

# 1 Evolution of diapause in the African turquoise killifish by 2 remodeling ancient gene regulatory landscape

3  
4  
5 Param Priya Singh<sup>1,5</sup>, G. Adam Reeves<sup>1,5</sup>, Kévin Contrepois<sup>1,2</sup>, Mathew Ellenberger<sup>1</sup>, Chi-Kuo  
6 Hu<sup>1</sup>, Michael P. Snyder<sup>1,2,3</sup>, Anne Brunet<sup>1,4,6</sup>

7  
8 <sup>1</sup>Department of Genetics, Stanford University, Stanford, CA, USA

9 <sup>2</sup>Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA

10 <sup>3</sup>Stanford Diabetes Research Center, Stanford University, Stanford, CA, USA

11 <sup>4</sup>Glenn Laboratories for the Biology of Aging, Stanford University, Stanford, CA, USA

12 <sup>5</sup>These authors contributed equally

13 <sup>6</sup>Corresponding author: [abrunet1@stanford.edu](mailto:abrunet1@stanford.edu)

## 14 15 **ABSTRACT**

16 Suspended animation states such as hibernation or diapause allow organisms to survive extreme  
17 environments. But the mechanisms underlying the evolution of these extreme survival states are  
18 unknown. The African turquoise killifish has evolved diapause as a form of suspended  
19 development to survive the complete drought that occurs every year in its habitat. Here we show  
20 that many gene duplicates – paralogs – exhibit specialized expression in diapause versus normal  
21 development in the African turquoise killifish. Surprisingly, paralogs with specialized expression  
22 in diapause are evolutionarily very ancient, and they are also present even in vertebrates that do  
23 not exhibit diapause. Profiling the chromatin accessibility landscape among different fish species  
24 reveals an evolutionarily recent increase in chromatin accessibility at these very ancient paralogs,  
25 suggesting rewiring of their regulatory landscape. The increase in chromatin accessibility in the  
26 African turquoise killifish is linked to the presence of new binding sites for transcription factors  
27 (e.g., FOXO, REST, and PPAR), due to both de novo mutations and transposable element  
28 insertion. Interestingly, accessible chromatin regions in diapause are enriched for lipid metabolism  
29 genes. By performing lipidomics in different fish species, we uncover a specific lipid profile in  
30 African turquoise killifish embryos in diapause. Notably, select very long-chain fatty acids are

31 high in diapause, suggesting they may be used for long-term survival in this state. Together, our  
32 multi-omic analysis indicates that diapause is driven by regulatory innovation of very ancient gene  
33 programs that are critical for survival. Our work also suggests a mechanism for how complex  
34 adaptations evolve in nature and offers strategies by which a suspended animation program could  
35 be reactivated in other species for long-term preservation.

36

37

## 38 MAIN TEXT

### 39 Introduction

40 Extremophiles – species that live in extreme environments – have evolved unique adaptations for  
41 survival. Understanding how extreme adaptations evolve can reveal new pathways with important  
42 ramifications for survival in all organisms. The African turquoise killifish *Nothobranchius furzeri*  
43 is an extremophile for embryo survival. This vertebrate species lives in ephemeral ponds in  
44 Zimbabwe and Mozambique that completely dry up for ~8 months each year (1). To survive this  
45 annual drought, the African turquoise killifish has evolved two key adaptations: a rapid life to  
46 successfully reproduce during the rainy season and a form of long suspended animation, with  
47 embryos entering diapause and subsisting in the mud during the dry season (2-5). Diapause  
48 embryos survive for months even years – longer than adult life – without any detectable tradeoff  
49 for future life (6). Remarkably, diapause embryos already have complex organs and tissues,  
50 including a developing brain and heart (6). Hence, diapause provides a unique form of long-term  
51 protection to a complex organism.

52 Like other suspended animation states (hibernation, torpor), diapause is a multifaceted and  
53 active adaptation. Diapause also exists in other vertebrate species, including mammals (e.g., bear,  
54 roe deer, mice) (7). As diapause is extreme in the African turquoise killifish, this species provides  
55 a model to understand the mechanism and evolution of this suspended animation trait in  
56 vertebrates. Many genes involved in chromatin remodeling, metabolism and stress resistance are  
57 upregulated in killifish (6, 8, 9). Yet, how such an extreme and coordinated program evolved in  
58 nature is unknown. Using the lens of evolution to understand diapause could uncover new  
59 protective mechanisms for long-term survival and offer a framework for the evolution of extreme  
60 adaptations in nature.

## 61 **Paralogs that specialize for expression in diapause are evolutionarily very ancient**

62 We asked when the genes expressed in diapause originated in evolutionary time. To this end, we  
63 focused on paralogs – duplicated copies of genes (*10, 11*). Paralogs are the primary mechanism by  
64 which new genes originate and specialize for new functions or states (*12, 13*) (Fig. 1A). Paralogs  
65 also allow for a precise timing of the evolutionary origin of specific genes and they could help  
66 explain how the killifish genome can support two seemingly antagonistic traits – rapid life vs.  
67 suspended animation. Using phylogenetic inference (see Methods), we find that the African  
68 turquoise killifish genome contains 20,091 paralog pairs. We used our previously generated RNA-  
69 seq datasets of development and diapause in the African turquoise killifish (*6*) to analyze if the  
70 expression pattern of paralogs has diverged in diapause versus normal development states.  
71 Interestingly, many paralog pairs show opposing expression, with one gene in the paralog pair  
72 highly expressed in diapause (*'diapause-specialized gene'* e.g., the chromatin modifier *EZH1*) and  
73 the other gene in the paralog pair highly expressed in development (*'development-specialized*  
74 *gene'* e.g., the chromatin modifier *EZH2*) (Fig. 1B and fig. S1). Overall, 6,247 paralog pairs show  
75 expression specialization in diapause versus development state (Fig. 1C and fig. S1, Data Files S1  
76 and S2).

77 We next asked whether paralogs that exhibit expression specialization in diapause are  
78 evolutionarily recent or ancient. Diapause in the African turquoise killifish is a relatively recent  
79 specialization that evolved less than 18 million years (MY) ago (*14*). To date the paralogs, we  
80 generated a paralog classification pipeline to identify the evolutionary time when each of the  
81 African turquoise killifish paralogs originate compared to other species (Fig. 1D and fig. S2) (*15*).  
82 We distinguished i) very ancient paralogs (shared with all vertebrates, including mammals) that  
83 originated more than 473 MY ago, ii) ancient paralogs (shared with all other fish) that originated  
84 between ~111-473 MY ago, and iii) recent/very recent paralogs (killifish/African turquoise  
85 killifish specific) that originated less than ~111 MY ago (Fig. 1D). Surprisingly, very ancient  
86 paralogs were significantly more likely to specialize for diapause compared to the genome-wide  
87 average, even though diapause originated recently (Fig. 1E). In contrast, ancient and especially  
88 recent/very recent (killifish-specific) paralogs were significantly less likely to specialize for  
89 diapause compared the genome-wide average, even though they originated around the time when  
90 diapause evolved (Fig. 1E). Consistently, paralogs that originate from very ancient vertebrate-  
91 specific whole genome duplication or ancient small-scale duplications were more likely to

92 specialize for diapause (fig. S3). The enrichment for very ancient paralog pairs for specialization  
93 in diapause was robust to varying outgroups, phylogeny, method to identify paralogs, and paralog  
94 family size (fig. S4, A to E). As a specificity control, such an enrichment was not observed for  
95 paralogs that are expressed at the same level during development and diapause (in fact, those  
96 exhibited a depletion for very ancient paralogs) (fig. S4F). The genes specialized for expression in  
97 diapause did not exhibit increased positive selection at the protein level, raising the possibility of  
98 regulatory rewiring (fig. S4G). Thus, very ancient paralogs are co-opted in diapause, suggesting  
99 that ancestral programs are harnessed for this suspended animation state – perhaps by remodeling  
100 of the regulatory landscape.

101

### 102 **Very ancient paralogs also specialize in diapause in other killifish species with diapause**

103 Many killifish species populate the world, and their ability to undergo diapause is linked to their  
104 environment. Killifish species that live in ephemeral ponds exhibit diapause (e.g., African  
105 turquoise killifish, South American killifish), whereas killifish species that live in constant water  
106 do not undergo diapause, and instead they continuously develop (e.g., red-striped killifish and  
107 lyretail killifish, both of which are from Africa) (16-19) (Fig. 2A). To assess whether the  
108 specialization of ancient paralogs in diapause is generalizable to other species that evolved  
109 diapause independently, we used available RNA-seq data from diapause and development in the  
110 South American killifish with diapause, *Austrofundulus limnaeus* (9). We also generated new  
111 RNA-seq data from the developing embryos of the red-striped killifish *Aphyosemion striatum* and  
112 the lyretail killifish *Aphyosemion australe* – the closest relatives of the African turquoise killifish  
113 *N. furzeri* but without diapause (Fig. 2A). We found that in the South American killifish, paralogs  
114 also showed specialized expression in diapause versus development (Fig. 2, B and C, and fig. S5),  
115 and that their specialized expression correlated with that of paralogs in the African turquoise  
116 killifish (Fig. 2D and fig. S6, A and B). In contrast, killifish species without diapause expressed  
117 both paralogs during development (fig. S6C). Importantly, paralogs with specific expression in  
118 diapause in the South American killifish were also enriched for very ancient gene duplicates (Fig.  
119 2E). Collectively, these results indicate that very ancient paralog pairs have been repeatedly co-  
120 opted for specialized expression in diapause during evolution. Together, our observations also  
121 raise the possibility that the recent specialization for diapause is driven at least in part by regulatory  
122 innovation of very ancient paralog pairs.

## 123 **Evolutionarily recent remodeling of the chromatin landscape at very ancient paralogs**

124 To characterize the regulatory landscape of the paralogs that specialize in diapause during  
125 evolution, we profiled the chromatin accessibility landscape in different species of killifish. We  
126 performed ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing), which  
127 assesses chromatin accessibility genome-wide (20), on embryos during diapause and development  
128 in killifish species with diapause (African turquoise killifish, South American killifish) and  
129 embryos during development in killifish species without diapause (lyretail killifish and red-striped  
130 killifish) at a similar developmental stage (Fig. 3A). We also used available ATAC-seq data for  
131 medaka and zebrafish development (21). We verified the quality of our ATAC-seq samples by  
132 transcription start site enrichment of open chromatin and fragment size periodicity (See Methods,  
133 fig. S8). Chromatin states easily separated diapause and development embryos by Principal  
134 Component Analysis (PCA) in the African turquoise killifish (Fig. 3B). Chromatin states also  
135 separated diapause and developmental samples of different killifish species (Fig. 3B). In the  
136 African turquoise killifish, 6,490 genomic regions were differentially accessible in diapause  
137 compared to development genome-wide (fig. S7A, Data Files S1 and S3), and they were located  
138 mostly in promoter, intronic, or distal intergenic (e.g., enhancer) regions (fig. S7B). There was a  
139 positive correlation between chromatin accessibility and gene expression levels in diapause in the  
140 African turquoise killifish (fig. S7, C and D). Together, these results indicate that our datasets for  
141 chromatin state landscape in diapause and development in several killifish species are of good  
142 quality.

143 We next examined accessible chromatin regions (ATAC-seq peaks) at paralogs that are  
144 differentially expressed in diapause versus development (Fig. 3C) (e.g., *DNAJA4* and *DNAJA2*,  
145 Fig. 3D and fig. S1B). As paralogs that specialize in diapause are very ancient (> 473 MY), we  
146 therefore asked when chromatin accessibility occurred in evolutionary time (Fig. 3C and fig. S9).  
147 To quantify chromatin accessibility over evolutionary time, we developed a pipeline to identify  
148 the relative evolutionary origin of ATAC-seq peak based on multi-genome alignment (see  
149 Methods), and classified each ATAC-seq peak as i) ancient/very ancient (i.e. chromatin accessible  
150 in all fish species evaluated, such as *CBX8*) (Fig. 3E and fig. S9), ii) recent (i.e. chromatin  
151 accessible only in killifish species, such as *HNRNPA3*), and very recent (chromatin accessible only  
152 in the African turquoise killifish, such as *OSBPL5*) (Fig. 3E and fig. S9). Interestingly, most  
153 regulatory regions of very ancient paralogs (> 473 MY) that are differentially regulated in diapause

154 exhibited chromatin accessibility very recently (~18 MY), only in the African turquoise killifish  
155 (Fig. 3C and fig. S10). The very recent chromatin accessibility at very ancient paralogs specialized  
156 in diapause was generalizable to non-paralog genes (fig. S10A) and was more pronounced at distal  
157 regulatory elements (likely enhancers) (fig. S10B). Thus, the African turquoise killifish exhibits  
158 an evolutionary recent remodeling of the chromatin accessibility landscape at very ancient genes.

159

### 160 **Mechanisms underlying the evolution of chromatin accessibility in diapause**

161 What are the mechanisms connecting evolutionary recent chromatin accessibility with diapause?  
162 Chromatin regions that opened recently in diapause paralogs in the African turquoise killifish were  
163 enriched for transcription factor motifs, including REST, NR2F2, Forkhead transcription factors  
164 (e.g., FOXA1, FOXO3), and PPAR (e.g., PPARA) (Fig. 4A). Most diapause-specific transcription  
165 factor binding motifs were only enriched in the African turquoise killifish, but not in other closely  
166 related fish without diapause (Fig. 4B and fig. S11). Thus, these transcription factor motifs arose  
167 very recently in the African turquoise killifish after divergence from other killifish species without  
168 diapause and could underlie the evolutionarily recent opening of chromatin at very ancient  
169 paralogs.

170 New transcription binding motifs can arise *de novo* by point mutation or transposable  
171 element (TE) insertion (22, 23) (Fig. 4C). A majority (81%) of the transcription factor binding  
172 motifs associated with diapause accessible chromatin at specialized paralogs in the African  
173 turquoise killifish evolved *de novo*, via mutation of the ancestral sequence (Fig. 4D). For example,  
174 transcription factor motifs (e.g., FOXO3 motifs) were canonical binding sites (as defined by  
175 HOMER (24)) in the African turquoise killifish sequence but were slightly divergent in closely  
176 related fish without diapause and even more divergent or even absent in more distant fish species  
177 (Fig. 4, E and F, and fig. S12). Importantly, we found a signature of positive selection (25) at many  
178 of the diapause-specific accessible chromatin regions in the African turquoise killifish, with  
179 enrichment for binding motifs for FOXO3, REST, and PPAR (Fig. 4G and fig. S13). Thus, the  
180 African turquoise killifish may have selected for canonical transcription factor binding motifs at  
181 regulatory regions of genes beneficial for diapause.

182 Intriguingly, some binding motifs associated with diapause in the African turquoise  
183 killifish paralogs overlapped with transposable elements and were unique to this species (Fig. 4D).  
184 While overlaps with transposable elements represents a minority of cases (5%) (Fig. 4D),

185 transposable elements can deliver a transcription factor binding motif to new regulatory sites faster  
186 than gradual mutation and selection. As transposable elements have exploded in the African  
187 turquoise killifish genome (26), they may represent an evolutionary mechanism to co-opt genes  
188 into the diapause expression program. Several transposable element families (e.g., DNA  
189 transposons and LINES) were highly enriched at accessible chromatin regions in diapause in the  
190 African turquoise killifish (Fig. 4H). In some cases, these regions contained both a transposable  
191 element and a transcription factor binding site (e.g., PPAR) (Fig. 4I and fig. S14). Hence,  
192 transcription factor binding motifs underlying diapause-specialized paralogs may have originated  
193 not only through mutation and selection but also via a recent burst, in the African turquoise  
194 killifish, of transposon-mediated reshuffling.

195

### 196 **Functional enrichment and lipidomics reveal specific lipids in the diapause state**

197 We asked whether ancient paralogs that are recently repurposed for diapause in the African  
198 turquoise killifish are associated with a specific biological function. Analysis of gene expression  
199 and chromatin accessibility at diapause-specific paralogs showed enrichment of several functions  
200 related to lipid metabolism (e.g., lipid storage, very long chain fatty acid metabolism and  
201 regulation of fatty acid beta oxidation) (Fig. 5A, Data Files S5 and S6). Both gene expression and  
202 chromatin accessibility datasets showed enrichment of upstream regulators of lipid metabolism  
203 (e.g., FOXO1 (27)) or transcriptional sensors of fatty acids (e.g., PPAR) (28) (Fig. 5B, Data File  
204 S7). These observations raise the possibility that lipids play an important role in diapause.

205 While some lipids and metabolites have been examined in killifish embryos and adults (29-  
206 31), a systematic profiling of lipids in diapause vs. development in killifish species with and  
207 without diapause has not been done. We therefore performed untargeted lipidomics on the African  
208 turquoise killifish embryos at different times: pre-diapause and diapause at different times (6 days  
209 and 1 month). As a comparison, we also performed lipidomics on embryos of another killifish  
210 species that does not undergo diapause (red-striped killifish) (Fig. 5C). The lipidome separated  
211 diapause from development in the African turquoise killifish and development in the red-striped  
212 killifish by PCA (Fig. 5D). Glycerophospholipids (e.g., phosphatidylcholines [PC]), which are  
213 membrane lipids, and triglycerides (TGs), which are storage lipids, were both changed in diapause  
214 in comparison to development stages (fig. S15A). The triglyceride changes in diapause are  
215 consistent with expression differences of genes, including specialized paralogs, encoding

216 triglyceride metabolism enzymes and regulators (fig. S15B). Interestingly, we observed an  
217 enrichment of TGs containing very long chain fatty acids (fatty acids with chain lengths of 22) in  
218 diapause compared to development, and the majority of these very long chain fatty acids have 5  
219 (docosapentaenoic acid, DPA) and 6 (docosahexaenoic acid, DHA) double bonds (Fig. 5, E and  
220 F). The same TGs with very long-chain fatty acids were also more abundant in African turquoise  
221 killifish embryos, even at pre-diapause, than in red-striped killifish at the equivalent state of  
222 development (Fig. 5G and fig. S15, C and D). As very long-chain fatty acids are processed by  
223 peroxisomes and subsequently by mitochondria to produce energy (32), they may serve as a long-  
224 term energy reserve for diapause. Other lipids, such as many ether-linked glycerophospholipids  
225 (plasmalogens), which can protect brain and hearts from oxidative stress (33-35), are more  
226 abundant in diapause than development and higher in the African turquoise killifish compared to  
227 the red-striped killifish (fig. S15D, Data File S8). Collectively, these data suggest that the African  
228 turquoise killifish has evolved to pack specific lipids, including long-chain fatty acids and  
229 membrane lipids with antioxidant properties, in their embryos. The rewiring of key transcription  
230 factor binding sites (e.g., FOXO1 or PPAR binding sites) at specialized paralogs (and other genes)  
231 involved in lipid metabolism could modulate lipid management for long-term protection and  
232 efficient storage and usage of specific fatty acids (Fig. 5H).

233

234

## 235 **Discussion**

236 Our study shows for the first time that although diapause evolved recently (less than 18 MY ago),  
237 the paralogs that specialized for diapause are ancestral and shared by most vertebrates (>473 MY  
238 old). This paralog specialization in the African turquoise killifish diapause is likely achieved by  
239 recent co-opting of conserved transcription factors (such as REST, FOXOs, and PPAR) and  
240 repurposing of their regulatory landscape by mutations and selection and transposon element  
241 insertion. Our multi-omics analysis of the diapause state (transcriptomics, chromatin states,  
242 lipidomics) and comparative analysis with several fish species suggests a model for diapause  
243 evolution via very ancient paralog-specialization. After duplication in the ancestor of most  
244 vertebrates, these very ancient paralogs likely specialize in the transient response to harsh  
245 environment (e.g., transient lack of food or temperature change, or other changes), which ensures  
246 their long-term maintenance in the genome. When the ancestors of African turquoise killifish



247 transitioned to ephemeral ponds over 18 million years (14), these paralogs evolved new  
248 transcription factor binding motifs driving further specialization, notably for lipid metabolism  
249 genes, for survival under extreme conditions in diapause (Fig 5H).

250 Elucidating the mechanisms underlying the origin of complex adaptations and phenotypes  
251 (e.g., ‘suspended animation’, novel cell types/tissues, etc.) is a central challenge of evolutionary  
252 biology (36). Gene duplication is the primary mechanism to generate new genes, and these act as  
253 substrate to evolve new functions. For example, ancient gene duplicates (paralogs) are specialized  
254 for expression in different tissues (37, 38). Ancient gene duplicates (paralogs) can also contribute  
255 to the evolution of new organs such as electric organ (39) and placenta (40). Gene duplicates have  
256 also been correlated with exceptional resistance to cancer in long-lived species (41-43). The  
257 specialization of paralogs might also explain how the African turquoise killifish genome can  
258 support two seemingly antagonistic complex traits – e.g., rapid life and long suspended animation.  
259 However, the mechanisms of how divergence of duplicated genes or paralogs contribute to the  
260 evolution of complex adaptations are still poorly understood. Emerging evidence, including our  
261 study, suggests that complex adaptations can arise by rewiring gene expression by unique  
262 regulatory elements (22, 23, 44, 45). Our results indicate that such a rewiring can be achieved  
263 using de novo regulatory elements and in some cases transposon insertion.

264 Cis-regulatory elements such as enhancers and promoters are known to evolve rapidly (46-  
265 49), and they can in turn facilitate complex adaptations with the same set of conserved genes.  
266 Transposon insertion can be even faster in promoting the rearrangement of regulatory regions (50-  
267 57). Rapid reshuffling of regulatory regions by mutation or transposon insertion provides a  
268 framework for the evolution of complex trait in nature. Such a mechanism could extend to the  
269 evolution of other complex traits, including regenerative capacity, which involves new enhancers  
270 in killifish (58), although other mechanisms may also contribute, including positive selection of  
271 specific genes (8, 9, 26, 59-61).

272 Our work also reveals specific genes and lipids that could be critical for long-term survival  
273 in suspended development. Lipids that accumulate in a state of ‘suspended animation’ (e.g., very  
274 long-chain fatty acids) could serve as key substrates for long-lasting survival (62, 63).  
275 Alternatively, they could also serve as new signals to affect specific aspects of the diapause state  
276 (in addition to known signals in other species, such as vitamin D or dafachronic acid in the South  
277 American killifish (64) and in *C. elegans*, respectively) (65, 66). The pathways and regulatory

278 mechanisms we identified could also apply to other states of suspended animation and even to  
279 adult longevity. For example, transcription factors whose motifs are enriched in the diapause state  
280 (e.g., FOXOs, PPARs) are genetically required for suspended animation states, such as *C. elegans*  
281 dauer (67-69), and are expressed in mammalian hibernation (70). Furthermore, lipids and lipid  
282 metabolism genes are expressed differentially in mammalian diapause (71) as well as hibernation  
283 and torpor (63, 72, 73), and are under positive selection in exceptionally long-lived mammals (74).  
284 Finally, several of the transcription factors we identified (e.g., FOXOs, REST) are genetically  
285 implicated in longevity in *C. elegans* and flies (75-80). Our results place these previous genetic  
286 and expression findings in a natural context and reveal how selective pressure can co-opt key  
287 metabolic programs to achieve extreme phenotypes. These observations also raise the possibility  
288 that a core program of lipid metabolism genes, regulated by specific transcription factors, can be  
289 deployed to achieve metabolic remodeling and stress resistance in diverse contexts, including in  
290 adults. Our study provides a new multi-omic resource for understanding the regulation and  
291 evolution of suspended animation states (hibernation, torpor, diapause). It also opens the  
292 possibility for strategies, including lipid-based interventions, to promote long-term tissue  
293 preservation and counter age-related diseases.

## 294 **ACKNOWLEDGMENTS**

295 We particularly thank J. Pritchard for invaluable discussion on paralogs and evolutionary analyses  
296 and for feedback on the manuscript. We thank J. Podrabsky and J. Wagner for providing *A.*  
297 *limnaeus* embryos and helpful suggestions with husbandry. We thank M. Robinson-Rechavi and  
298 J. Liu for help with positive selection analysis in ATAC-seq peaks. We thank J. Miklas, X. Zhao,  
299 K. Papsdorf, T. Ruetz, J. Chen, H. Fraser, G. Bejerano, S. Goenka, H. Chen, and all Brunet lab  
300 members for helpful discussion or comments. We thank C. Bedbrook, E. Sun, F. Boos, T. Ruetz,  
301 and X. Zhao for independent code-checking. Funding: This work was supported by the Glenn  
302 Foundation for Medical Research (A.B.), a fellowship from Stanford Center for Computational,  
303 Evolutionary, and Human Genomics (P.P.S.), and a National Science Foundation graduate  
304 fellowship (G.A.R.). Author contributions: P.P.S. designed the project with help from G.A.R. and  
305 A.B.. P.P.S. performed all the killifish experiments and computational analyses unless otherwise  
306 indicated. G.A.R. generated all ATAC-seq libraries, performed multi-genome integration of  
307 ATAC-seq and transposon analyses. K.C. and M.E. generated the lipidomics data and helped with  
308 the analysis under the supervision of M.P.S.. C-K.H. provided African turquoise killifish RNA-

309 seq datasets pre-publication. P.P.S. wrote the manuscript with the help from G.A.R. and A.B.. All  
310 the authors provided intellectual input and commented on the manuscript.

## 311 SUPPLEMENTARY MATERIALS

312 Material and Methods

313 References 81–131

314 Figs. S1 to S17

315 Table S1

316 Data Files S1 to S8

317

## 318 REFERENCES

- 319 1. A. Cellerino, D. R. Valenzano, M. Reichard, From the bush to the bench: the annual  
320 Nothobranchius fishes as a new model system in biology. *Biol Rev Camb Philos Soc* **91**,  
321 511-533 (2016).
- 322 2. M. Platzer, C. Englert, Nothobranchius furzeri: A Model for Aging Research and More.  
323 *Trends Genet* **32**, 543-552 (2016).
- 324 3. M. Vrtilek, J. Zak, M. Psenicka, M. Reichard, Extremely rapid maturation of a wild  
325 African annual fish. *Curr Biol* **28**, R822-R824 (2018).
- 326 4. C. K. Hu, A. Brunet, The African turquoise killifish: A research organism to study  
327 vertebrate aging and diapause. *Aging Cell* **17**, e12757 (2018).
- 328 5. M. Reichard, M. Polacik, Nothobranchius furzeri, an 'instant' fish from an ephemeral  
329 habitat. *Elife* **8**, (2019).
- 330 6. C. K. Hu *et al.*, Vertebrate diapause preserves organisms long term through Polycomb  
331 complex members. *Science* **367**, 870-874 (2020).
- 332 7. M. B. Renfree, J. C. Fenelon, The enigma of embryonic diapause. *Development* **144**, 3199-  
333 3210 (2017).
- 334 8. K. Reichwald *et al.*, Insights into Sex Chromosome Evolution and Aging from the Genome  
335 of a Short-Lived Fish. *Cell* **163**, 1527-1538 (2015).
- 336 9. J. T. Wagner *et al.*, The genome of Austrofundulus limnaeus offers insights into extreme  
337 vertebrate stress tolerance and embryonic development. *BMC Genomics* **19**, 155 (2018).
- 338 10. Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy. *Nat*  
339 *Rev Genet* **18**, 411-424 (2017).
- 340 11. L. Z. Holland, Evolution of new characters after whole genome duplications: insights from  
341 amphioxus. *Semin Cell Dev Biol* **24**, 101-109 (2013).
- 342 12. S. Ohno, U. Wolf, N. B. Atkin, Evolution from fish to mammals by gene duplication.  
343 *Hereditas* **59**, 169-187 (1968).
- 344 13. G. C. Conant, K. H. Wolfe, Turning a hobby into a job: how duplicated genes find new  
345 functions. *Nat Rev Genet* **9**, 938-950 (2008).

- 346 14. R. Cui *et al.*, Ancestral transoceanic colonization and recent population reduction in a  
347 nonannual killifish from the Seychelles archipelago. *Mol Ecol*, (2021).
- 348 15. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative  
349 genomics. *Genome Biol* **20**, 238 (2019).
- 350 16. W. J. Murphy, G. E. Collier, A molecular phylogeny for aplocheiloid fishes  
351 (Atherinomorpha, Cyprinodontiformes): the role of vicariance and the origins of  
352 annualism. *Mol Biol Evol* **14**, 790-799 (1997).
- 353 17. A. I. Furness, D. N. Reznick, M. S. Springer, R. W. Meredith, Convergent evolution of  
354 alternative developmental trajectories associated with diapause in African and South  
355 American killifish. *Proc Biol Sci* **282**, (2015).
- 356 18. K. L. M. Martin, J. E. Podrabsky, Hit pause: Developmental arrest in annual killifishes and  
357 their close relatives. *Dev Dyn* **246**, 858-866 (2017).
- 358 19. J. E. Podrabsky, M. Arezo, Annual killifishes as model systems for advancing  
359 understanding of evolution and developmental biology. *Dev Dyn* **246**, 778 (2017).
- 360 20. J. D. Buenrostro, P. G. Giresi, L. C. Zaba, H. Y. Chang, W. J. Greenleaf, Transposition of  
361 native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-  
362 binding proteins and nucleosome position. *Nat Methods* **10**, 1213-1218 (2013).
- 363 21. F. Marletaz *et al.*, Amphioxus functional genomics and the origins of vertebrate gene  
364 regulation. *Nature* **564**, 64-70 (2018).
- 365 22. M. Rebeiz, M. Tsiantis, Enhancer evolution and the origins of morphological novelty. *Curr*  
366 *Opin Genet Dev* **45**, 115-123 (2017).
- 367 23. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory activities of transposable elements:  
368 from conflicts to benefits. *Nat Rev Genet* **18**, 71-86 (2017).
- 369 24. S. Heinz *et al.*, Simple combinations of lineage-determining transcription factors prime cis-  
370 regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589  
371 (2010).
- 372 25. J. Liu, M. Robinson-Rechavi, Robust inference of positive selection on regulatory  
373 sequences in the human brain. *Sci Adv* **6**, (2020).
- 374 26. R. Cui *et al.*, Relaxed Selection Limits Lifespan by Increasing Mutation Load. *Cell* **178**,  
375 385-399 e320 (2019).
- 376 27. S. Kousteni, FoxO1, the transcriptional chief of staff of energy metabolism. *Bone* **50**, 437-  
377 443 (2012).
- 378 28. L. Poulsen, M. Siersbaek, S. Mandrup, PPARs: fatty acid sensors controlling metabolism.  
379 *Semin Cell Dev Biol* **23**, 631-639 (2012).
- 380 29. P. F. Almáida-Pagan *et al.*, Age-related changes in mitochondrial membrane composition  
381 of *Nothobranchius furzeri*: comparison with a longer-living *Nothobranchius* species.  
382 *Biogerontology* **20**, 83-92 (2019).

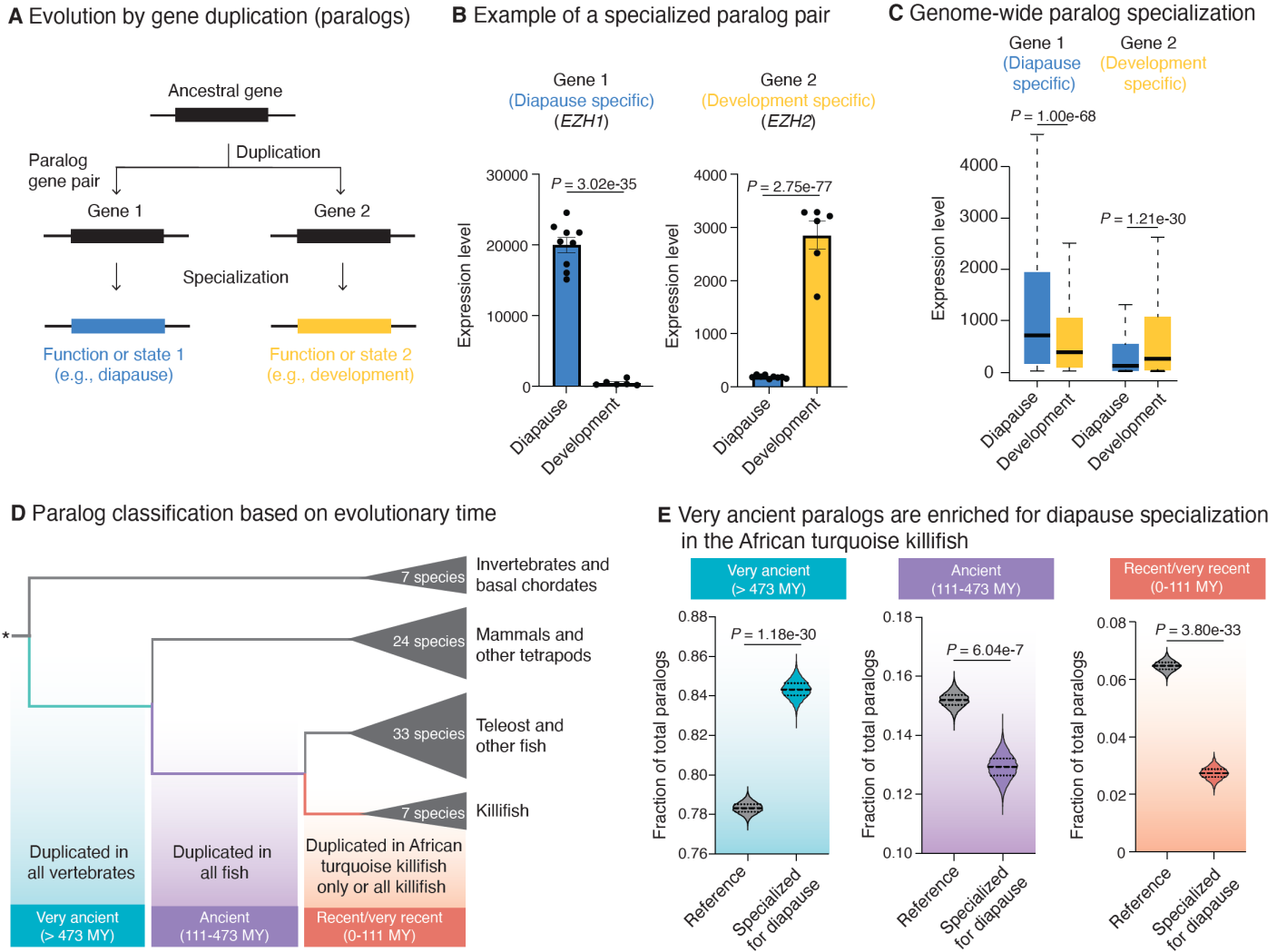
- 383 30. R. Dabrowski, R. Ripa, C. Latza, A. Annibal, A. Antebi, Optimization of mass  
384 spectrometry settings for steroidomic analysis in young and old killifish. *Anal Bioanal*  
385 *Chem* **412**, 4089-4099 (2020).
- 386 31. D. E. Zajic, J. E. Podrabsky, Metabolomics analysis of annual killifish (*Austrofundulus*  
387 *limnaeus*) embryos during aerial dehydration stress. *Physiol Genomics* **52**, 408-422 (2020).
- 388 32. M. T. Nakamura, B. E. Yudell, J. J. Loor, Regulation of energy metabolism by long-chain  
389 fatty acids. *Prog Lipid Res* **53**, 124-144 (2014).
- 390 33. P. J. Sindelar, Z. Guan, G. Dallner, L. Ernster, The protective role of plasmalogens in iron-  
391 induced lipid peroxidation. *Free Radic Biol Med* **26**, 318-324 (1999).
- 392 34. N. E. Braverman, A. B. Moser, Functions of plasmalogen lipids in health and disease.  
393 *Biochim Biophys Acta* **1822**, 1442-1452 (2012).
- 394 35. S. Paul, G. I. Lancaster, P. J. Meikle, Plasmalogens: A potential therapeutic target for  
395 neurodegenerative and cardiometabolic disease. *Prog Lipid Res* **74**, 186-195 (2019).
- 396 36. A. Monteiro, O. Podlaha, Wings, horns, and butterfly eyespots: how do complex traits  
397 evolve? *PLoS Biol* **7**, e37 (2009).
- 398 37. X. Lan, J. K. Pritchard, Coregulation of tandem duplicate genes slows evolution of  
399 subfunctionalization in mammals. *Science* **352**, 1009-1013 (2016).
- 400 38. K. Guschanski, M. Warnefors, H. Kaessmann, The evolution of duplicate gene expression  
401 in mammalian organs. *Genome Res* **27**, 1461-1474 (2017).
- 402 39. M. E. Arnegard, D. J. Zwickl, Y. Lu, H. H. Zakon, Old gene duplication facilitates origin  
403 and diversification of an innovative communication system--twice. *Proc Natl Acad Sci U S*  
404 *A* **107**, 22172-22177 (2010).
- 405 40. K. Knox, J. C. Baker, Genomic evolution of the placenta using co-option and duplication  
406 and divergence. *Genome Res* **18**, 695-705 (2008).
- 407 41. A. Doherty, J. P. de Magalhaes, Has gene duplication impacted the evolution of Eutherian  
408 longevity? *Aging Cell* **15**, 978-980 (2016).
- 409 42. J. M. Vazquez, V. J. Lynch, Pervasive duplication of tumor suppressors in Afrotherians  
410 during the evolution of large bodies and reduced cancer risk. *Elife* **10**, (2021).
- 411 43. D. Tejada-Martinez, J. P. de Magalhaes, J. C. Opazo, Positive selection and gene  
412 duplications in tumour suppressor genes reveal clues about how cetaceans resist cancer.  
413 *Proc Biol Sci* **288**, 20202592 (2021).
- 414 44. F. C. Jones *et al.*, The genomic basis of adaptive evolution in threespine sticklebacks.  
415 *Nature* **484**, 55-61 (2012).
- 416 45. D. Villar, P. Flicek, D. T. Odom, Evolution of transcription factor binding in metazoans -  
417 mechanisms and functional implications. *Nat Rev Genet* **15**, 221-233 (2014).
- 418 46. J. Vierstra *et al.*, Mouse regulatory DNA landscapes reveal global principles of cis-  
419 regulatory evolution. *Science* **346**, 1007-1012 (2014).
- 420 47. D. Villar *et al.*, Enhancer evolution across 20 mammalian species. *Cell* **160**, 554-566 (2015).

- 421 48. V. B. Indjeian *et al.*, Evolving New Skeletal Traits by cis-Regulatory Changes in Bone  
422 Morphogenetic Proteins. *Cell* **164**, 45-56 (2016).
- 423 49. J. P. Verta, F. C. Jones, Predominance of cis-regulatory changes in parallel expression  
424 divergence of sticklebacks. *Elife* **8**, (2019).
- 425 50. V. J. Lynch *et al.*, Ancient transposable elements transformed the uterine regulatory  
426 landscape and transcriptome during the evolution of mammalian pregnancy. *Cell Rep* **10**,  
427 551-561 (2015).
- 428 51. E. B. Chuong, N. C. Elde, C. Feschotte, Regulatory evolution of innate immunity through  
429 co-option of endogenous retroviruses. *Science* **351**, 1083-1087 (2016).
- 430 52. J. H. Notwell, T. Chung, W. Heavner, G. Bejerano, A family of transposable elements co-  
431 opted into developmental enhancers in the mouse neocortex. *Nat Commun* **6**, 6644 (2015).
- 432 53. M. Trizzino *et al.*, Transposable elements are the primary source of novelty in primate gene  
433 regulation. *Genome Res* **27**, 1623-1633 (2017).
- 434 54. V. Sundaram, J. Wysocka, Transposable elements as a potent source of diverse cis-  
435 regulatory sequences in mammalian genomes. *Philos Trans R Soc Lond B Biol Sci* **375**,  
436 20190347 (2020).
- 437 55. M. Roller *et al.*, LINE retrotransposons characterize mammalian tissue-specific and  
438 evolutionarily dynamic regulatory regions. *Genome Biol* **22**, 62 (2021).
- 439 56. R. L. Cosby *et al.*, Recurrent evolution of vertebrate transcription factors by transposase  
440 capture. *Science* **371**, (2021).
- 441 57. J. Judd, H. Sanderson, C. Feschotte, Evolution of mouse circadian enhancers from  
442 transposable elements. *Genome Biol* **22**, 193 (2021).
- 443 58. W. Wang *et al.*, Changes in regeneration-responsive enhancers shape regenerative  
444 capacities in vertebrates. *Science* **369**, (2020).
- 445 59. D. R. Valenzano *et al.*, The African Turquoise Killifish Genome Provides Insights into  
446 Evolution and Genetic Architecture of Lifespan. *Cell* **163**, 1539-1554 (2015).
- 447 60. A. Sahm, M. Bens, M. Platzer, A. Cellerino, Parallel evolution of genes controlling  
448 mitonuclear balance in short-lived annual fishes. *Aging Cell* **16**, 488-496 (2017).
- 449 61. M. B. Emanuel Barth, Luca Dolfi, Rongfeng Cui, Marco Groth, Roberto Ripa, Aurora  
450 Savino, Dario R. Valenzano, Matthias Platzer, Manja Marz, Alessandro Cellerino, Analysis  
451 of microRNA expression reveals convergent evolution of the molecular control of diapause  
452 in annual fish. DOI:10.21203/rs.3.rs-744922/v1.
- 453 62. M. Vrtilik, M. Polacik, M. Reichard, The role of energetic reserves during embryonic  
454 development of an annual killifish. *Dev Dyn* **246**, 838-847 (2017).
- 455 63. L. Olsen, E. Thum, N. Rohner, Lipid metabolism in adaptation to extreme nutritional  
456 challenges. *Dev Cell* **56**, 1417-1429 (2021).

- 457 64. A. L. T. Romney, E. M. Davis, M. M. Corona, J. T. Wagner, J. E. Podrabsky, Temperature-  
458 dependent vitamin D signaling regulates developmental trajectory associated with diapause  
459 in an annual killifish. *Proc Natl Acad Sci U S A* **115**, 12763-12768 (2018).
- 460 65. A. Antebi, W. H. Yeh, D. Tait, E. M. Hedgecock, D. L. Riddle, daf-12 encodes a nuclear  
461 receptor that regulates the dauer diapause and developmental age in *C. elegans*. *Genes Dev*  
462 **14**, 1512-1527 (2000).
- 463 66. B. Gerisch, C. Weitzel, C. Kober-Eisermann, V. Rottiers, A. Antebi, A hormonal signaling  
464 pathway influencing *C. elegans* metabolism, reproductive development, and life span. *Dev*  
465 *Cell* **1**, 841-851 (2001).
- 466 67. C. Kenyon, J. Chang, E. Gensch, A. Rudner, R. Tabtiang, A *C. elegans* mutant that lives  
467 twice as long as wild type. *Nature* **366**, 461-464 (1993).
- 468 68. P. L. Larsen, P. S. Albert, D. L. Riddle, Genes that regulate both development and  
469 longevity in *Caenorhabditis elegans*. *Genetics* **139**, 1567-1583 (1995).
- 470 69. S. Gottlieb, G. Ruvkun, daf-2, daf-16 and daf-23: genetically interacting genes controlling  
471 Dauer formation in *Caenorhabditis elegans*. *Genetics* **137**, 107-120 (1994).
- 472 70. C. J. Nelson, J. P. Otis, H. V. Carey, A role for nuclear receptors in mammalian  
473 hibernation. *J Physiol* **587**, 1863-1870 (2009).
- 474 71. R. Arena *et al.*, Lipid droplets in mammalian eggs are utilized during embryonic diapause.  
475 *Proc Natl Acad Sci U S A* **118**, (2021).
- 476 72. H. T. Jansen *et al.*, Hibernation induces widespread transcriptional remodeling in metabolic  
477 tissues of the grizzly bear. *Commun Biol* **2**, 336 (2019).
- 478 73. Y. Xu *et al.*, Molecular signatures of mammalian hibernation: comparisons with alternative  
479 phenotypes. *BMC Genomics* **14**, 567 (2013).
- 480 74. X. Zhou *et al.*, Beaver and Naked Mole Rat Genomes Reveal Common Paths to Longevity.  
481 *Cell Rep* **32**, 107949 (2020).
- 482 75. K. Lin, J. B. Dorman, A. Rodan, C. Kenyon, daf-16: An HNF-3/forkhead family member  
483 that can function to double the life-span of *Caenorhabditis elegans*. *Science* **278**, 1319-1322  
484 (1997).
- 485 76. S. Ogg *et al.*, The Fork head transcription factor DAF-16 transduces insulin-like metabolic  
486 and longevity signals in *C. elegans*. *Nature* **389**, 994-999 (1997).
- 487 77. M. E. Giannakou *et al.*, Long-lived *Drosophila* with overexpressed dFOXO in adult fat  
488 body. *Science* **305**, 361 (2004).
- 489 78. D. S. Hwangbo, B. Gershman, M. P. Tu, M. Palmer, M. Tatar, *Drosophila* dFOXO controls  
490 lifespan and regulates insulin signalling in brain and fat body. *Nature* **429**, 562-566 (2004).
- 491 79. T. Lu *et al.*, REST and stress resistance in ageing and Alzheimer's disease. *Nature* **507**,  
492 448-454 (2014).
- 493 80. J. M. Zullo *et al.*, Regulation of lifespan by neural excitation and REST. *Nature* **574**, 359-  
494 364 (2019).

## MAIN FIGURES

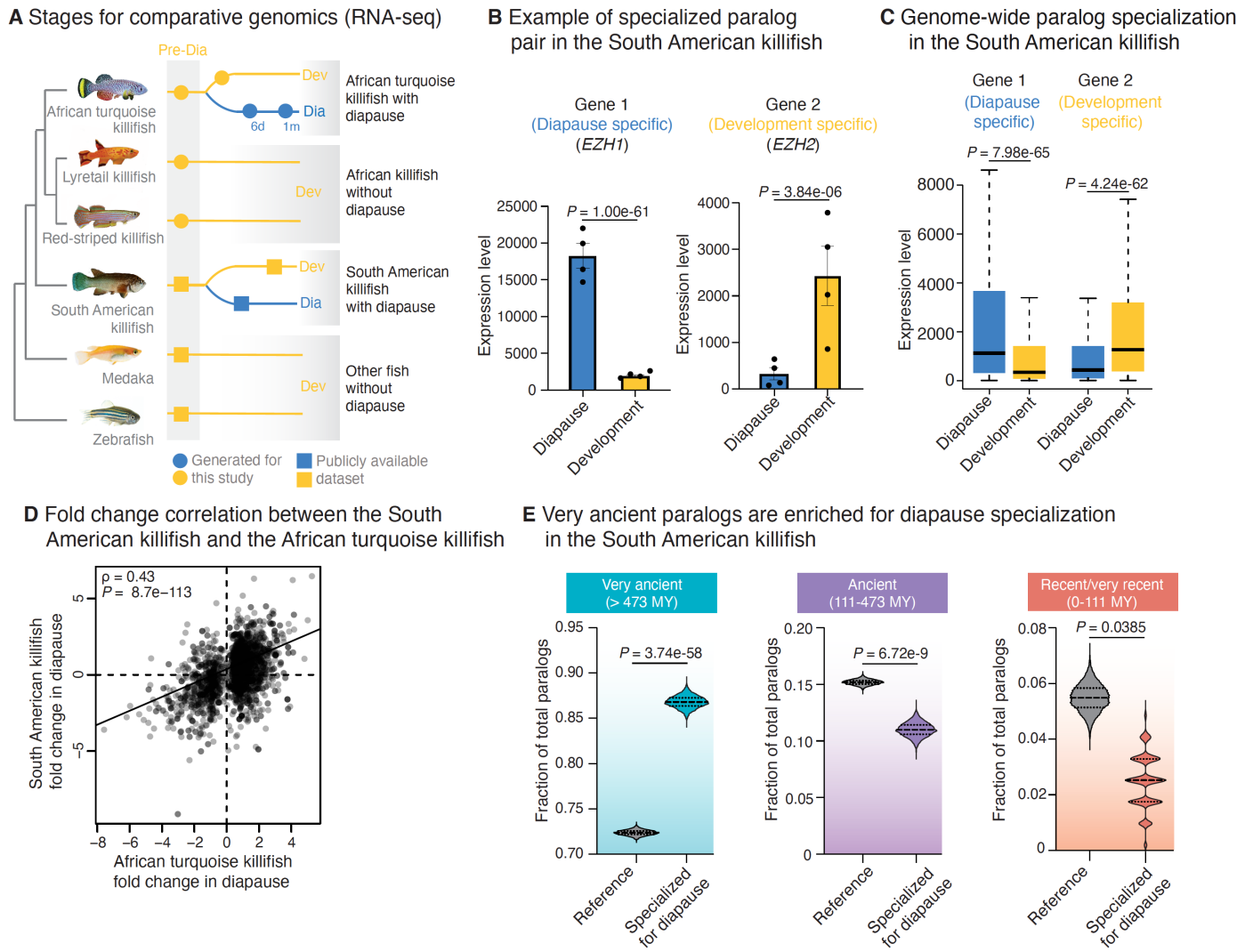
Figure 1





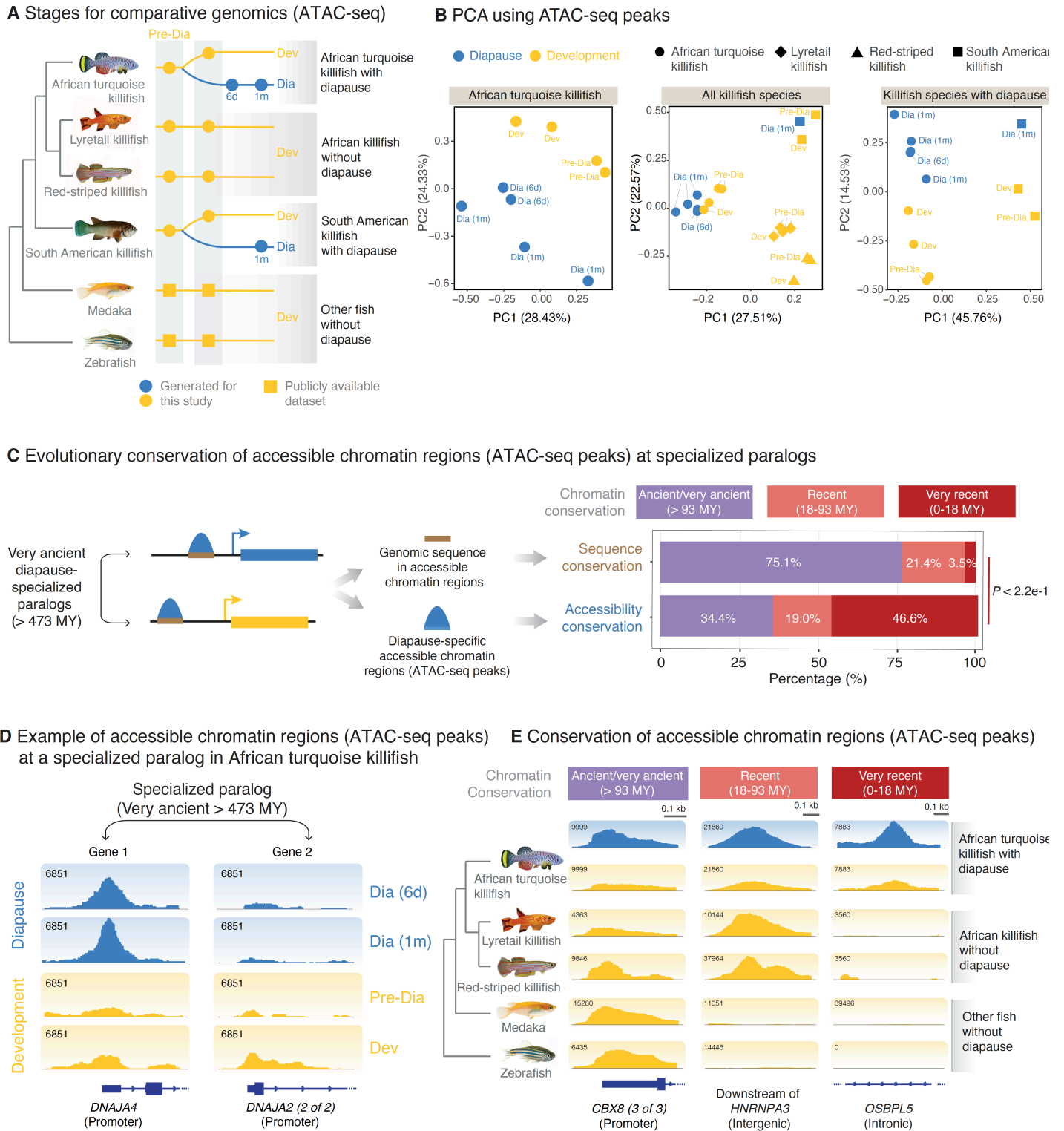
495 **Figure 1. Specialization of very ancient paralogs for expression in diapause in African turquoise**  
496 **killifish.** (A) Schematic of paralog specialization after gene duplication. After duplication from the same  
497 ancestral gene, genes from a paralog gene pair can specialize for different functions or states (e.g., diapause  
498 vs. development). (B) Examples of a paralog gene pair, with specialized expression of gene 1 in diapause  
499 (blue, *EZH1*) and gene 2 in development (yellow, *EZH2*) in the African turquoise killifish (*Nothobranchius*  
500 *furzeri*). Bars represent mean expression level (normalized DESeq2 count) across replicates in diapause or  
501 development state. Dots show normalized DESeq2 counts in each replicate. Error bar is standard error of  
502 mean (SEM). Corrected *P*-values (median from pairwise comparisons) from DESeq2 Wald test. (C) Box  
503 plots showing expression levels (normalized DESeq2 counts) of all the specialized paralog pairs in diapause  
504 and development in the African turquoise killifish genome. Boxes show the median and interquartile ranges,  
505 and whiskers indicate maximum 1.5 interquartile range. Gene 1 of the paralog pair has a higher expression  
506 on average in diapause (blue) compared to development (yellow), whereas gene 2 has a higher expression  
507 on average in development (yellow) compared to diapause (blue). *P*-values from Kolmogorov-Smirnov  
508 test. (D) Schematic for binning paralog duplication time into 3 categories based on OrthoFinder pipeline  
509 with 71 species (see Methods and fig. S2 for complete tree). Divergence time estimates are from species  
510 tree in Ensembl. Binned categories include genes that were duplicated in the common ancestor of (1) all  
511 vertebrates or earlier (very ancient, > 473 million years [MY]), (2) all fish (ancient, 111-473 MY), and (3)  
512 all killifish or African turquoise killifish exclusively (recent/very recent, 0-111 MY). (E) Fraction of total  
513 paralog pairs within each of the very ancient (left), ancient (middle), and recent/very recent (right) binned  
514 categories. Violin plots represent the distribution of observed vs expected specialized paralog fractions  
515 generated through 10,000 bootstrapped random sampling. Median and quartiles are indicated by dashed  
516 lines. The enrichment of diapause-specialized paralog pairs within each bin is compared to genome-wide  
517 expectation (see Methods). Compared to the reference, paralogs with specialization in diapause are enriched  
518 among genes with very ancient duplication times and depleted among genes with ancient and recent/very  
519 recent duplication times respectively. *P*-values from Chi-square test.

Figure 2



520 **Figure 2. Very ancient paralogs also specialize for expression in diapause in other killifish species**  
521 **with diapause.** (A) Experimental design for analysis of RNA-seq datasets either publicly available (square)  
522 or *de-novo* generated for this study (circles). Killifish species are from Africa (with and without diapause)  
523 and South America (with diapause). Medaka and zebrafish are other teleost fish without diapause. The  
524 development stage (yellow square) in the South American killifish corresponds to post-diapause  
525 development. Pre-Dia, pre-diapause; Dia, diapause; Dev, development; 6d, 6 days in diapause; 1m, 1 month  
526 in diapause. (B) Examples of paralog gene pair, with specialized expression of gene 1 in diapause (blue,  
527 *EZH1*) and gene 2 in development (yellow, *EZH2*) in South American killifish (*Austrofundulus limnaeus*).  
528 Gene names displayed are the names assigned to the ortholog in African turquoise killifish. Bars represent  
529 the mean expression level (normalized DESeq2 count) across replicates in diapause or post-diapause  
530 development state. Error bar is standard error of mean (SEM). Each dot represents the normalized  
531 expression level for all sample replicates in diapause or post diapause development. *P*-values from DESeq2  
532 Wald test. (C) Box plots showing the expression levels (normalized DESeq2 counts) of all the specialized  
533 paralog pairs in diapause and development in South American killifish. Boxes show the median and  
534 interquartile ranges, and whiskers indicate maximum 1.5 interquartile range. Gene 1 of the paralog pair has  
535 a higher expression on average in diapause (blue) compared to development (yellow), whereas gene 2 has  
536 a higher expression on average in development (yellow) compared to diapause (blue). *P*-values from  
537 Kolmogorov-Smirnov test. (D) Spearman's rank correlation between ortholog genes that change with  
538 diapause in African turquoise killifish and South American killifish. Dots represent the fold change values  
539 of ortholog genes in diapause compared to development in the two species. Spearman's correlation  
540 coefficient ( $\rho$ ) and *P*-values from t-test are shown on the plot. (E) Fraction of the total paralog pairs within  
541 each of the very ancient (left), ancient (middle), and recent/very recent (right) binned categories as  
542 described in Fig. 1D. Violin plots represent the distribution of observed vs expected specialized paralog  
543 fractions generated through 10,000 bootstrapped random sampling. Median and quartiles are indicated by  
544 dashed lines. The enrichment of diapause-specialized paralogs pairs within each bin is compared to  
545 genome-wide expectation (reference). Compared to the reference, paralogs with specialization in diapause  
546 are enriched among genes with very ancient duplication times and depleted among genes with ancient and  
547 recent or very recent duplication times respectively. *P*-values from Chi-square test.

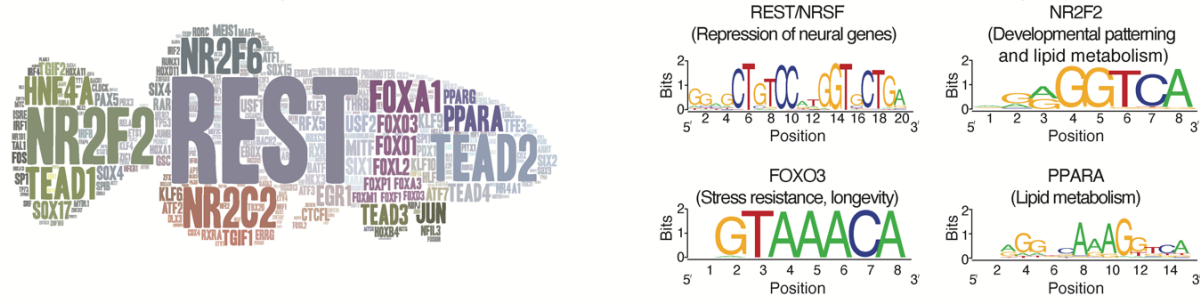
Figure 3



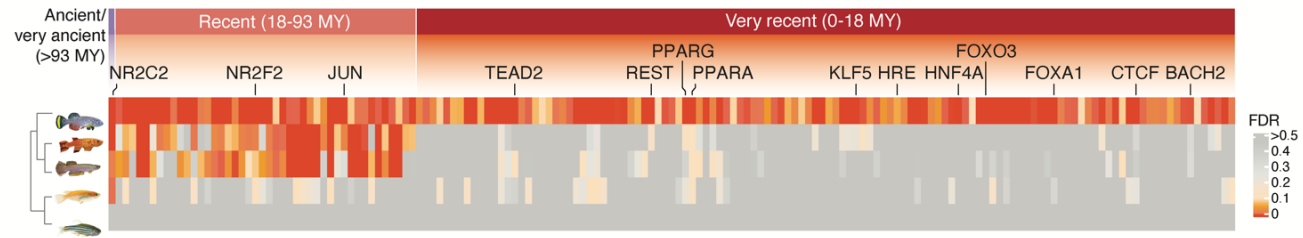
548 **Figure 3. Evolutionarily recent remodeling of genome-wide chromatin landscape drives specialized**  
549 **expression of very ancient paralogs in diapause.** (A) Experimental design for the ATAC-seq datasets  
550 either publicly available (squares) or de novo generated in this study (circles) (see also Data File S2). Pre-  
551 Dia, pre-diapause; Dia, diapause; Dev, development; 6d, 6 days in diapause; 1m, 1 month in diapause. (B)  
552 Principal component analysis (PCA) on all chromatin accessibility regions in each species group: African  
553 turquoise killifish only (left), all killifish species (center), and diapause-capable killifish (right). Each point  
554 represents the consensus ATAC-seq peaks (chromatin accessibility) from an individual replicate of a given  
555 species at a given developmental or diapause state. Percentage of the variance explained by each Principal  
556 Component (PC) is shown in parentheses. (C) Conservation analysis of genomic sequence and chromatin  
557 accessibility at very ancient paralogs with specialization in diapause vs. development. Left: Schematic of  
558 the analysis. Right: Percentage (e.g., conservation) of alignable regions containing diapause-specific  
559 chromatin accessibility (upper) and the conservation of diapause-specific chromatin accessibility (lower)  
560 near specialized ancient paralogs (see also fig. S10). While the majority of genomic sequences are ancient,  
561 the chromatin accessibility at those peaks evolved recently in the African turquoise killifish. *P*-values from  
562 Chi-square test. (D) Genome browser (IGV) visualization of chromatin accessibility regions (ATAC-seq  
563 peaks) at the promoter of genes from a very ancient paralog pair, with one gene specialized for expression  
564 in diapause (*DNAJA4*) and the other in development (*DNAJA2*). Replicates within each condition were  
565 aggregated by summation for visualization. Blue boxes and lines represent genomic features (exons and  
566 introns, respectively). (E) Genome-browser (IGV) visualization of representative diapause-specific  
567 chromatin accessibility regions that are: ancient/very ancient (conserved across killifish and medaka or  
568 zebrafish; left), recent (conserved in killifish species; middle), or very recent (specific to African turquoise  
569 killifish; right). Blue boxes and lines at the bottom represent genomic features for the closest gene (exons  
570 and introns, respectively). To generate tracks, RPKM-normalized reads were summed across replicates and  
571 biological timepoints (e.g., diapause and development separately) to obtain single tracks for each species.

Figure 4

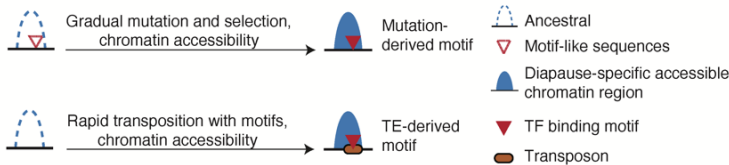
**A** Transcription factor (TF) binding motifs at diapause-specific accessible chromatin near specialized paralogs



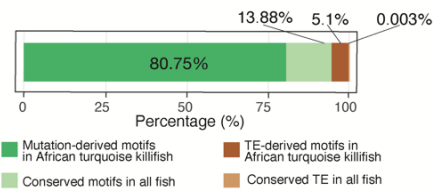
**B** Evolution of TF binding sites at diapause-specific accessible chromatin near specialized paralogs



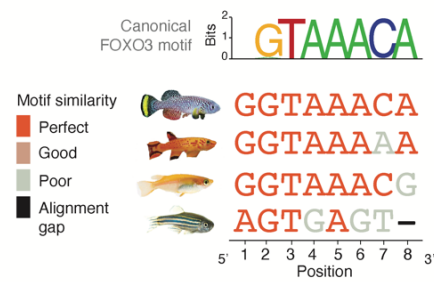
**C** Schematic of TF binding motif evolution



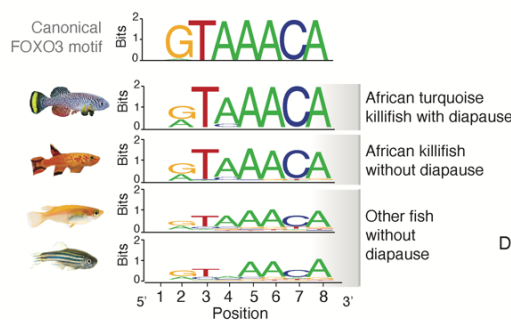
**D** Diapause motif evolution breakdown



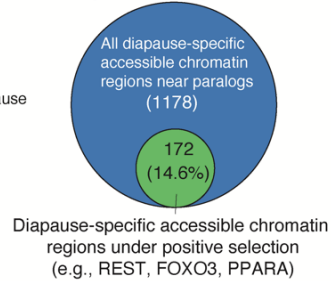
**E** Example of TF binding motif evolution



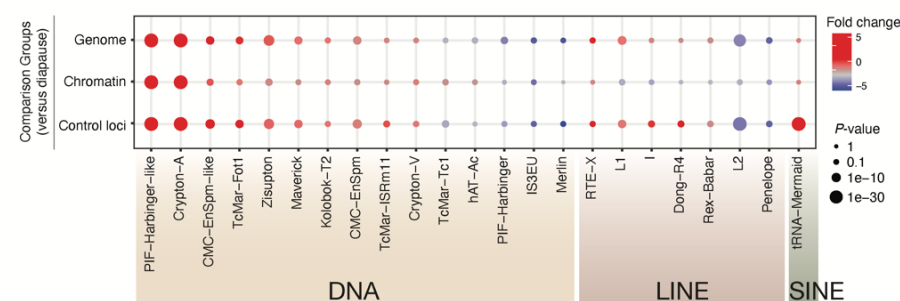
**F** Motif evolution genome-wide



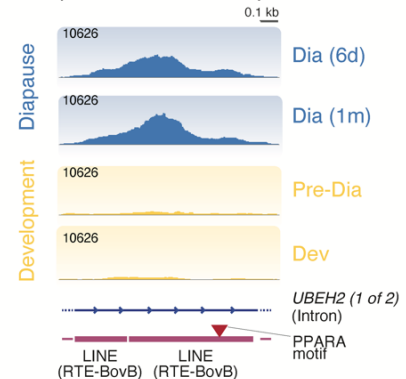
**G** Positive selection of regulatory regions



**H** Transposon enrichment in diapause-specific accessible chromatin regions



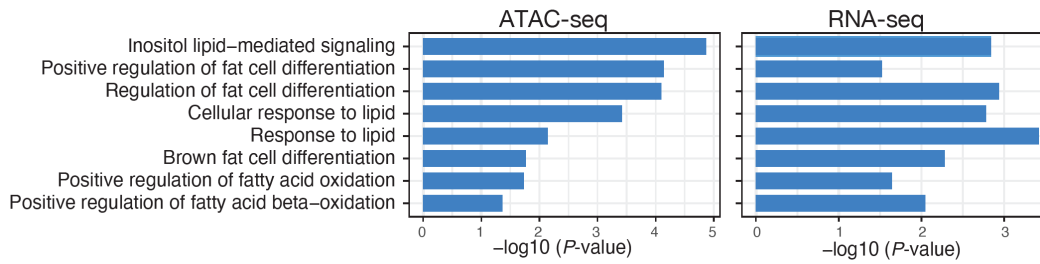
**I** Example motif that overlaps with a TE



572 **Figure 4. Mechanisms underlying the evolution of chromatin accessibility in diapause at specialized**  
573 **paralogs.** (A) Left: Word cloud for transcription-factor binding motifs enriched in the diapause-specific  
574 chromatin accessible regions (ATAC-seq peaks) at specialized paralogs using HOMER. Right: Motif logos  
575 generated by weblogo for selected transcription factor examples are shown on the right. Y-axis is formatted  
576 using *informational content* (i.e., bits) which scales based on single-base overrepresentation in the binding  
577 sequence (0: bases are represented equally at 25% each in reference sequences; 2: a single base dominates  
578 the entirety of reference sequences at 100%). (B) Conservation in other fish species of transcription-factor  
579 binding motifs enriched in the diapause-specific chromatin accessible regions in the African turquoise  
580 killifish. The majority of diapause-specific motifs are very recent (i.e., specific to the African turquoise  
581 killifish) and not enriched in killifish species without diapause. Selected representative motifs are  
582 highlighted (see also figs. S11, S17). (C) Schematic of two possible mechanisms for the evolution of the  
583 diapause-specific transcription factor binding motifs in the African turquoise killifish genome. Upper:  
584 Gradual mutations paired with selective pressure leads to the formation of binding sites for specific  
585 transcription factors and accompanying chromatin accessibility. Lower: A site experiences a transposable  
586 element (TE) insertion event, providing a novel sequence that contains a binding site for a specific  
587 transcription factor and accompanying chromatin accessibility. (D) Contribution (in percentage) of the two  
588 possible evolutionary mechanisms (mutation and TE insertion) for motif evolution in diapause-specific  
589 chromatin at specialized paralogs. Diapause-specific motifs likely acquired by mutation in the African  
590 turquoise killifish (dark green) are those that are expected to bind their transcription factors exclusively in  
591 the African turquoise killifish based on the HOMER log odds ratio binding criteria (see Methods).  
592 Conserved motifs likely acquired by mutation in multiple species (light green) are those that are expected  
593 to bind their transcription factor motifs in at least one other fish species evaluated. Diapause-specific motifs  
594 likely acquired by TE insertion (brown) are those that overlap with an annotated TE sequence and are  
595 expected to bind their transcription factor exclusively in the African turquoise killifish. Conserved motifs  
596 likely acquired by TE insertion (light brown) are those that overlap with a TE sequence in the African  
597 turquoise killifish and at least one other species in addition to being expected to bind their transcription  
598 factor in both the African turquoise killifish and at least one additional species. A large majority of motifs  
599 likely evolved through mutations in the African turquoise killifish. (E) Example of a transcription factor  
600 binding motif that likely evolved via mutation: *FOXO3* binding motif in a diapause-specific chromatin  
601 accessible region of the African turquoise killifish genome near *GPC3* (*LOC107379575*) and aligned  
602 regions in other fish species. Aligned sequences colored based on the similarity to the canonical *FOXO3*  
603 motif from HOMER (top track). (F) Aggregated informational content across all *FOXO3* transcription  
604 factor binding sites in diapause-specific accessible chromatin regions and aligned regions in other species.  
605 Y-axis is formatted using *informational content* (i.e., bits). The African turquoise killifish motif exhibits  
606 greater similarity to the canonical *FOXO3* binding motif (see also fig. S12). (G) Fraction of diapause-  
607 specific chromatin accessible regions under positive selection in the African turquoise killifish (see  
608 Methods) at FDR < 0.1, which are enriched for selected motifs (e.g., REST, FOXO3, PPARA). (H)  
609 Enrichment or depletion of specific transposable elements (TEs) in the diapause-specific chromatin  
610 accessible regions (ATAC-seq peaks) in the African turquoise killifish genome as compared to the overall  
611 genomic abundance of the given TE (“Genome”), compared to the abundance of the given TE in all  
612 chromatin accessible regions (“Chromatin”), and compared to the abundance of the given TE in size-  
613 matched control regions 10 kb away from ATAC-seq peak of interest (“Control loci”). (I) Example of a  
614 diapause-specific chromatin accessibility region containing a PPARA binding site overlapping with a TE.  
615 Almost all TE-derived motifs are specific to the African turquoise killifish.

Figure 5

**A Lipid-related functions in diapause**

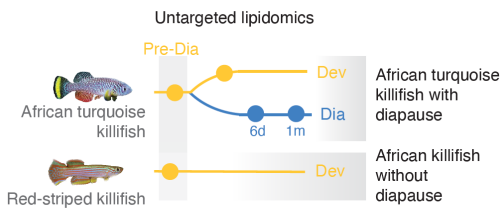


**B Upstream regulators**

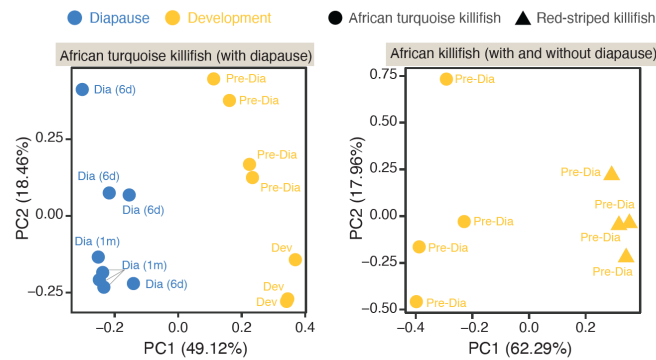
Regulator	FDR
JUN	5.2e-10
PPARG	2.60e-9
REST	7.04e-8
FOXO3	1.02e-5
FOXO1	2.13e-5
TEAD2	3.78e-2

◀ Lipid metabolism regulators

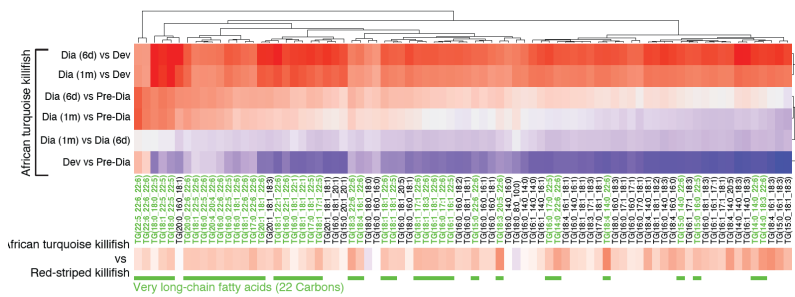
**C Stages for lipidomics**



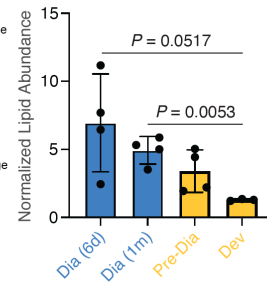
**D PCA based on lipidomics**



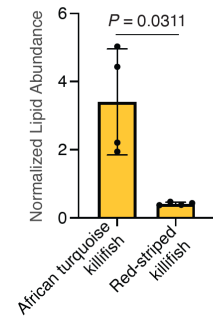
**E Triglyceride enriched in the African turquoise killifish and red-striped killifish**



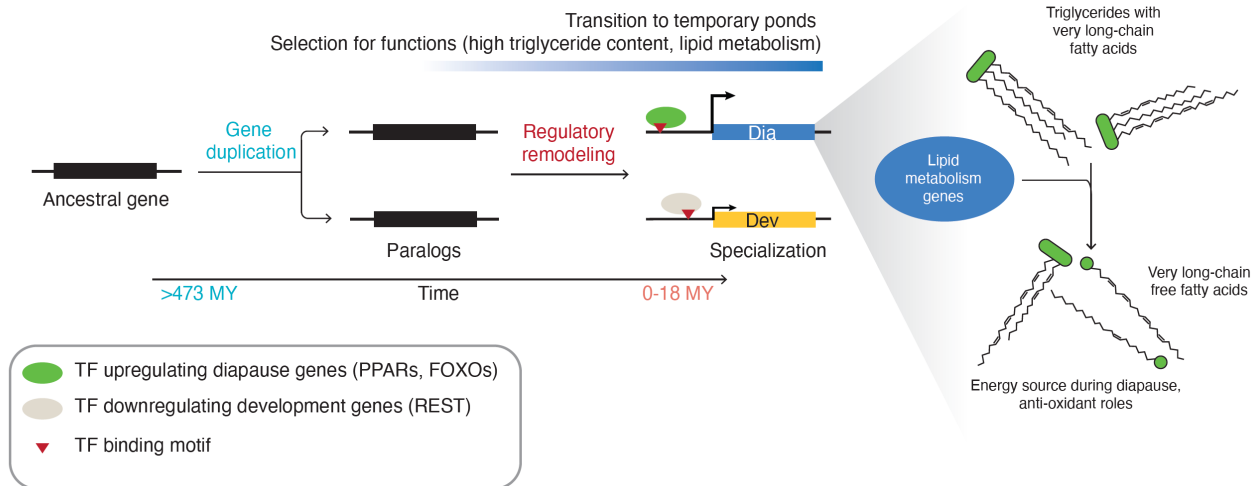
**F Very long-chain fatty acids in triglyceride abundance in the African turquoise killifish**



**G Very long-chain fatty acids in triglyceride abundance between species**



**H Model for the evolution of diapause**





616 **Figure 5. Functional enrichment and lipidomics reveal specific lipids in the diapause state.** (A) Gene  
617 Ontology enrichment for the lipid-related terms enriched in both expressed genes (diapause specialized  
618 paralogs) and chromatin accessible regions in diapause (see Data File S4 for full list). (B) Selected upstream  
619 regulators of the paralog genes specialized in diapause using Ingenuity Pathway Analysis (see Data File S4  
620 for full list). Regulators related to lipid metabolism are highlighted (red arrows). Several upstream  
621 regulatory transcription factors in this analysis overlap with transcription factor binding motifs enriched in  
622 diapause-specific chromatin accessible regions (see Fig. 3). (C) Experimental design for untargeted  
623 lipidomics in the African turquoise killifish (with diapause) and the red-striped killifish (without diapause).  
624 The two development and diapause time points in the African turquoise and pre-diapause in red-striped  
625 killifish are identical to those for ATAC-seq (Fig. 3A). (D) Principal component analysis (PCA) on the  
626 estimated concentrations for all detected lipids for the African turquoise killifish only (left) and for both  
627 killifish (right). Each point represents an individual replicate from a given species at a given developmental  
628 or diapause stage. Variance explained by each Principal Component (PC) is shown in parentheses. (E)  
629 Heatmap representing the fold change of significant triglycerides between diapause vs. development in the  
630 African turquoise killifish (upper) and between the African turquoise killifish vs. red-striped killifish  
631 (development only, lower). Fold change values are plotted between each pair-wise comparison between  
632 diapause and development time points, or the two development or diapause time points. For most  
633 triglycerides, levels are higher at both 1-month (1m) and 6-days (6d) diapause relative to development. The  
634 bottom panel shows the fold change values of the same lipids in the African turquoise killifish compared  
635 to the red-striped killifish; levels of most triglycerides are higher in African turquoise killifish. Green boxes  
636 and text below heatmap blocks denote very long-chain fatty acids among displayed triglycerides species.  
637 (F) Lipid abundance counts for very long-chain fatty acids in triglycerides (TGs) in the African turquoise  
638 killifish during development (yellow) and diapause (blue) timepoints. Bars represent mean values in each  
639 condition error bars represent standard error of the mean (SEM). Dots represent individual replicates. *P*-  
640 values from Welch's t-test. (G) Normalized lipid abundance counts for very long-chain fatty acids in  
641 triglycerides in the African turquoise killifish (left) and red-striped killifish (right) during matched  
642 developmental timepoints (Pre-Dia). Data represented as in F. *P*-value from Welch's t-test. (H) Putative  
643 model for the evolution of diapause-specialized paralogs in the African turquoise killifish, in which very  
644 ancient paralogs (>473 MY) have undergone regulatory rewiring (TF binding site occurrence and  
645 chromatin-accessibility change) that generates a diapause-specialized (blue) and development-specialized  
646 (yellow) gene in the paralog pair. Some of these diapause expressed paralog genes are important for lipid  
647 metabolism, such as the processing of very long-chain triglycerides. These very long-chain triglycerides  
648 are abundant in the African turquoise killifish and accumulate in diapause, and they could be critical for  
649 long-term survival.