sample using the Illumina TruSeq PCR-free DNA HT sample preparation kit with 550bp 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.

413 *Sequencing and read processing*

414 The two WGS libraries were pooled and sequenced according to the Illumina protocol over
415 two lanes of an Illumina HiSeq 2500, generating 125 bp, paired-end sequence reads. This yielded
416 approximately 300 million raw, paired-end reads per sample. We removed read duplicates using
417 custom scripts (NYGC), then used Cutadapt 1.8.1 (Martin 2011) to trim adapters and low-quality
418 bases (Q=30). Next, we performed Enterobacteria phage phiX decontamination by mapping the
419 reads against the phiX174 reference genome (GenBank accession number: NC_001422.1) with
420 GEM mapper (Marco-Sola et al. 2012). Finally, we performed read error correction using Lighter
421 1.0.7 (Song et al. 2014).

422

423 *Reference Mapping and Degenerative Gene Evolution*

424 To assess genomic variation in *L. lethops* that could contribute to its cryptophthalmic
425 phenotype, we examined *lethops*-specific genetic variation in relation to other cichlids. To do this,
426 we first mapped the trimmed and filtered read sets to a reference genome of the Lake Tanganyikan
427 lamprologine, *Neolamprologus brichardi* (NeoBri1.0, GCF_000239395.1, Brawand et al. 2014)
428 using BWA-MEM (v. 0.7.15) with default settings (Li 2013). Although most duplicated sequences
429 were removed in the initial processing of the data, additional duplicates were identified and marked
430 by using MarkDuplicates from Picard (v. 1.77; http://broadinstitute.github.io/picard/). This was
431 followed by indel realignment using IndelRealigner from the Genome Analysis Toolkit (GATK v.
432 3.8; McKenna et al. 2010). Variant calling was done using GATK's HaplotypeCaller (specific
433 flags envoked: --emitRefConfidence GVCF, --variant_index_type LINEAR, --
434 variant_index_parameter 128000 -rf BadCigar). The resulting VCFs of the two species were then
435 combined and variants genotyped with GATK's GenotypeGVCFs. Next, we separated single

436 nucleotide variants (SNVs) and insertion/deletions (INDELs) into two separate VCF files. 'Mixed'
437 variants (SNVs & INDELs combined) were included in the INDEL variant file.

438 We removed SNVs that had a quality by depth less than 10 (QD < 10.0), mapping quality
439 less than 40 (MQ < 40.0), Fisher strand bias greater than 60 (FS > 60.0), a strand odds ratio greater
440 than 3 (SOR > 3.0), mapping quality rank sum less than -5 (MQRankSum < -5.0), and read position
441 rank sum less than -6 (ReadPosRankSum < -6.0). All filtering options were based on the
442 developer's recommended cutoffs, with adjustments for QD, ReadPosRankSum, and
443 MQRankSum based on the observed distributions for these parameters (Supplementary Fig. 4).
444 We removed INDELs and mixed INDELs/SNVs with quality by depth less than 10 (QD < 10.0),
445 Fisher strand bias greater than 200 (FS > 200.0), strand odds ratio greater than 10 (SOR > 10.0),
446 and read position rank sum greater than 20 (ReadPosRankSum > 20.0). Again, these thresholds
447 were chosen based on the developer's recommendations with adjustments made based on the
448 observed distributions (Supplementary Fig. 5). After filtering, SNVs, INDELs and mixed variants
449 were recombined into a single gVCF.

450 Next, we used the program SnpEff (v. 4.3m, Cingolani et al. 2012) with a custom database
451 for the gene annotations of *N. brichardi* to annotate our filtered variants. With SnpEff, we used
452 the 'classic' option, with the addition of the 'no-intergenic', 'no-intron' and 'no-utr' options. All
453 other parameters were default. After annotation, we filtered the variants to leave only those that
454 were fixed and divergent from the reference in *L. lethops* but fixed and matching the reference in
455 *L. tigripictilis*. The former was assumed to be derived, and the latter ancestral. We then further
456 sorted our list of *lethops*-specific variants to those that were likely to be gene inactivating
457 mutations. These were variants of 'high' impact as determined with SnpEff, and included
458 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 (IGV; 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 for the location, depth of coverage, and read orientation for each high impact variant analyzed here). For three of the genes (oca2, spexin and RP1L1), we independently verified loss-of-function mutations with PCR and Sanger sequencing (see Supplementary Methods).

In conjunction with this analysis, we compared the annotated gene set of *N. brichardi* with the gene set of the zebrafish, *Danio 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 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 shown in Supplementary Table 1). 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 and 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

482 settings and a kmer (k) value of 96 (Supplementary Table 5). After assembly, we used the program
483 BUSCO (v. 3.0.1, Simão et al. 2015) to assess the completeness of these two *de novo* assembled
484 genomes in comparison to the Actinopterygian (ray-finned fishes) dataset (Supplementary Table
485 5). We also performed this same analysis on all available cichlid genomes for the purposes of
486 phylogenetic comparison and to estimate a divergence time for *L. lethops* (Supplementary Fig. 6).
487 BUSCO genes that were determined to be complete and single copy in all ten cichlid
488 genomes, were consolidated into a single FASTA file (one FASTA per gene), then aligned based
489 on amino acid similarity using the program TranslatorX (Abascal et al. 2010). We removed poorly
490 aligned regions with Gblocks (Castresana 2000, Talavera & Castresana 2007). We also used a
491 custom Perl script to flag any potentially misaligned sequences based on high rates of
492 consecutively divergent sites in any sample relative to the others. Flagged sequences were
493 examined by eye for proper alignment, then either realigned or else removed from further analyses.
494 Next, we produced individual gene trees for each of are aligned BUSCO gene regions
495 (2468 regions). To do this we first selected the most appropriate model of divergence for the region
496 based on corrected AIC score using jModelTest 2 (v. 2.1.10, Darriba et al. 2012). Then with this
497 model, we performed a phylogenetic analysis for the genetic region using PhyML (v. 3.1, Guindon
498 et al. 2010). Finally, we selected 200 regions first based on bipartition support (relative to the
499 species tree), and root-to-tip variance (preferentially selecting regions with lower variance). These
500 values for each gene were calculated using the program SortaDate (Smith et al. 2018).
501 To estimate approximate species' splitting times we concatenated these 200 genetic
502 regions, then partitioned this dataset into the first, second and third codon positions. We used the
503 Bayesian Evolutionary Analysis Utility tool (BEAUti v. 2.3.2) to create the XML file for
504 implementation in BEAST. We ran two MCMC chains of 10^8 iterations in the program BEAST

505 (v. 2.5.2, Drummond and Rambaut 2007), with an estimated, strict molecular clock. We used the
506 Generalized Time Reversible substitution model ('GTR', Tavaré 1986), with 1% of invariant sites
507 and four gamma categories, and a calibrated Yule model (Heled & Drummond 2012). To help
508 calibrate the analysis and determine when *L. lethops* arose, we used a prior of 2.36 MYA with a
509 log normal distribution and an S parameter of 0.006 for the origin of the tribe Lamprologini (Irisarri
510 et al. 2018). In our analysis the tribe Lamprologini is represented by *L. lethops*, *L. tigripictilis*, and
511 *N. brichardi.*

512 We used a burn-in period of the initial 10% of states, and parameters were logged every
513 1000 iterations. LogCombiner v.1.8.1 was used to merge two separate runs. Log files were checked
514 using Tracer (v. 1.7.1; Rambaut et al. 2014) to ensure that an effective sampling size (ESS) greater
515 than 200 were achieved for each parameter. Divergence times were estimated based on the 95%
516 highest posterior density (HPD) interval.

517

518 *Lamprologus Demographic History*

519 Population reductions, particularly in the case of severe bottlenecks, can result in rapid
520 fixation of substitutions due to accelerated genetic drift. It is probable that the riverine reach
521 between Bulu and Luozi was colonized by a relatively small number of individuals representing
522 the progenitor of *L. lethops*. Moreover, given the environmental and physiological challenges
523 presented by their putative isolated habitat, we hypothesize that this population remained at a small
524 size relative to other cichlid species. For these reasons, it is possible that the emergence of
525 inactivating mutations leading to degenerative trait evolution may have spread and been
526 maintained in this species through neutral demographic processes alone. To examine the
527 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 2011). 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 SNPs. 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 more accurately make relative comparisons between *L. lethops* and *L. tigripictilis*.

Time parameters were set to 4 + 25 x 2 + 4 + 6 (Li and Durbin 2011). 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 and colleagues (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 in the reference, *N. brichardi,* greater than 10Kb in length.

Selection on Inactivated Genes of Interest

To examine selection, we first produced maximum-likelihood estimates for the ratio of non-synonymous to synonymous divergence (dN/dS) 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 21 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 2017).

We used branch-specific models with PAML's codeml program (PAML v. 4.8, Yang 2007) to examine dN/dS ratios in *L. lethops* compared to 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), 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 non-synonymous evolution for all species (one dN/dS ratio for all branches, codeml model = 0), while the alternative model allowed the *L. lethops* branch to have a different rate of evolutionary change (one dN/dS for *L. lethops* and one dN/dS ratio for all other branches, codeml model = 2). Likelihood ratio tests were used to determine if the likelihoods of the models were statistically different.

Data Accessibility

571 Raw sequencing data are available from the NCBI Short Read Archive under the BioProject
572 accession number PRJNA577474. Alignments and chromatograms from PCR/Sanger sequencing
573 confirmation are available in Dryad (https://doi.org/10.5061/dryad.3ffbg79db).
574
575 **References**

576 Aspiras AC, Rohner N, Martineau B, Borowsky RL, Tabin CJ. 2015. Melanocortin 4 receptor
577 mutations contribute to the adaptation of cavefish to nutrient-poor conditions. P Natl Acad
578 Sci USA. 112:9668–9673.
579 Bandah-Rozenfeld D et al. 2010. Mutations in IMPG2, encoding interphotoreceptor matrix
580 proteoglycan 2, cause autosomal-recessive retinitis pigmentosa. Am J Hum Genet. 87:199–
581 208.
582 Barr HJ 1964. The Epistemology of Causality from the Point of View of Evolutionary Biology.
583 Philos Sci. 31, 286–288.
584 Beale AD, Whitmore D, Moran D. 2016. Life in a dark biosphere: a review of circadian physiology
585 in "arrhythmic" environments. J Comp Physiol B. 186:947–968.
586 Berman JR, Skariah G, Maro GS, Mignot E, Mourrain P. 2009. Characterization of two melanin-
587 concentrating hormone genes in zebrafish reveals evolutionary and physiological links with
588 the mammalian MCH system. J Comp Neurol. 517:695–710.
589 Bilandžija H, Ma L, Parkhurst A, Jeffery WR. 2013. A potential benefit of albinism in Astyanax
590 cavefish: downregulation of the oca2 gene increases tyrosine and catecholamine levels as an
591 alternative to melanin synthesis. PLoS One. 8:e80823.

592 Bilandžija H, Abraham L, Ma L, Renner KJ, Jeffery WR. 2018. Behavioural changes controlled
593 by catecholaminergic systems explain recurrent loss of pigmentation in cavefish. Proc Royal
594 Soc B. 285:20180243.
595 Bonnefille R. 2010. Cenozoic vegetation, climate changes and hominid evolution in tropical
596 Africa. Global Planet Change. 72:390-411.
597 Borowsky R, Cohen D. 2013. Genomic consequences of ecological speciation in Astyanax
598 cavefish. PloS one. 8:e79903.
599 Brawand D et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish.
600 Nature. 513:375–381.
601 Cai X, Patel S. 2010. Degeneration of an intracellular ion channel in the primate lineage by
602 relaxation of selective constraints. Mol Biol Evol. 27:2352–2359.
603 Cavallari N et al. 2011. A blind circadian clock in cavefish reveals that opsins mediate peripheral
604 clock photoreception. PLoS Biol. 9:e1001142.
605 Cingolani P, et al. 2012. A program for annotating and predicting the effects of single nucleotide
606 polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-
607 2; iso-3. Fly. 6:80–92.
608 Culver DC, Kane TC, Fong DW. 1995. *Adaptation and natural selection in caves: the evolution*
609 *of Gammarus minus*. Harvard University Press.
610 Culver DC, Wilkens H. 2000. Critical review of the relevant theories of the evolution of
611 subterranean animals. Ecosystems of the World. 381–398.
612 Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics
613 and parallel computing. Nat Methods. 9:772.

614 Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees.
615 BMC Evol Biol. 7:214.
616 Dupont LM, Donner B, Schneider R, Wefer G. 2001. Mid-Pleistocene environmental change in
617 tropical Africa began as early as 1.05 Ma. Geology 29:195-8.
618 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput.
619 Nucleic acids resear Espinasa L, Jeffery WR. 2006. Conservation of retinal circadian rhythms
620 during cavefish eye degeneration. Evol Dev. 8:16–22.
621 1792–1797.
622 Fischer RM, et al. 2013. Co-expression of VAL-and TMT-opsins uncovers ancient photosensory
623 interneurons and motorneurons in the vertebrate brain. PLoS Biol. 11:e1001585.
624 Friedman JS, et al. 2006. Premature truncation of a novel protein, RD3, exhibiting subnuclear
625 localization is associated with retinal degeneration. Am J Hum Genet. 79:1059–1070.
626 Galtier N, Gouy M, Gautier C. 1996. SEAVIEW and PHYLO_WIN: two graphic tools for
627 sequence alignment and molecular phylogeny. Bioinformatics. 12:543–548.
628 Gross JB, Wilkens H. 2013. Albinism in phylogenetically and geographically distinct populations
629 of Astyanax cavefish arises through the same loss-of-function Oca2 allele. Heredity. 111:122.
630 Gross JB, Borowsky R, Tabin CJ. 2009. A novel role for Mc1r in the parallel evolution of
631 depigmentation in independent populations of the cavefish Astyanax mexicanus. PLoS Genet.
632 5:e1000326.
633 Gross JB, Furterer A, Carlson BM, Stahl BA. 2013. An integrated transcriptome-wide analysis of
634 cave and surface dwelling Astyanax mexicanus. PLoS One. 8:e55659.
635 Guindon S, et al. 2010. New algorithms and methods to estimate maximum-likelihood
636 phylogenies: assessing the performance of PhyML 3.0. Systematic biology. 59:307–321.

637 Hinaux H, et al. 2013. De novo sequencing of Astyanax mexicanus surface fish and Pachón
638 cavefish transcriptomes reveals enrichment of mutations in cavefish putative eye genes. PloS
639 One. 8:e53553.
640 Howe DG, et al. 2012. ZFIN, the Zebrafish Model Organism Database: increased support for
641 mutants and transgenics. Nucleic Acids Res. 41:D854–D860.
642 Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ. 2010. Vertebrate pigmentation: from
643 underlying genes to adaptive function. Trends Genet. 26:231–239.
644 Irisarri I, et al. 2018. Phylogenomics uncovers early hybridization and adaptive loci shaping the
645 radiation of Lake Tanganyika cichlid fishes. Nat Commun. 9:1–12.
646 Jeffery WR. 2005. Adaptive evolution of eye degeneration in the Mexican blind cavefish. J
647 Heredity. 96:185–196.
648 Kawauchi H. 2006. Functions of melanin-concentrating hormone in fish. J Exp Zool Part A.
649 305:751–760.
650 Kim EB, et al. 2011. Genome sequencing reveals insights into physiology and longevity of the
651 naked mole rat. Nature. 479:223.
652 Klaassen H, Wang Y, Adamski K, Rohner N, Kowalko JE. 2018. CRISPR mutagenesis confirms
653 the role of oca2 in melanin pigmentation in Astyanax mexicanus. Dev Biol. 441:313–318.
654 Klaus S, et al. 2013. Rapid evolution of troglomorphic characters suggests selection rather than
655 neutral mutation as a driver of eye reduction in cave crabs. Biol Lett. 9:20121098.
656 Kolodziejski PA, et al. 2018. Serum levels of spexin and kisspeptin negatively correlate with
657 obesity and insulin resistance in women. Physiol Res. 67:45–56.
658 Kozak M. 1984. Point mutations close to the AUG initiator codon affect the efficiency of
659 translation of rat preproinsulin in vivo. Nature 308:241–246.

660 Kratochwil CF, Urban S, Meyer A. Genome of the Malawi golden cichlid fish (Melanochromis
661 auratus) reveals exon loss of oca2 in an amelanistic morph. Pigm Cell Melanoma R. 0:1-5.
662 doi: 10.1111/pcmr.12799.
663 Kumar S, et al. 2016. Decreased Circulating Levels of Spexin in Obese Children. J Clinic Endocrin
664 Metab. 101:2931–2936.
665 Langecker TG, Schmale H, Wilkens H. 1993. Transcription of the opsin gene in degenerate eyes
666 of cave-dwelling Astyanax fasciatus (Teleostei, Characidae) and of its conspecific epigean
667 ancestor during early ontogeny. Cell Tissue Res 273:183-192.
668 Lartillot N, Poujol R. 2011. A Phylogenetic Model for Investigating Correlated Evolution of
669 Substitution Rates and Continuous Phenotypic Characters. Mol Biol Evol. 28:729–744.
670 Lee J, Gross JM. 2007. Laminin β1 and γ1 Containing Laminins Are Essential for Basement
671 Membrane Integrity in the Zebrafish Eye. Invest Ophthalmol Vis Sci. 48:2483–2490.
672 Leys Remko, Cooper Steven J.B, Strecker Ulrike, Wilkens Horst. 2005. Regressive evolution of
673 an eye pigment gene in independently evolved eyeless subterranean diving beetles. Biol Lett.
674 1:496–499.
675 Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
676 arXiv:1303.3997 [q-bio]. http://arxiv.org/abs/1303.3997 (Accessed August 12, 2019).
677 Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform.
678 Bioinformatics. 25:1754–1760.
679 Liu Y, et al. 2010. Convergent sequence evolution between echolocating bats and dolphins. Curr
680 Biol. 20:R53–R54.
681 Liu YP, et al. 2017. Putative digenic inheritance of heterozygous RP1L1 and C2orf71 null
682 mutations in syndromic retinal dystrophy. Ophthal Genet. 38:127–132.

683 Malinsky M, et al. 2018. Whole-genome sequences of Malawi cichlids reveal multiple radiations
684 interconnected by gene flow. Nat Ecol Evol. 2:1940–1955.
685 Marcovitz A, Jia R, Bejerano G. 2016. “Reverse Genomics” Predicts Function of Human
686 Conserved Noncoding Elements. Mol Biol Evol. 33:1358–1369.
687 Matsuda K, et al. 2006. Central administration of melanin-concentrating hormone (MCH)
688 suppresses food intake, but not locomotor activity, in the goldfish, Carassius auratus. Neurosci
689 Lett. 399:259–263.
690 Matsuda K. 2009. Recent Advances in the Regulation of Feeding Behavior by Neuropeptides in
691 Fish. Ann NY Acad Sci. 1163:241–250.
692 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads.
693 EMBnet.journal. 17:10–12.
694 Marco-Sola S, Sammeth M, Guigó R, Ribeca P. 2012. The GEM mapper: fast, accurate and
695 versatile alignment by filtration. Nat Methods. 9:1185–1188.
696 Matos-Cruz V, et al. 2011. Unexpected Diversity and Photoperiod Dependence of the Zebrafish
697 Melanopsin System. PLOS ONE. 6:e25111.
698 McKenna A, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing
699 next-generation DNA sequencing data. Genome Res. 20:1297–1303.
700 Meunier I, et al. 2014. Frequency and Clinical Pattern of Vitelliform Macular Dystrophy Caused
701 by Mutations of Interphotoreceptor Matrix IMPG1 and IMPG2 Genes. Ophthalmology.
702 121:2406–2414.
703 Moran D, Softley R, Warrant EJ. 2015. The energetic cost of vision and the evolution of eyeless
704 Mexican cavefish. Sci Advances. 1:e1500363.

705 Near TJ, et al. 2012. Resolution of ray-finned fish phylogeny and timing of diversification. PNAS
706 USA. 109:13698–13703.
707 Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and
708 nonsynonymous nucleotide substitutions. Mol Biol Evol. 3:418–426.
709 Nelson SM, Park L, Stenkamp DL. 2009. Retinal homeobox 1 is required for retinal neurogenesis
710 and photoreceptor differentiation in embryonic zebrafish. Dev Biol. 328:24–39.
711 Niemiller ML, Fitzpatrick BM, Shah P, Schmitz L, Near TJ. 2013. Evidence for Repeated Loss of
712 Selective Constraint in Rhodopsin of Amblyopsid Cavefishes (Teleostei: Amblyopsidae).
713 Evolution. 67:732–748.
714 Niemiller ML, Soares D. 2015. Cave Environments. In: Extremophile Fishes: Ecology, Evolution,
715 and Physiology of Teleosts in Extreme Environments. Riesch, R, Tobler, M, & Plath, M,
716 editors. Springer International Publishing: Cham pp. 161–191.
717 Peter BD. 2004. African climate change and faunal evolution during the Pliocene–Pleistocene.
718 Earth Planet Sci Lett. 220:3-24.
719 Poulson TL, White WB. 1969. The Cave Environment. Science. 165:971–981.
720 Protas M, et al. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution
721 of albinism. Nat Genet. 38:107–111.
722 Protas M, Conrad M, Gross JB, Tabin C, Borowsky R. 2007. Regressive Evolution in the Mexican
723 Cave Tetra, *Astyanax mexicanus*. Curr Biol. 17:452–454.
724 Protas M, et al. 2008. Multi-trait evolution in a cave fish, *Astyanax mexicanus*. Evol Dev. 10:196–
725 209.
726 R Core Team. R: A language and environment for statistical computing. R Foundation for
727 Statistical Computing, Vienna, Austria. URL: https://www.R-project.org/ (2018).

728 Rambaut A, Suchard MA, Xie D, Drummond AJ. 2014. Tracer v1.6,
729 http://beast.bio.ed.ac.uk/Tracer.
730 Scrima A, et al. 2008. Structural Basis of UV DNA-Damage Recognition by the DDB1–DDB2
731 Complex. Cell. 135:1213–1223.
732 Smith SA, Brown JW, Walker JF. 2018. So many genes, so little time: A practical approach to
733 divergence-time estimation in the genomic era. PLOS ONE. 13:e0197433.
734 Song L, Florea L, Langmead B. 2014. Lighter: fast and memory-efficient sequencing error
735 correction without counting. Genome Biol. 15:509.
736 Stern DB, Crandall KA. 2018. The Evolution of Gene Expression Underlying Vision Loss in Cave
737 Animals. Mol Biol Evol. 35:2005–2014.
738 Stiassny MLJ, Alter SE. In press. Evolution in the fast lane: diversity, ecology, and speciation of
739 cichlids in the lower Congo River. In: The Behavior, Ecology and Evolution of Cichlid Fishes.
740 Springer Publishing.
741 Strickler AG, Famuditimi K, Jeffery WR. 2002. Retinal homeobox genes and the role of cell
742 proliferation in cavefish eye degeneration. Int J Dev Biol. 46:285–294.
743 Tang J, Chu G. 2002. Xeroderma pigmentosum complementation group E and UV-damaged
744 DNA-binding protein. DNA Repair. 1:601–616.
745 Tavaré S. 1986. Some probabilistic and statistical problems in the analysis of DNA sequences.
746 Lectures Math. Life Sci. 17:57-86.
747 Walewski JL, et al. 2014. Spexin is a novel human peptide that reduces adipocyte uptake of long
748 chain fatty acids and causes weight loss in rodents with diet-induced obesity. Obesity.
749 22:1643–1652.

Wang K, et al. 2019. Morphology and genome of a snailfish from the Mariana Trench provide insights into deep-sea adaptation. Nat Ecol Evol. 3:823–833.

Wessinger CA, Rausher MD. 2015. Ecological Transition Predictably Associated with Gene Degeneration. Mol Biol Evol. 32:347–354.

Wilkens H. 1971. Genetic Interpretation of Regressive Evolutionary Processes: Studies on Hybrid Eyes of Two Astyanax Cave Populations (Characidae, Pisces). Evolution. 25:530–544.

Wilkens H. 1988. Evolution and Genetics of Epigean and Cave Astyanax fasciatus (Characidae, Pisces). In: Evolutionary Biology: Volume 23. Hecht, MK & Wallace, B, editors. Evolutionary Biology Springer US: Boston, MA pp. 271–367.

Womack MC, Fiero TS, Hoke KL. 2018. Trait independence primes convergent trait loss. Evolution 72:679-687.

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. 330:200-211.

Yang Z. 2007. PAML4: Phylogenetic analysis by maximum likelihood. Mol Biol Evol. 24:1586-1591.

Zhao H, et al. 2018. Modulation of DNA repair systems in blind cavefish during evolution in constant darkness. Curr Biol. 28:3229-3243.

Zheng B, et al. 2017. Spexin suppress food intake in zebrafish: evidence from gene knockout study. Sci Rep. 7:14643.

Acknowledgements. We gratefully acknowledge assistance from Oliver Lucanus, Tobit Liyandja and Raoul Monsembula in obtaining specimens. Funding was provided by NSF/DEB Award 1655694 to SEA and MLJS, and PSC-CUNY Award 41. MLA was supported in part by a Gerstner Fellowship in Bioinformatics and Computational Biology to the American Museum of Natural History.

Author Contributions. MLJS and SEA conceived the study. MLA, MLJS and SEA designed the study. SEA performed laboratory work. MLA conducted the genome assembly and bioinformatics analyses. MLJS conducted the morphological comparisons. All authors prepared, edited and approved the manuscript.

Figures

Figure 1. Distribution of *L. lethops* and its loss of pigmentation. a, Map of the lower Congo River with geographical location indicator inset on left. Known distribution of *Lamprologus 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 to 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.

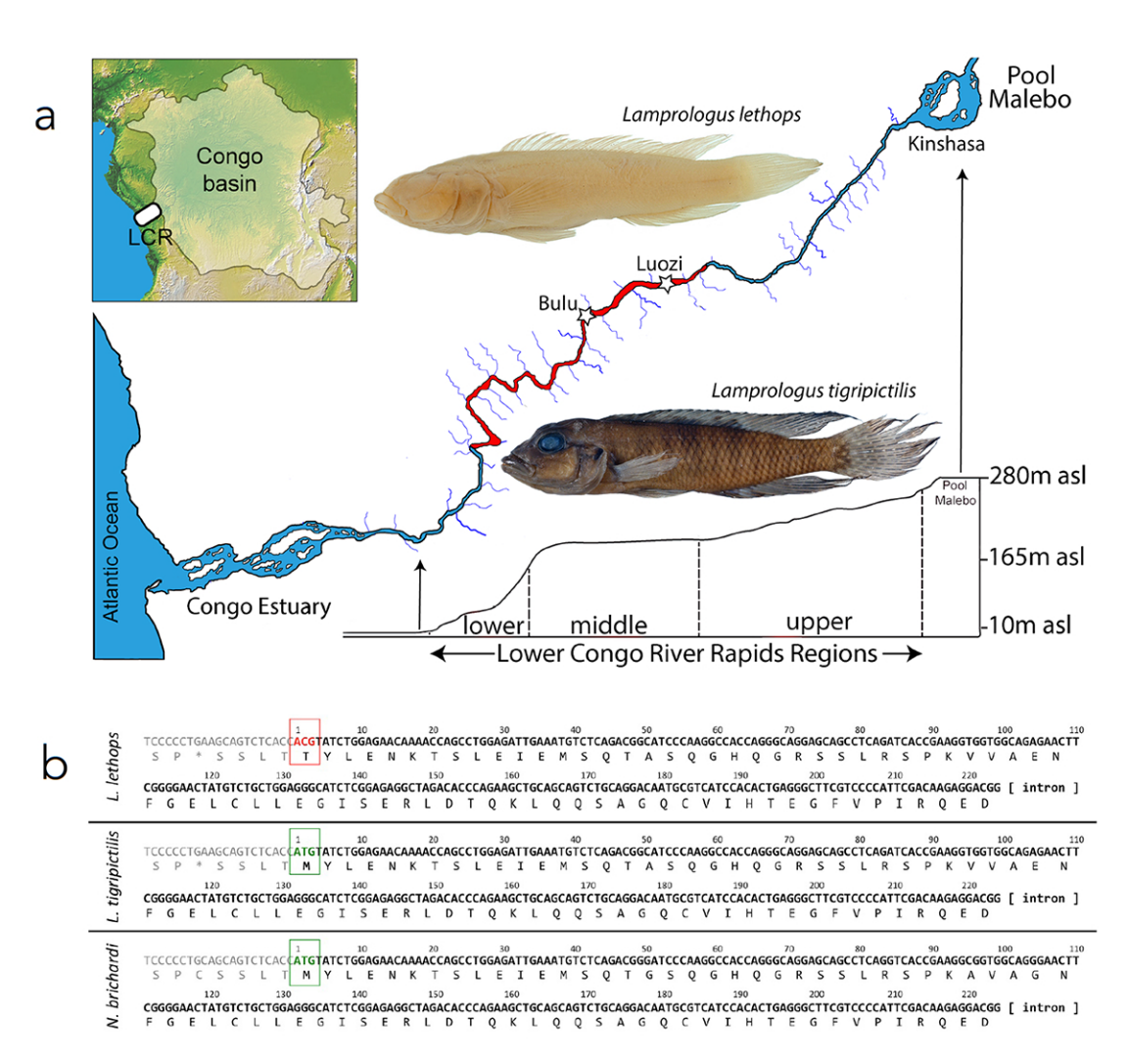
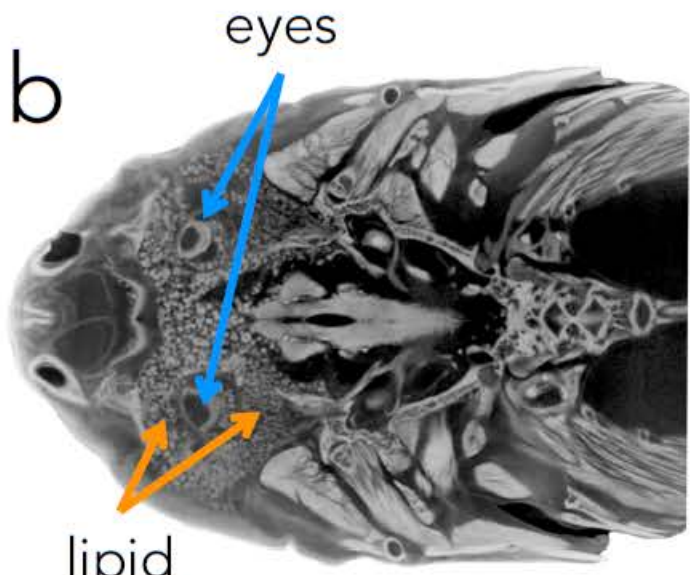

a
Congo basin
LCR
Lamprologus lethops
Pool Malebo
Kinshasa
Luozi
Bulu
Lamprologus tigripictilis
Atlantic Ocean
Congo Estuary
Pool Malebo
280m asl
165m asl
10m asl
lower
middle
upper
Lower Congo River Rapids Regions
b
L. lethops
TCCCCCTGAAGCAGTCTCACCACGTATCTGGAGAACAAAACCAGCCTGGAGATTGAAATGTCTCAGACGGCATCCCAAGGCCACCAGGGCAGGAGCAGCCTCAGATCACCGAAGGTGGTGGCAGAGAACTT
S P * S S L T T Y L E N K T S L E I E M S Q T A S Q G H Q G R S S L R S P K V V A E N
CGGGGAACTATGTCTGCTGGAGGGCATCTCGGAGAGGCTAGACACCCAGAAGCTGCAGCAGTCTGCAGGACAATGCGTCATCCACACTGAGGGCTTCGTCCCCATTCGACAAGAGGACGG [intron]
F G E L C L L E G I S E R L D T Q K L Q Q S A G Q C V I H T E G F V P I R Q E D
L. tigripictilis
TCCCCCTGAAGCAGTCTCACCATGTATCTGGAGAACAAAACCAGCCTGGAGATTGAAATGTCTCAGACGGCATCCCAAGGCCACCAGGGCAGGAGCAGCCTCAGATCACCGAAGGTGGTGGCAGAGAACTT
S P * S S L T M Y L E N K T S L E I E M S Q T A S Q G H Q G R S S L R S P K V V A E N
CGGGGAACTATGTCTGCTGGAGGGCATCTCGGAGAGGCTAGACACCCAGAAGCTGCAGCAGTCTGCAGGACAATGCGTCATCCACACTGAGGGCTTCGTCCCCATTCGACAAGAGGACGG [intron]
F G E L C L L E G I S E R L D T Q K L Q Q S A G Q C V I H T E G F V P I R Q E D
N. brichardi
TCCCCCTGCAGCAGTCTCACCATGTATCTGGAGAACAAAACCAGCCTGGAGATTGAAATGTCTCAGACGGGATCCCAAGGCCACCAGGGCAGGAGCAGCCTCAGGTCACCGAAGGCGGTGGCAGGGAACTT
S P C S S L T M Y L E N K T S L E I E M S Q T G S Q G H Q G R S S L R S P K A V A G N
CGGGGAACTATGTCTGCTGGAGGGCATCTCGGAGAGGCTAGACACCCAGAAGCTGCAGCAGTCTGCAGGACAATGCGTCATCCACACTGAGGGCTTCGTCCCCATTCGACAAGAGGACGG [intron]
F G E L C L L E G I S E R L D T Q K L Q Q S A G Q C V I H T E G F V P I R Q E D

Figure 2. Eye morphology and genes affecting vision. a, external morphology and **b,** transverse histological sections through cranium and eyes (CT scan after incubation in phosphotungstic acid) of *L. lethops* (the location of the degenerate eyes and high concentrations of fat globules are indicated by the arrows), and **c,** external morphology and **d,** transverse histological sections through cranium and eyes (CT scan after incubation in phosphotungstic acid) of *L. tigripictilis* (the location 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 S2.

a

lateral view

b

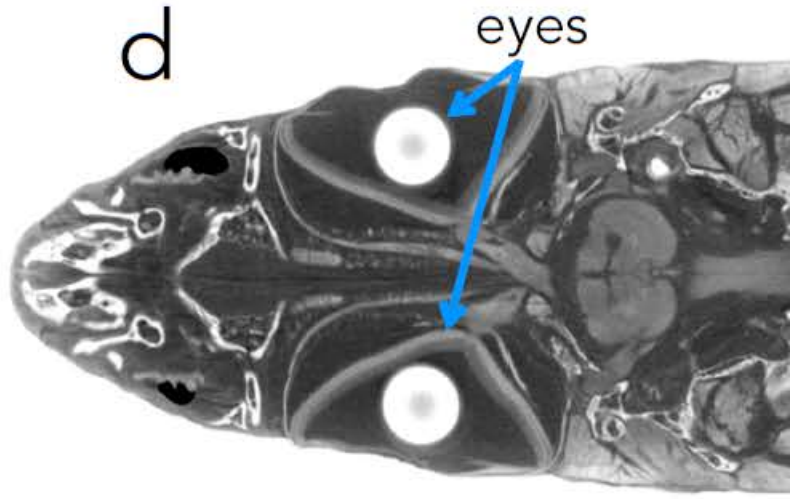


dorsal view

c

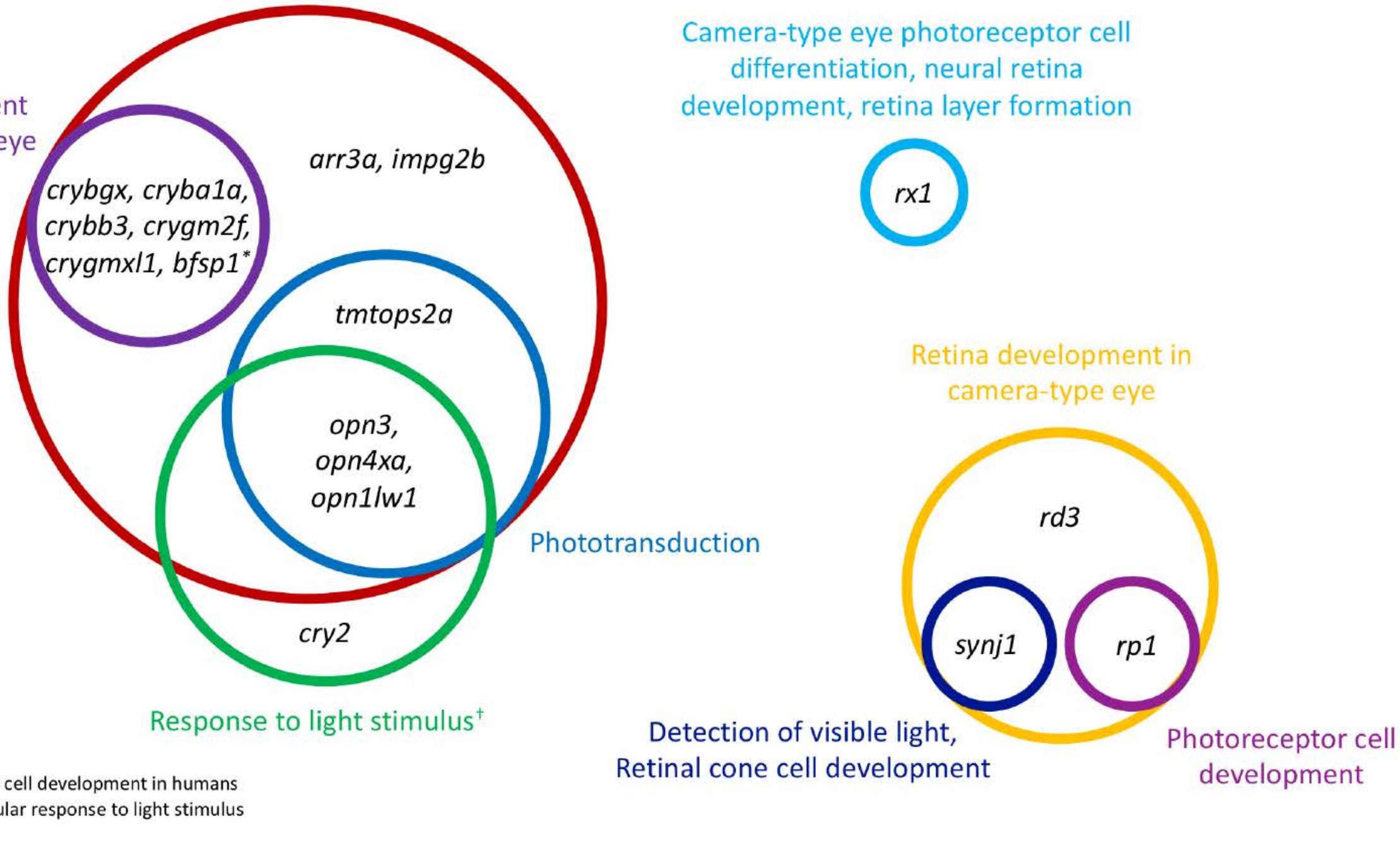
lateral view

d



dorsal view

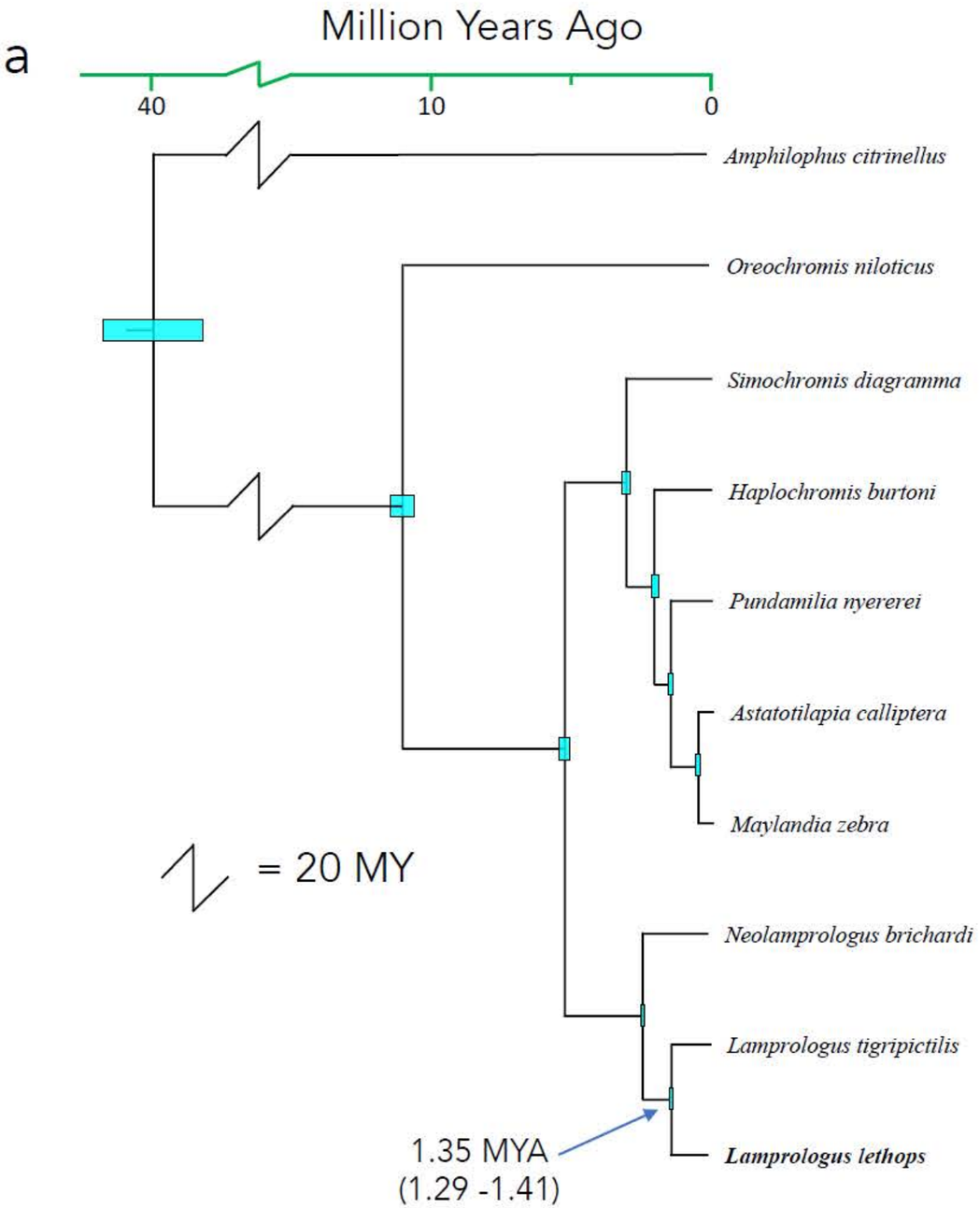
e



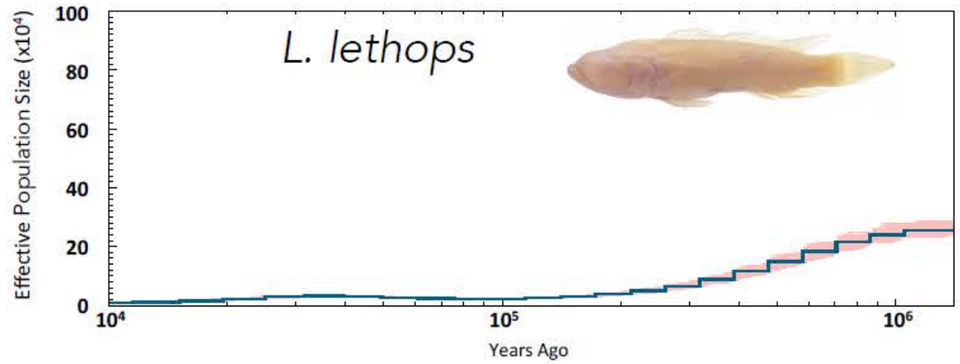
* Implicated in lens fiber cell development in humans
† Includes GO term: cellular response to light stimulus

Figure 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. Pairwise sequentially Markovian coalescent (PSMC) analysis of effective population size changes in **(b)** *L. lethops* and **(c)** *L. tigripictilis* over the last 1.3 million years, indicating their different demographic histories after species divergence. **d,** A comparison if scaffold-wide chromosome levels (for all scaffolds larger than 10Kb in the reference genome, *N. 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 non-synonymous versus synonymous evolutionary changes (as dN/dS) 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 dN/dS ratio in *L. tigripictilis* then in *L. lethops.*

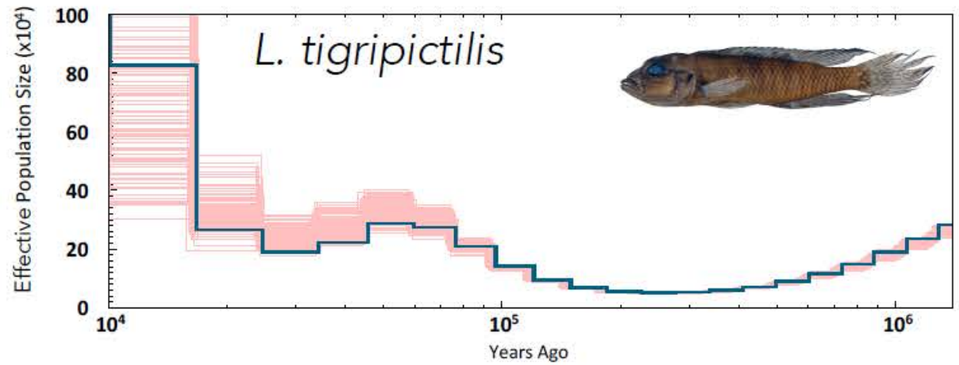
a


Million Years Ago
40
10
0
Amphilophus citrinellus
Oreochromis niloticus
Simochromis diagramma
Haplochromis burtoni
Pundamilia nyererei
Astatotilapia calliptera
Maylandia zebra
= 20 MY
Neolamprologus brichardi
Lamprologus tigripictilis
1.35 MYA
(1.29 -1.41)
Lamprologus lethops

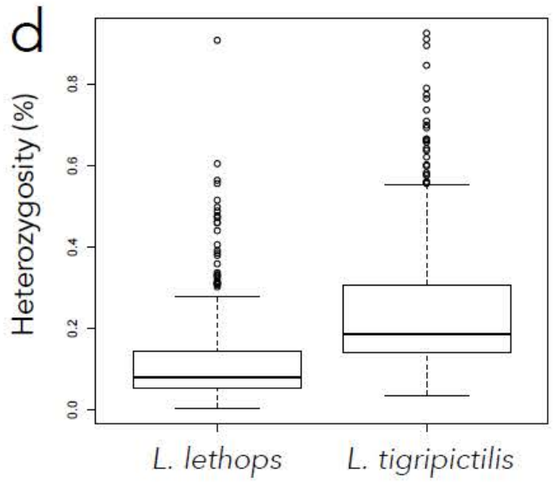
b


L. lethops
Effective Population Size (x10⁴)
100
80
60
40
20
0
10⁴
10⁵
10⁶
Years Ago

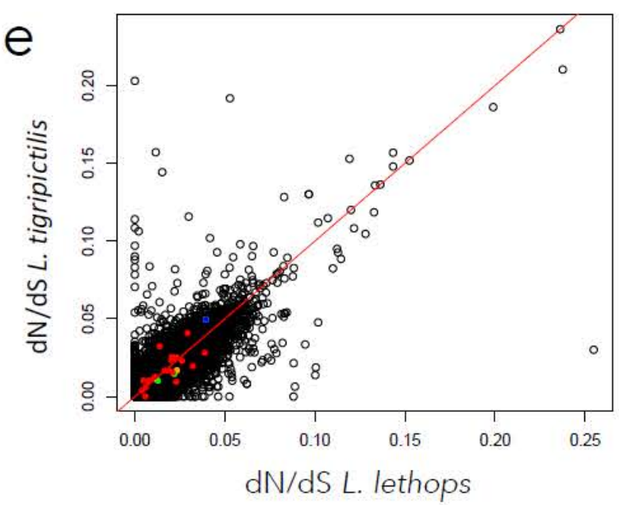
c


L. tigripictilis
Effective Population Size (x10⁴)
100
80
60
40
20
0
10⁴
10⁵
10⁶
Years Ago

d


Heterozygosity (%)
0.0
0.2
0.4
0.6
0.8
L. lethops
L. tigripictilis

e


dN/dS L. tigripictilis
0.00
0.05
0.10
0.15
0.20
0.25
dN/dS L. lethops

Supplementary Tables

Table S1. GO terms used to locate genes that potentially influence our phenotypes of interest: pigmentation, eye formation, metabolism/feeding, UV repair, and circadian rhythms.

Table S2. Genes found to contain inactivating mutations and that had GO terms associated with one or more of our phenotypes of interest (see Supplementary Table 1). Gene names and symbols come from *Danio rerio*. The type of inactivating mutation(s) observed in each gene, associated GO terms, associated gene in *Danio rerio* (from the Zfin database; Howe et al. 2012) are indicated. Finally, if the gene has been implicated in similar phenotypic changes in other vertebrate species (predominately the cavefish, *Astyanax mexicanus*), these references are given.

Table S3. The ratios of non-synonymous changes to synonymous changes (dN/dS) for the 22 genes found to contain inactivating mutations which may also influence important phenotypes in *L. lethops.* These ratios were determined in relation to *N. birchardi* using Maximum-likelihood methods as implemented in PAML (Yang 2007). Means and standard deviations for these 22 genes appear in the second to last row. The ratios were not significantly different between the two species (paired students t-test, t = -0.21812, df = 21, p-value = 0.8294). In the last row are the means and standard deviations for all other annotated genes in these two species.

Table S4. The results of our PAML analyses of the genes of interest. This analysis compared the rate of evolution (as dN/dS ratio) for the *L.* lethops branch of the tree to those of all the other cichlids to see if it was significantly different. Significance was determined by the likelihood ratio test. These genes are those in which confident alignment for the 10 cichlid species was possible

842 (these were the Cichlid species with available genomes at the time of the study). Gaps in the
843 sequences and regions with missing data were removed before the analyses. 'Codons analyzed'
844 indicates the number of codons that remained after gap removal. Only one gene, opsin 1, was
845 found to have a significant difference in evolutionary rates between *L. lethops* and the other nine
846 cichlids.
847
848 **Table S5.** Summary statistics for *de novo* genome assemblies of *L. lethops* and *L. tigripictilis*.
849
850 **Supplementary Figures**
851 **Figure S1. Morphological comparison of *L. lethops* and the closely related *L. tigripictilis***.
852 Transverse section through cranium and eyes (hematoxylin and eosin stained) of (**a)** *L. lethops*
853 compared with **(b)** *L. tigripictilis*. Laterosensory canals (outlined in white) and pores (orange
854 shading) of **(c)** *L. lethops* and **(d)** *L. tigripictilis*. The cleared and stained dorsal view showing the
855 extent of the gas bladder outlined in **(e)** *L. lethops* and **(f)** *L. tigripictilis*. Transverse section through
856 the gas bladder (CT scan after incubation in phosphotungstic acid) of **(g)** *L. lethops* and **(h)** *L.*
857 *tigripictilis.*
858
859 **Figure S2. Caudal fin of *L. lethops*.** Photograph of the caudal fin of a *L. lethops* specimen
860 showing the absence of visible melanophores or melanin.
861
862 **Figure S3.** Phylogenetic relationships of the ten cichlid species used in this study, showing the
863 95% Highest Posterior Density (HPD) intervals indicating estimated divergence time at each node.
864

Figure S4. The observed distributions of single nucleotide variant (SNVs) quality measures in our genomes. The distributions in our four-fold degenerate sites dataset were nearly identical (not shown). The red triangles indicate the filtering thresholds chosen. SNPs that had a quality by depth less than 2 (QD < 2.0), Fisher strand bias greater than 40 (FS > 40.0), mapping quality less than 55 (MQ < 55.0), mapping quality rank sum less than -0.2 (MQRankSum < -0.2), read position rank sum less than -2 (ReadPosRankSum < -2.0), and a strand odds ratio greater than 3 (SOR > 3.0), were all removed from our datasets.

Figure S5. The observed distributions of single nucleotide variant (SNVs) quality measures in our genomes. The distributions in our four-fold degenerate sites dataset were nearly identical (not shown). The red triangles indicate the filtering thresholds chosen. SNPs that had a quality by depth less than 2 (QD < 2.0), Fisher strand bias greater than 40 (FS > 40.0), mapping quality less than 55 (MQ < 55.0), mapping quality rank sum less than -0.2 (MQRankSum < -0.2), read position rank sum less than -2 (ReadPosRankSum < -2.0), and a strand odds ratio greater than 3 (SOR > 3.0), were all removed from our datasets.

Figure S6. Summary of the results from the BUSCO analysis of the ten cichlid genomes used in this study. Complete and single copy genes were used in our phylogenetic analysis and estimation of divergence times. This figure was produced using the python script 'generate_plot.py', available with the BUSCO package.