A new fish-specific gene evolves new functions in cichlid fishes

Langyu Gu[1,2,3,*], Chenzheng Li[2], Xiaobing Mao[2], Zongfang Wei[2], Youkui Huang[2], Ximin He[1], Wenjun Zhou[1], Li Li[2,*], Deshou Wang[2,*]

[1] State Key Laboratory of Biocontrol, School of Life Sciences and School of Ecology, Sun Yat-sen University, Guangzhou, Guangdong, China.

[2] Key Laboratory of Freshwater Fish Reproduction and Development (Ministry of Education), Key Laboratory of Aquatic Science of Chongqing, School of Life Sciences, Southwest University, Chongqing, China

[3] Zoological Institute, Department of Environmental Sciences, University of Basel, Basel, Switzerland

*Correspondence:

Langyu Gu: langyugu@gmail.com

Li Li: alisir@swu.edu.cn

Deshou Wang: wdeshou@swu.edu.cn

Abstract

Genes whose dysfunction does not affect normal survival are common. Are they only meaningless residues during evolution? Here, we identified a new fish-specific gene, which we named *lg*. Gene knockout resulted in no obvious phenotype in zebrafish, but *lg* evolved an amino acid mutation that was under positive selection in the modern haplochromine (MH) cichlid fish lineage, the well-known adaptive radiative lineage. Moreover, the cichlid fish-specific upstream region of *lg* drove new eGFP expression in tissues related to adaptation. Noticeably, this homologous region from different cichlid fishes drove different patterns, which is simply due to three MH-segregated SNP mutations that are predicted to bind a hormone-related transcription factor. We thus revealed an initially redundant gene evolving new functions in an adaptive radiative lineage. This further illuminates the mechanism of the emergence of new gene functions with respect to evo-devo in a broad way.

Key words

lg, new gene, lineage-specific, evo-devo, fish

From an evolutionary point of view, the most adaptive organism spends the least resources but produces the most outputs. However, it seems that this rule does not apply to genome organization, which is occupied by a large proportion of genes whose dysfunction does not affect normal survival. Does this make sense?

Nothing in biology makes sense except in the light of evolution [1]. As raw materials for natural selection, the large portion of redundant genes provides good opportunities for functional innovation. This is especially true for lineage-specific genes, which can occupy up to one-third of the genome in many organisms [2]. Studies have gradually revealed that many of these genes can evolve new functions related to novelties and environmental adaptation in a particular lineage. For example, in early brain development in humans [3,4], in the creation of a novel tissue in the sea anemone [5], and in the development of a novel fan trait on the leg in a water spider genus to explore new environments [6]. There is even a whole lineage-specific gene family in water fleas related to environmental stimuli [7]. Lineage-specific genes thus provide a new perspective on the emergence of new gene function.

It is well accepted that different organisms share deep homologous developmental molecular bases (molecular toolkits), such as highly conserved signalling pathways and transcription factors [8–10]. Then, how do new genes gain new functions, and in what way are they integrated into the existing networks? Although usually restricted due to pleiotropic effects [11], protein coding region evolution can take place [12–14]. In comparison, regulatory evolution is a well-known strategy, since changes at the expression level are more tolerable and can offer immediate phenotypic consequences [11,15]. Even a single mutation in the upstream region can induce new expression profiles [16,17]. The relaxed selection

85 pressure induced by the changing expression profiles can further prompt the
86 accumulation of mutations at the protein level.
87
88 In this study, we identified a new unannotated fish-specific gene, which
89 we named *lg* (ENSONIT00000026203). Comparative -omics analyses and
90 functional experiments revealed that multiple losses of *lg* occurred in different
91 fish lineages, and no obvious phenotype was detected in gene knockout
92 zebrafish, which led us to assume that the ancestral function of *lg* is not
93 indispensable. *lg* was mainly expressed in the brain and eyes in most fishes, but
94 it evolved new expression patterns, such as in pharyngeal cartilages and
95 epidermal serotonin-positive neuroepithelia cells, in the cichlid fish lineage,
96 especially in the most species-rich lineage, the modern haplochromine (MH)
97 lineage. Further comparative genomic analyses across the East African cichlid
98 fish phylogeny revealed that *lg* evolved a cichlid fish-specific upstream region
99 possessing three MH lineage-segregated SNPs and one MH lineage-segregated
100 amino acid mutation that was under positive selection. Transgenic experiments
101 confirm the new regulatory functions of this upstream region to drive eGFP
102 expression in zebrafish. Noticeably, the homologous sequence from different
103 cichlid fishes drove different eGFP expression patterns, which was due to the
104 three MH lineage-segregated SNP mutations that are predicted to bind to a
105 hormone-related transcription factor. Taken together, our work revealed for the
106 first time an initially redundant new fish-specific gene that evolved new
107 functions in an adaptive radiative cichlid fish lineage.
108
109 **Results**
110
111 ***lg* is a new fish-specific gene**
112 To see *lg* sequence evolution, we conducted a tblastx search in all species
113 with available genomes in Ensembl 100. The results showed that *lg* only

114 appeared in fishes (Fig. 1A, Table S1). To further characterize its evolution
115 within teleost fishes, we conducted a tblastx search and gene predictions in the
116 draft genome assembly across the whole phylogeny of teleost fishes [18]. The
117 results showed that *lg* was lost multiple times in different lineages (Fig. 1B),
118 which inspired the assumption that the ancestral function of *lg* is not
119 indispensable for fish survival.
120

121 **_lg_ evolved new expression profiles in cichlid fishes, especially in the MH**
122 **lineage**

123 To identify the tissues where *lg* was expressed, we investigated its
124 expression profiles based on the combined results of available RNAseq data
125 retrieved from NCBI and *in situ* hybridization in different fishes (Fig. 2, Table
126 S2). *lg* is mainly expressed in the brain and eyes in most fishes. Interestingly, it
127 evolved new expression profiles in cichlid fishes in pharyngeal cartilages, as
128 well as in the skin of the most species-rich lineage of cichlid fishes—the MH
129 lineage [19]. These results suggest that the ancestral function of *lg* can be related
130 to neurons, but new functions can evolve in other tissues in cichlid fishes,
131 especially in the MH lineage.
132

133 **No obvious phenotype was detected in _lg_ gene knockout zebrafish**

134 To further investigate the function of *lg*, gene knockout was conducted
135 using CRISPR/Cas9 in zebrafish. We successfully obtained F2 homozygous
136 mutant (MUT) zebrafish and their wild-type siblings (WT). Sanger sequencing
137 and Western blot analyses confirmed the successful gene dysfunction with
138 frameshift (5 bp deletion) and protein interruption (stop codon appeared at the
139 18th amino acid) (Fig. 3A). To determine the affected gene expression after gene
140 knockout, comparative transcriptomic analysis was conducted between MUT
141 and WT zebrafish at 2 dpf (days post fertilization) and 4 dpf, respectively. The
142 results showed that most differentially expressed (DE) genes were related to

143 immunity, neurons and energy (Table S3). To determine whether there are
144 different immune responses between MUT and WT, we examined the immune
145 responses of larval zebrafish to lipopolysaccharide (LPS) by counting the
146 numbers of *lyz*$^{+}$ neutrophils in the caudal haematopoietic tissue one day after
147 injection [20]. No significant difference was detected (Fig. 3B). To determine
148 whether gene knockout affects neuronal development in the brain and eyes
149 where *lg* is expressed in zebrafish, we compared the phenotypic changes
150 between MUT and WT using the brain neuron marker HuC [21] and still no
151 difference was detected (Fig. 3C). Then, we turned our attention to one top DE
152 gene, *slc16a6*, which has been reported to be related to energy metabolism
153 under fasting conditions [22]. Under the same strategy, however, we still could not
154 find obvious metabolic phenotypic changes between MUT and WT (Fig. 3C).
155 This can be explained by the fact that *slc16a6* is only downregulated here
156 instead of entirely disrupted, as in the reference [22], so that it can still execute its
157 function. Overall, mutation of *lg* resulted in no obvious phenotypes in zebrafish.
158 These results further support our assumption that the ancestral function of *lg* is
159 not indispensable.
160

161 ***lg* evolved a cichlid fish-specific upstream region possessing MH-segregated**
162 **SNPs and one MH-segregated amino acid mutation that is under positive**
163 **selection**

164 Noticeably, as we mentioned above, *lg* evolved new gene expression
165 profiles in cichlid fishes (Fig. 2), the classical model species for studying
166 adaptive radiation, which explosively radiated into hundreds to thousands of
167 species within a short time with very closely related genomes [23]. To see whether
168 there are cichlids-specific region that might conduct the new regulatory
169 function, genomic alignment of *lg* among cichlid fishes and other teleost fishes
170 were conducted (Fig. S1). Indeed, cichlid fishes showed a large portion of
171 cichlid fish-specific CNS (Conserved Non-Coding Sequences) in the upstream

172 region. To further characterize its sequence evolution within cichlid fishes, we
173 amplified the whole *lg* gene, including this CNS upstream region, in a total of
174 20 cichlid fishes representing main lineages across the East African cichlid fish
175 phylogeny [24] (Fig. 4A). Interestingly, both the upstream and the coding regions
176 exhibited segregation patterns with the MH lineage, the most species-rich
177 lineage of cichlid fishes [25] (Fig. 4A). For example, within the cichlid fish
178 specific CNS upstream region, there are three MH lineage-segregated SNP
179 mutations (Fig. 4A). Transcription factors (TFs) binding prediction revealed
180 that these three SNPs can set up binding opportunities for a hormone-related
181 TF, the glucocorticoid modulatory element binding protein 2 (GMEB2) (Fig.
182 4A). In addition, there is one nonsynonymous amino acid mutation (A to S)
183 located at the 38th residue in the coding region specific to the MH lineage that is
184 under positive selection (Fig. 4A, Table S4). These results prompted us to pay
185 special attention to the *lg* of cichlid fishes in the next steps.
186

187 **The cichlid fish-specific upstream region indeed has regulatory functions**
188 **and performs different roles in different cichlid fish lineages**

189 To determine whether the cichlid fish-specific upstream region of *lg*
190 indeed has regulatory functions, and whether this region of the MH lineage has
191 different regulatory roles compared to other cichlid fish lineages, we
192 constructed two transgenic lines with the homologous upstream regions from
193 different species as the promoters to drive eGFP expression in zebrafish (Fig.
194 4A, Fig. S2). One line is with the upstream region of an MH cichlid fish,
195 *Pundamilia nyererei* (pT2AL-Pn-eGFP transgenic line, Pn line). The other line
196 is with the corresponding homologous region of a cichlid fish belonging to a
197 sister lineage of East African Great Lakes cichlid fishes, Nile tilapia
198 (*Oreochromis niloticus*) [23] (pT2AL-On-eGFP transgenic line, On line). We
199 successfully obtained F2 generations for both transgenic lines. Both lines can
200 drive eGFP expression in zebrafish, confirming their regulatory functions, but
201 with different patterns (Fig. 4B and Fig. 4C). The Pn line showed strong
202 expression in neurons in most regions of the brain (olfactory epithelium,

203 olfactory bulb, forebrain, optic tectum, cerebellum and hindbrain), eyes, spinal
204 cord, and heart valves. In addition, the Pn line also showed new strong eGFP
205 expression in the hatching gland in embryos, pharyngeal arches, chondrocyte
206 cells in the fin and scattered epidermal serotonin-positive cells in the skin. The
207 positive eGFP cells were further confirmed by available antibodies (Fig. 4B).
208 Interestingly, compared to the Pn line, eGFP fluroscence was weak in the brain
209 and pharyngeal arches in the On line ($p<0.01$), and was hard to detect in
210 epidermal serotonin-positive cells and fin chondrocytes (Fig. 4C). Considering
211 that we obtained different F1 individuals of each line possessing the same stable
212 eGFP expression patterns, and that the eGFP expression patterns of different
213 lines were consistent with the *in situ* hybridization and RNAseq results above,
214 the transgenic lines we constructed were stable and credible. These results
215 indicated that the cichlid fish-specific upstream region of *lg* indeed evolved new
216 regulatory functions, i.e., not only executing the ancestral expression profiles in
217 neurons but also evolving new expression patterns in the hatching gland and
218 pharyngeal arches in cichlid fishes. Strong and new regulatory functions further
219 evolved in the scattered epidermal serotonin-positive cells in the MH cichlid
220 fish lineage.
221
222 **Different regulatory roles of cichlid fish-specific upstream regions are**
223 **simply due to three MH lineage-segregated SNPs**
224 To determine whether the strong and new expression profiles in the Pn
225 transgenic line were caused by the three MH lineage-segregated SNPs in the
226 upstream region, we constructed another transgenic line with the same upstream
227 region as in the Pn line but deleted these three SNPs (pT2AL-Pn-noSNPs-eGFP
228 line, Pn-noSNPs line) (Fig. 4C, Fig. S2). Interestingly, only with these three
229 SNP deletions did the Pn-noSNPs line clearly show significantly different
230 expression patterns from the Pn line but similar patterns to the On line (Fig. 4C,
231 Fig. 4D and Table S5). These results confirmed that these SNPs are the key
232 reason for the different expression patterns between the Pn and On lines.
233

Discussion

Overall, we unravelled the evolution of a fish-specific new gene, *lg*, for the first time. No obvious phenotypic changes were detected in the *lg* gene knockout zebrafish. This can be due to compensation by other genes in the correponding gene network. This robust nature is not unexpected for immune and brain network systems [26,27], as the top DE genes of the *lg* gene knockout line are mainly neurons and immunity-related genes. In addition, lineage-specific new genes are initially usually located at the edge of the gene network [28], and their disruption thus does not greatly affect the core gene network. Regardless, mutation of *lg* does not affect zebrafish survival. Together with its multiple losses in different fish lineages during evolution, we propose that the ancestral function of *lg* is not indispensable for fish survival.

Interestingly, *lg* evolved new regulatory functions in cichlid fishes, especially in its most species-rich lineage—the MH lineage. For example, in the hatching gland, which is a key innovation in regulating the hatching period of embryos to become free-living larvae [29]; in pharyngeal cartilages, the well-known tissues related to feeding and ecological niche exploration are key to cichlid fish adaptive radiation [30–32]; in scattered epidermal serotonin-positive cells in the skin which are neuroepithelia cells related to environmental stimuli, such as by water flow, pheromones and hormones [33,34]; and in the olfactory epithelium, which is well known for the detection of odorants, such as amino acids, hormones and pheromones [35]. Given that some of these tissues are derived from a key vertebrate innovation—neural crest cells [36]—*lg* can also be causally linked to neural crest cell development and differentiation. Together, *lg* gained new regulatory functions in most species-rich lineages of cichlid fishes, the MH lineage, which can be causally linked to adaptive radiation. Noticeably, although we focused mainly on the regulatory function evolution of *lg* in cichlid fishes here, the newly important function can also evolve in their coding region. Molecular evolutionary analyses have already shown that there is one amino

265 acid mutation segregated with the MH cichlid fish lineage and that it is under
266 positive selection; this mutation might be causally linked to adaptive radiation
267 of cichlid fishes. Further gene knockout of *lg* in MH cichlid fishes will help test
268 this hypothesis and unravel the corresponding involved pathways. In addition,
269 the evolution of *lg* gene function may be ongoing in cichlid fishes, considering
270 that they are evolutionarily young species [37].
271
272 DNA in the genome that was originally thought of as junk, such as
273 transposable elements, noncoding regions and nonfunctional genes, remains
274 surprising. How can a new gene function be evolved and integrated into the
275 existing gene network? Here, we propose one hypothesis of the functional
276 evolution of *lg* that can be applied in a broad way (Fig. 5): the relaxed selection
277 pressure on the lineage-specific gene in the ancesral lineage frees it from
278 accumulating mutations in both the regulatory and the coding regions. The
279 evolution of the regulatory region in *lg* offers immediate effects, i.e., only three
280 SNP mutations in the upstream region can provide new TFs binding
281 opportunities to build new gene network connections to affect gene expression
282 profiles in the most species-rich MH lineage of cichlid fishes. It will be
283 interesting to test our predictions of the binding of a glucocorticoid hormone
284 related TF with these SNPs in the future. If this is true, these SNPs can
285 efficiently connect the organism with external environments, since
286 glucocorticoids are well-known stress-related hormones to regulate animal
287 behaviour in response to the environment [38,39]. The *lg* gene we investigated here
288 thus provides a good example of new gene function evolution in an initially
289 redundant lineage-specific gene with respect to evo-devo in a broad way.
290
291 **Methods**
292
293 **sampling**
294 Samples of cichlid fishes for comparative genomic analyses were
295 provided by Prof. Walter Salzburger in University of Basel, Switzerland; cichlid

296 fish *Pundamila nyererei* for transgene plasmids construction was obtained from
297 IGV (Interessengemeinschaft Viktoriasee-Cichliden); Strains of AB (for
298 transgenic lines construction) and ABTü (for gene knockout) zebrafish (*Danio*
299 *rerio*) were raised under standard condition (28.5°C, 14/10 h light/dark). Tilapia
300 (*Oreochromis niloticus*) were raised at 26°C, 14/10 h light/dark. Animal
301 experiments have been approved by the cantonal veterinary office in Basel
302 under permit number 2317 in University of Basel, Switzerland; the Committee
303 of Laboratory Animal Experimentation at Sun Yat-sen University and
304 Southwest University, China.
305
306 ***-omics analyses to characterize the evolution of lg across fish phylogeny***
307 To characterize the sequence evolution of *lg*, tblastx was conducted with
308 -*e* 0.001 in fishes with good genome assembly in Ensembl 100 (Table S1) and
309 local blast v2.9.0 (tblastx) was conducted with draft genomes assembly across
310 the phylogeny according to the reference [18]. After obtaining hits, the scaffolds
311 possessing the corresponding sequences were extracted to predict the coding
312 regions with Augustus 3.3.3 [40,41]. The predicted *lg* coding sequences were
313 further confirmed by aligning the translated amino acids with the available
314 sequences from Ensembl 100 using Geneious 8.0.5 [42].
315 .
316 ***-omics analyses to characterize the evolution of lg within cichlid fishes***
317 To further characterize the sequence evolution of *lg* within East African
318 cichlid fishes, we amplified the whole gene sequence and its upstream region
319 (ca.10 kb in total) in different cichlid fishes. Combined with available cichlid
320 fish genomes, we studied the sequence evolution in a total of 20 cichlid fishes
321 representing main lineages across the phylogeny [24]. Genomic DNA was
322 extracted using a DNeasy Blood & Tissue kit (Qiagen). PCR primers were
323 designed using Primer premier 5 [43]. High-fidelity PCR Master Mix (BioLabs)
324 and a touch-down annealing process were used for amplification. PGM Ion

325 Torrent was used to sequence long amplicons (400 bp) with CHIP 316 (Life
326 Technologies). Galaxy v20.09 [44] was used to trim low-quality reads (window
327 size 2, step size 1, minimum quality score 20) and filter short reads with a
328 minimum size of 40 bp. The sequence of a cichlid fish, tilapia, retrieved from
329 the Ensembl database was used as the reference to recheck the assembled
330 sequences. The sequencing reads of the *lg* gene across the cichlid fish
331 phylogeny are available at NCBI (BioProject ID PRJNA611690). The ion
332 torrent sequencing of most species was successful except *Limnochromis*
333 *abeelei*, *Ophthalmotilapia ventralis*, and *Bathybates graueri*, whose sequences
334 were further obtained from Sanger sequencing (Primers can be found in Table
335 S6).
336
337 Genomic alignment of *lg* of cichlid fishes and other teleost fishes was
338 conducted with mVISTA [45,46] using the LAGAN alignment tool [47] with *P.*
339 *nyererei* as the reference. We applied the repeat masking option with fugu as the
340 reference. Transcription factors (TFs) binding prediction analyses were
341 conducted with MatInspector v8.4.1 with the genomatix software suite
342 (https://www.genomatix.de) using default parameters with zebrafish as the reference
343 [48,49].
344
345 **Positive selection detection of *lg* in cichlid fishes**
346 To detect whether positive selection occurred in the evolution of *lg* in a
347 specific branch, a branch-site model within the codeml in PAML v3.14 was
348 performed [50,51]. Fixed branch lengths (fix_blength = 2) retrieved under the M0
349 model were used. Alignment gaps and ambiguous characters were removed
350 (Cleandata = 1). The tree for the PAML test is based on the phylogeny from the
351 study [24]. Rates of nonsynonymous to synonymous substitutions (ω or dN/dS)
352 significantly larger than 1 indicate that the sites in a specific target branch
353 (foreground branch) are under positive selection. A *chi*-squared approximation

354 was calculated to test statistical significance. Detail statistics of branch-site
355 model comparisons can be found in Table S4.
356
357 **CRISPR/Cas9 gene knockout in zebrafish**
358 The *lg* mutant line was generated by the CRISPR/Cas9 system. The guide
359 RNA (gRNA) target site was selected using the CRISPOR v4.98 [52] with a 20 bp
360 NGG PAM motif. The synthesized gRNA and Cas9 mRNA were coinjected
361 into one-cell stage zebrafish embryos at a concentration of 50 ng/μl. Ten
362 randomly selected embryos at 1 dpf (days post fertilization) were lysed and used
363 to assess functional gene knockout using PCR amplification with primers
364 designed at the target knockout site (Table S6). Heterozygous F1 fish were
365 obtained by crossing F0 and wild-type (WT) zebrafish and genotyped with the
366 same primers above using Sanger sequencing at the target site (Table S6).
367 Siblings of the F1 generation harbouring the same mutations were mated to
368 generate homozygous F2 mutants (MUT) and WT sisters (as the control).
369
370 **Western blot**
371 Since there was no available antibody for *lg*, we added an HA tag
372 (ATGTACCCATACGACGTCCCAGACTACGCT) in front of the coding
373 region of *lg* by PCR amplification (Table S6). The antibody of the HA tag can
374 thus detect the fused *lg* protein if there is no disruption. To ligate the amplicons
375 of HA-tagged coding regions of WT and MUT into the pCS2 vector, restriction
376 enzyme sites (*Bam*H I and *Xba* I) were added at the two ends of the amplicons
377 by PCR amplification (Table S6), followed by a double digestion system (NEB)
378 and ligation (Promega) (Table S6). The plasmids pCS2, pCS2-lg-HAtag-WT
379 and pCS2-lg-HAtag-MUT (2 μg) (Fig. S2) were then transfected into the
380 HEK293T cell line (ab255449, Abcam) separately. Protein was collected after
381 36 h of transfection to perform Western blot according to the standard
382 procedure [53]. Mouse anti-HA (901501 Biolegend, 1:2000) and mouse anti-

383 GAPDH (60004-1-lg Proteintech, 1:2000) antibodies were used for detection.
384 Detections were performed with goat anti-mouse secondary antibody (31430
385 Thermo 1:5000) and enhanced chemiluminescence (ECL plus) reagent.
386
387 **RNAseq analyses**
388 To determine the gene expression profiles of *lg*, the available raw
389 transcriptomic data of different tissues in different fishes were extracted from
390 the SRA database (Table S2). The raw reads were mapped to the corresponding
391 transcriptome retrieved from Ensembl using Hisat2-2.1.0 [54] with default
392 parameters. The reported SAM results were converted into BAM format, and
393 finally recorded as count files using SAMtools v1.10 [55] followed by TPM value
394 calculation [56].
395
396 To identify the affected genes after knockout, comparative transcriptomic
397 analyses were conducted between MUT and WT (Table S3). RNA extraction
398 was performed (each replicate contained 30 pooled individuals) using TRIzol®
399 reagent (Invitrogen). Six biological replicates of WT and MUT of the F2
400 generation were used for RNAseq analyses. Library construction and
401 sequencing were performed using Illumina TrueSeq RNAseq technology at the
402 Novogene Company, China, with a paired end read length of 350 bp. The raw
403 reads were mapped to the zebrafish transcriptome assembly using the same
404 methods mentioned above to obtain the read count files. The differentially
405 expressed (DE) transcripts between WT and MUT were further analysed using
406 the edgeR v3.20.7 package (FDR<0.01). The raw data were deposited at NCBI
407 BioProject ID PRJNA611476. The top 10 up- and downregulated DE transcripts
408 lists can be found in Table S3.
409
410 ***In situ* hybridization**

411 The gene fragments in the coding region of *lg* in zebrafish and tilapia
412 were amplified using PrimerSTAR (Takara) (Table S6) and *lyz* probe synthesis
413 was conducted as described in the reference [57]. The amplified products were
414 cloned into the pGEM®-T Easy vector (Promega). Plasmids were extracted and
415 purified using a TIANprep Mini Plasmid Kit II (Tiangen). The insertion and
416 direction of probes were confirmed with Sanger sequencing using M13 PCR
417 products (Tsingke). RNA probes were synthesized with the DIG RNA labelling
418 kit (SP6/T7) (Roche). The experiment was conducted as described in the
419 reference [58] with biological replicates (n=30 for zebrafish, n=20 for tilapia, and
420 n=10 for *P. nyererei*). The embryos or larvae were mounted in 1% low melting
421 point agarose (GBCBIO) and imaged with a Zeiss Axio Imager Z2 microscope.
422
423 **Transgenic plasmids and transgenic lines construction**
424 To determine the regulatory functions of cichlid fish-specific upstream
425 regions, we constructed plasmids with different upstream regions as promoters
426 to drive eGFP expression in zebrafish (Fig. 4A and Fig. S2), i.e., 1) species in
427 the MH lineage (*P. nyererei*) (Pn line); and 2) the corresponding region of
428 species in the sister lineage of East African Great Lakes cichlid fishes (tilapia)
429 (On line). 3) To determine whether the three MH-segregated SNPs in the cichlid
430 fish-specific upstream region are important, we specifically deleted these SNPs
431 using site-directed mutagenesis by PCR primer extension (Table S6) [59] to
432 construct another transgenic line (Pn-noSNPs line). A schematic of the plasmid
433 construction can be found in Fig. S2.
434
435 DNA fragments for different plasmid constructs were amplified with
436 PCR primers supplemented with the corresponding restriction endonuclease
437 sites (Fig. S2, Table S6). Fragments were then recovered (Tiangen), digested
438 (NEB) and ligated (Promega) into the pT2AL vector following the instructions
439 accordingly. Successfully constructed recombinant plasmids were further

440 confirmed by Sanger sequencing (Tsingke). Injections were performed into the
441 one-cell stage of zebrafish embryos. By outcrossing with WT zebrafish, we
442 successfully obtained F2 stable transgenic lines for all these lines (Pn line, On
443 line and Pn-noSNPs line). The fluroscence signals were imaged with a Zeiss
444 880 confocal microscope.
445
446 **Immunohistochemistry (IHC)**
447 The following primary antibodies were used: mouse anti-Pvalb7, 1:1000
448 provided by Prof. Masahiko Hibi, Nagoya University, Japan [60]; rabbit anti-
449 serotonin, 1:400 (s5545, Sigma); rabbit anti-calretinin, 1:400 (ab702, Abcam);
450 goat anti-GFP, 1:400 (ab6673, Abcam); mouse anti-HuC/HuD, 1:1000 (16A11,
451 Invitrogen); rabbit anti-SOX10, 1:400 (ab229331, Abcam); rabbit anti-PAX6
452 1:400 (ab209836, Abcam); mouse anti-collagen type II (II-II6B3, DSHB); and
453 rabbit anti-cathepsin L/MEP (ab209708, Abcam). The secondary antibodies
454 were donkey to goat IgG H&L (Alexa Fluor® 488) (ab150133, Abcam),
455 donkey to mouse IgG H&L (Alexa Fluor® 647) (ab150107, Abcam), donkey to
456 rabbit IgG H&L (Alexa Fluor® 647) (ab150063, Abcam) and DAPI (ab228549,
457 Abcam). Fish larvae were fixed in 4% paraformaldehyde (Sigma) for 4 h at
458 room temperature, followed by a detailed protocol as described previously [61].
459 The fluroscence signals were imaged with a Zeiss 880 confocal microscope.
460
461 **Immune response detection**
462 To determine whether there are different immune responses between
463 MUT and WT zebrafish, we injected lipopolysaccharide (LPS), the key
464 component of Gram-negative bacteria, into the circulation of 2 dpf zebrafish
465 and counted the numbers of *lyz*$^+$ neutrophils in the caudal haematopoietic tissue
466 at one day post injection. The *lyz*$^+$ whole-mount RNA *in situ* hybridization
467 signals across eight somites of the larval CHT (caudal haematopoietic tissue)
468 regions were manually counted. Statistical analyses can be found in Table S7.

469

470 **Data Availability**

471

472 The publicly available datasets that this study used are provided in the
473 supplementary information. The RNAseq data of WT and MUT are available at
474 NCBI (BioProject ID PRJNA611476). The ion torrent sequencing reads of the
475 *lg* gene are available at NCBI (BioProject ID PRJNA611690). The scaffolds
476 possessing *lg* of teleost fishes across the phylogeny, the assembly of *lg*
477 including the upstream region of cichlid fishes across the phylogeny retrieved
478 from ion torrent sequencing and Sanger sequencing can be found in Dryad
479 https://doi.org/10.5061/dryad.fxpnvx0qb. Source associated data of Fig. 3 and Fig. 4
480 are provided in the supplementary files. Other data are available upon
481 reasonable request.

482

483 **References**

484

485 1. Dobzhansky, T. Nothing in biology makes sense except in the light of
486 evolution. *Am. Biol. Teach.* **March,** 125–129 (1973).
487 2. Tautz, D. & Domazet-Lošo, T. The evolutionary origin of orphan genes.
488 *Nat. Rev. Genet.* **12,** 692–702 (2011).
489 3. Bae, B.-I., Jayaraman, D. & Walsh, C. A. Genetic changes shaping the
490 human brain. *Dev. Cell* **32,** 423–34 (2015).
491 4. Zhang, Y. E. & Long, M. New genes contribute to genetic and phenotypic
492 novelties in human evolution. *Curr. Opin. Genet. Dev.* **29,** 90–6 (2014).
493 5. Babonis, L. S., Martindale, M. Q. & Ryan, J. F. Do novel genes drive
494 morphological novelty? An investigation of the nematosomes in the sea
495 anemone Nematostella vectensis. *BMC Evol. Biol.* **16,** 114 (2016).
496 6. Santos, M. E., Le Bouquin, A., Crumière, A. J. J. & Khila, A. Taxon-
497 restricted genes at the origin of a novel trait allowing access to a new
498 environment. *Science (80-.).* **358,** 386–390 (2017).
499 7. Khalturin, K. *et al.* A novel gene family controls species-specific
500 morphological traits in Hydra. *PLoS Biol.* **6,** e278 (2008).
501 8. Garcia-Fernàndez, J. The genesis and evolution of homeobox gene
502 clusters. *Nat. Rev. Genet.* **6,** 881–92 (2005).
503 9. Wu, M., Chen, G. & Li, Y.-P. TGF-β and BMP signaling in osteoblast,
504 skeletal development, and bone formation, homeostasis and disease. *Bone*

505 *Res.* **4,** 16009 (2016).
506 10. Wagner, G. P. & Lynch, V. J. The gene regulatory logic of transcription
507 factor evolution. *Trends Ecol. Evol.* **23,** 377–85 (2008).
508 11. Wray, G. A. The evolutionary significance of cis-regulatory mutations.
509 *Nat. Rev. Genet.* **8,** 206–216 (2007).
510 12. Zhang, J. Parallel adaptive origins of digestive RNases in Asian and
511 African leaf monkeys. *Nat. Genet.* **38,** 819–23 (2006).
512 13. Hoekstra, H. E., Hirschmann, R. J., Bundey, R. A., Insel, P. A. &
513 Crossland, J. P. A single amino acid mutation contributes to adaptive
514 beach mouse color pattern. *Science* **313,** 101–4 (2006).
515 14. Chen, L., DeVries, A. L. & Cheng, C.-H. C. Evolution of antifreeze
516 glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish.
517 *Proc. Natl. Acad. Sci.* **94,** 3811–3816 (1997).
518 15. Carroll, S. B. Evolution at Two Levels: On Genes and Form. *PLoS Biol.* **3,**
519 e245 (2005).
520 16. Hu, W., Chen, S., Zhang, R. & Lin, Y. Single nucleotide polymorphisms
521 in the upstream regulatory region alter the expression of myostatin. *Vitr.*
522 *Cell. Dev. Biol. - Anim.* **49,** 417–423 (2013).
523 17. Marceca, C., Pfob, M., Schelling, G., Steinlein, O. K. & Eggert, M. Single
524 nucleotide polymorphism creating a variable upstream open reading frame
525 regulates glucocorticoid receptor expression. *Gene* **563,** 24–28 (2015).
526 18. Malmstrøm, M. *et al.* Evolution of the immune system influences
527 speciation rates in teleost fishes. *Nat. Genet.* **48,** 1204–1210 (2016).
528 19. Salzburger, W. The interaction of sexually and naturally selected traits in
529 the adaptive radiations of cichlid fishes. *Mol. Ecol.* **18,** 169–85 (2009).
530 20. SH, O. & LI, Z. Hematopoiesis: An Evolving Paradigm for Stem Cell
531 Biology. *Cell* **132,** 631–44 (2008).
532 21. Kim, C.-H. *et al.* Zebrafish elav/HuC homologue as a very early neuronal
533 marker. *Neurosci. Lett.* **216,** 109–112 (1996).
534 22. Hugo, S. E. *et al.* A monocarboxylate transporter required for hepatocyte
535 secretion of ketone bodies during fasting. *Genes Dev.* **26,** 282–93 (2012).
536 23. Salzburger, W. Understanding explosive diversification through cichlid
537 fish genomics. *Nat. Rev. Genet.* (2018). doi:10.1038/s41576-018-0043-9
538 24. Meyer, B. S., Matschiner, M. & Salzburger, W. A tribal level phylogeny
539 of Lake Tanganyika cichlid fishes based on a genomic multi-marker
540 approach. *Mol. Phylogenet. Evol.* **83,** 56–71 (2015).
541 25. Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. Out of Tanganyika:
542 genesis, explosive speciation, key-innovations and phylogeography of the
543 haplochromine cichlid fishes. *BMC Evol. Biol.* **5,** 17 (2005).
544 26. Hubel, P. *et al.* A protein-interaction network of interferon-stimulated
545 genes extends the innate immune system landscape. *Nat. Immunol.* **20,**
546 493–502 (2019).
547 27. Denève, S., Alemi, A. & Bourdoukan, R. The Brain as an Efficient and

Robust Adaptive Learner. *Neuron* **94,** 969–977 (2017).
28. Zhang, W., Landback, P., Gschwend, A. R., Shen, B. & Long, M. New genes drive the evolution of gene interaction networks in the human and mouse genomes. *Genome Biol.* **16,** 202 (2015).
29. Nagasawa, T. *et al.* Evolutionary Changes in the Developmental Origin of Hatching Gland Cells in Basal Ray-Finned Fishes. *Zoolog. Sci.* **33,** 272 (2016).
30. Conith, A. J., Kidd, M. R., Kocher, T. D. & Albertson, R. C. Ecomorphological divergence and habitat lability in the context of robust patterns of modularity in the cichlid feeding apparatus. *BMC Evol. Biol.* **20,** 95 (2020).
31. Roberts, R. B., Hu, Y., Albertson, R. C. & Kocher, T. D. Craniofacial divergence and ongoing adaptation via the hedgehog pathway. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 13194–9 (2011).
32. Albertson, R. C., Streelman, J. T., Kocher, T. D. & Yelick, P. C. Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 16287–92 (2005).
33. König, D. *et al.* Distribution and Restoration of Serotonin-Immunoreactive Paraneuronal Cells During Caudal Fin Regeneration in Zebrafish. *Front. Mol. Neurosci.* **12,** 227 (2019).
34. Coccimiglio, M. L. & Jonz, M. G. Serotonergic neuroepithelial cells of the skin in developing zebrafish: morphology, innervation and oxygen-sensitive properties. *J. Exp. Biol.* **215,** 3881–3894 (2012).
35. Kermen, F., Franco, L. M., Wyatt, C. & Yaksi, E. Neural circuits mediating olfactory-driven behavior in fish. *Front. Neural Circuits* **7,** 62 (2013).
36. Green, S. A., Simoes-Costa, M. & Bronner, M. E. Evolution of vertebrates as viewed from the crest. *Nature* **520,** 474–82 (2015).
37. Salzburger, W. Understanding explosive diversification through cichlid fish genomics. *Nat. Rev. Genet.* **19,** 705–717 (2018).
38. Nyman, C., Fischer, S., Aubin-Horth, N. & Taborsky, B. Evolutionary conserved neural signature of early life stress affects animal social competence. *Proceedings. Biol. Sci.* **285,** (2018).
39. Culbert, B. M., Gilmour, K. M. & Balshine, S. Stress axis regulation during social ascension in a group-living cichlid fish. *Horm. Behav.* **103,** 121–128 (2018).
40. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33,** W465–7 (2005).
41. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.* **32,** W309–12 (2004).

591 42. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop
592 software platform for the organization and analysis of sequence data.
593 *Bioinformatics* **28,** 1647–9 (2012).
594 43. Lalitha, S. Primer Premier 5. *Biotech Software&Internet Rep.* **1,** (2000).
595 44. Afgan, E. *et al.* The Galaxy platform for accessible, reproducible and
596 collaborative biomedical analyses: 2018 update. *Nucleic Acids Res.* **46,**
597 W537–W544 (2018).
598 45. Mayor, C. *et al.* VISTA : visualizing global DNA sequence alignments of
599 arbitrary length. *Bioinformatics* **16,** 1046–1047 (2000).
600 46. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I.
601 VISTA: computational tools for comparative genomics. *Nucleic Acids*
602 *Res.* **32,** W273–9 (2004).
603 47. Brudno, M. *et al.* LAGAN and Multi-LAGAN: efficient tools for large-
604 scale multiple alignment of genomic DNA. *Genome Res.* **13,** 721–31
605 (2003).
606 48. Quandt, K., Frech, K., Karas, H., Wingender, E. & Werner, T. MatInd and
607 MatInspector: new fast and versatile tools for detection of consensus
608 matches in nucleotide sequence data. *Nucleic Acids Res.* **23,** 4878–84
609 (1995).
610 49. Cartharius, K. *et al.* MatInspector and beyond: promoter analysis based on
611 transcription factor binding sites. *Bioinformatics* **21,** 2933–2942 (2005).
612 50. Yang, Z. PAML: a program package for phylogenetic analysis by
613 maximum likelihood. *Comput. Appl. Biosci.* **13,** 555–6 (1997).
614 51. Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol.*
615 *Biol. Evol.* **24,** 1586–91 (2007).
616 52. Concordet, J.-P. & Haeussler, M. CRISPOR: intuitive guide selection for
617 CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids*
618 *Res.* **46,** W242–W245 (2018).
619 53. Zhao, F. *et al.* Irf8 regulates the progression of myeloproliferative
620 neoplasm-like syndrome via Mertk signaling in zebrafish. *Leukemia* **32,**
621 149–158 (2018).
622 54. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner
623 with low memory requirements. *Nat. Methods* **12,** 357–360 (2015).
624 55. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools.
625 *Bioinformatics* **25,** 2078–9 (2009).
626 56. Wagner, G. P., Kin, K. & Lynch, V. J. Measurement of mRNA abundance
627 using RNA-seq data: RPKM measure is inconsistent among samples.
628 *Theory Biosci.* **131,** 281–285 (2012).
629 57. Hou, Y. *et al.* Systemic inoculation of Escherichia coli causes emergency
630 myelopoiesis in zebrafish larval caudal hematopoietic tissue. *Sci. Rep.* **6,**
631 36853 (2016).
632 58. Westerfield, M. *The zebrafish book. A guide for the laboratory use of*
633 *zebrafish (Danio rerio). 4th edition.* (University of Oregon Press, Eugene,

2000).
59. Reikofski, J. & Tao, B. Y. Polymerase chain reaction (PCR) techniques for site-directed mutagenesis. *Biotechnol. Adv.* **10,** 535–47 (1992).
60. Nimura, T. *et al.* Role of Reelin in cell positioning in the cerebellum and the cerebellum-like structure in zebrafish. *Dev. Biol.* **455,** 393–408 (2019).
61. Barresi, M. J. F., Stickney, H. L. & Devoto, S. H. The zebrafish slow-muscle-omitted gene product is required for Hedgehog signal transduction and the development of slow muscle identity. *Development* **127,** 2189–2199 (2000).

Acknowledgements

Many thanks to Prof. Walter Salzburger, Attila Rüegg, Zhen Liu, Zhiqiang Zhou and Hesheng Xiao for the sampling and suggestions. We thank Prof. Masahiko Hibi and Prof. Kiyoshi Naruse for providing the antibody Pvalb7. We also thank Prof. Lingfei Luo, Prof. Hua Ruan, Prof. Yang Liu and Prof. Tienming Lee for the laboratory supports. Many thanks to Prof. Zuzana Musilova, Dr. Xiaogui Yi, Dr. Yandong Zhan, Dr. Fangying Zhao, Yuepeng He, Xinkai Cheng, Qing Li and Yue Gao for the supports. Thanks to Huansheng Tao and Jialing Xu for the Confocol microscope support, Zhao Long for fluroscent microscope support. This study was funded by grants from the National Natural Science Foundation of China (31802283), Guangdong Basic and Applied Basic Research Foundation (2020A1515010330); the Fundamental Research Funds for the Central Universities (20lgpy109); China Postdoctoral Science Foundation (2018 M633311) and Doctoral Scholarship of University of Basel, Switzerland to LG; Grants from the National Natural Science Foundation of China (31630082) to DW, and (31771623 and 31571500) to LL. Part of -omics analyses was supported by National Supercomputer Center in Guangzhou, China. The English language of this article was edited by Elsevier Language Editing Services.

Authors contributions

666

667 LG, DW and LL conceived this work. LG carried out –omics sequencing
668 and evolutionary analyses. LG, CL, XM and ZW conducted the *in situ*
669 hybridization. LG carried out the IHC. XH and WZ participated the IHC
670 experiments of serotonin-positive cells and olfactory epithelium cells. LG, CL
671 and XM constructed plasmids of the Pn line and the On line. LG constructed
672 plasmids of the Pn-noSNPs line. LG leads the transgenic lines and mutant lines
673 construction with the help from CL and XM. LG took the confocal photos. YH
674 performed the Western blot experiment. ZW carried out the LPS immune
675 response experiments. LG conducted the fasting ORO staining experiments. LG
676 wrote the initial manuscript draft, and all authors commented on the manuscript
677 and approved the final version.

678

679 **Competing interests**: The authors declare no competing interests.

680

681 **Figure legends**

682

683 **Fig. 1 Sequence evolution of *lg*.** a. Comparative genomic analyses showed that
684 *lg* is a fish-specific gene. b. Multiple losses occurred in different lineages within
685 teleost fishes. The reference tree is from [18].

686

687 **Fig. 2 Expression profiles of *lg* in fishes.** a. Transcriptomic analyses of *lg* gene
688 expression patterns in adult fish tissues. *lg* was mainly expressed in the brain
689 and eyes in most fishes but evolved new expression patterns in the skin in the
690 modern haplochromine (MH) lineage of cichlid fishes. b. *In situ* hybridization
691 in zebrafish and cichlid fish larvae. *lg* was mainly expressed in the brain,
692 ganglion and inner plexiform cell layers in the eyes of zebrafish. No expression
693 was detected in the hatching gland of 24 hpf (hours post fertilization) zebrafish
694 embryos. *lg* was highly expressed in the brain, eyes and pharyngeal arches in

695 the MH cichlid fish, *Pundamilia nyererei*, but weakly expressed in the cichlid
696 fish *Oreochromis niloticus*.
697
698 **Fig. 3 *lg* gene knockout in zebrafish and comparisons between mutant**
699 **(MUT) and wild-type (WT) zebrafish.** A. *lg* is successfully disrupted using
700 CRISPR/Cas9. This was confirmed by both Sanger sequencing (top) and
701 Western blot (bottom). B. Immune response comparisons between MUT and
702 WT using whole mount *in situ* hybridization in CHT (caudal haematopoietic
703 tissue) treated with LPS and PBS (n=10 for WT, and n=18 for MUT),
704 respectively. Quantification of the numbers of lyz^+ signals across eight somites
705 is shown as dots. n.s. not significant. Mean and SD error bars were shown. C.
706 Phenotype comparisons between MUT and WT zebrafish. Brain and eye
707 phenotypes in mutants were examined with the neuron marker HuC. No
708 obvious phenotype was found. Under fasting condition, no obvious mutant
709 phenotype was detected with ORO staining using the same strategy as described
710 in the literature [22].
711
712 **Fig. 4 Sequence evolution of *lg* across East African cichlid fish phylogeny**
713 **and regulatory function evolution in cichlid fishes.** Aa. Both the upstream
714 and coding regions exhibited SNPs segregated with the modern haplochromine
715 (MH) lineage, i.e., three SNPs (GTA) in the upstream region and one
716 nonsynonymous amino acid change (A to S) located at the 38th position was
717 under positive selection (Table S4). Transcription factors binding sites
718 prediction showed that these SNPs can set up the opportunity to bind to
719 glucocorticoid modulatory element binding protein 2 (GMEB2). The reference
720 tree is from literature [24]. Ab. Three types of transgenic plasmids construction.
721 The pT2AL-On-eGFP line (On line) adopted the upstream region of
722 *Oreochromis niloticus* as the promoter to drive eGFP expression. The pT2AL-
723 Pn-eGFP line (Pn line) used the homologous upstream region of *Pundamilia*

724 *nyererei* in the MH lineage as the promoter. The pT2AL-Pn-no-SNPs-eGFP line
725 (Pn-noSNPs line) used the same promoter as the Pn line but deleted the three
726 MH-segregated SNPs (GTA). B. The fluorescence images in the Pn line.
727 Merged figures of double staining of eGFP and cell marker antibodies are
728 shown if there are available antibodies. Red is from the cell marker antibody.
729 Green is from eGFP. Blue is from DAPI. Pictures were taken from 4 dpf (days
730 post fertilization) zebrafish except the hatching gland which is from 24 hpf
731 (hours post fertilization) embryos. C. Different expression patterns of three
732 transgenic lines in zebrafish. D. Statistics of expression pattern comparisons
733 among the three transgenic lines, focusing on the cerebellum, hindbrain and the
734 first two pharyngeal arches. The Pn line showed significantly stronger
735 fluroscence than the other two lines. With the three MH-segregated SNP
736 deletions, the Pn-noSNPs line showed weak fluroscence expressions, similar to
737 the On line. Mean and SD error bars were shown. n=6 for each measurement.
738 Detailed statistical analyses can be found in Table S5.
739
740 **Fig. 5 Hypothesis of *lg* sequence and function evolution with respect to evo-**
741 **devo.** *lg* is a fish-specific gene. Its ancestral function is not indispensable. In
742 cichlid fishes, *lg* evolves a cichlid fish-specific regulatory region, thus gained
743 new regulatory functions to drive gene expression in cells related to adaptation,
744 such as in pharyngeal cartilages and the hatching gland. Especially in the most
745 species-richness lineage, the modern haplochromine (MH) lineage, this
746 upstream can not only drive strong eGFP expression patterns, but also further
747 evolved new profiles in epidermal serotonin-positive cells. This can be simply
748 due to the three MH-segregated SNPs in the upstream region that are predicted
749 to bind to a hormone related transcription factor, the glucocorticoid modulatory
750 element binding protein 2 (GMEB2). If this is true, it opens a new opportunity
751 to response to the external environemnt. In addition, the coding region of *lg* also
752 evolved one MH- segregated amino acid mutation that is under positive

selection, indicating that new function might also evolved in the coding region which deserves further investigation. *lg* thus provides a good example of new gene function emergence in a redundant lineage-specific gene with respect to evo-devo.

Figures

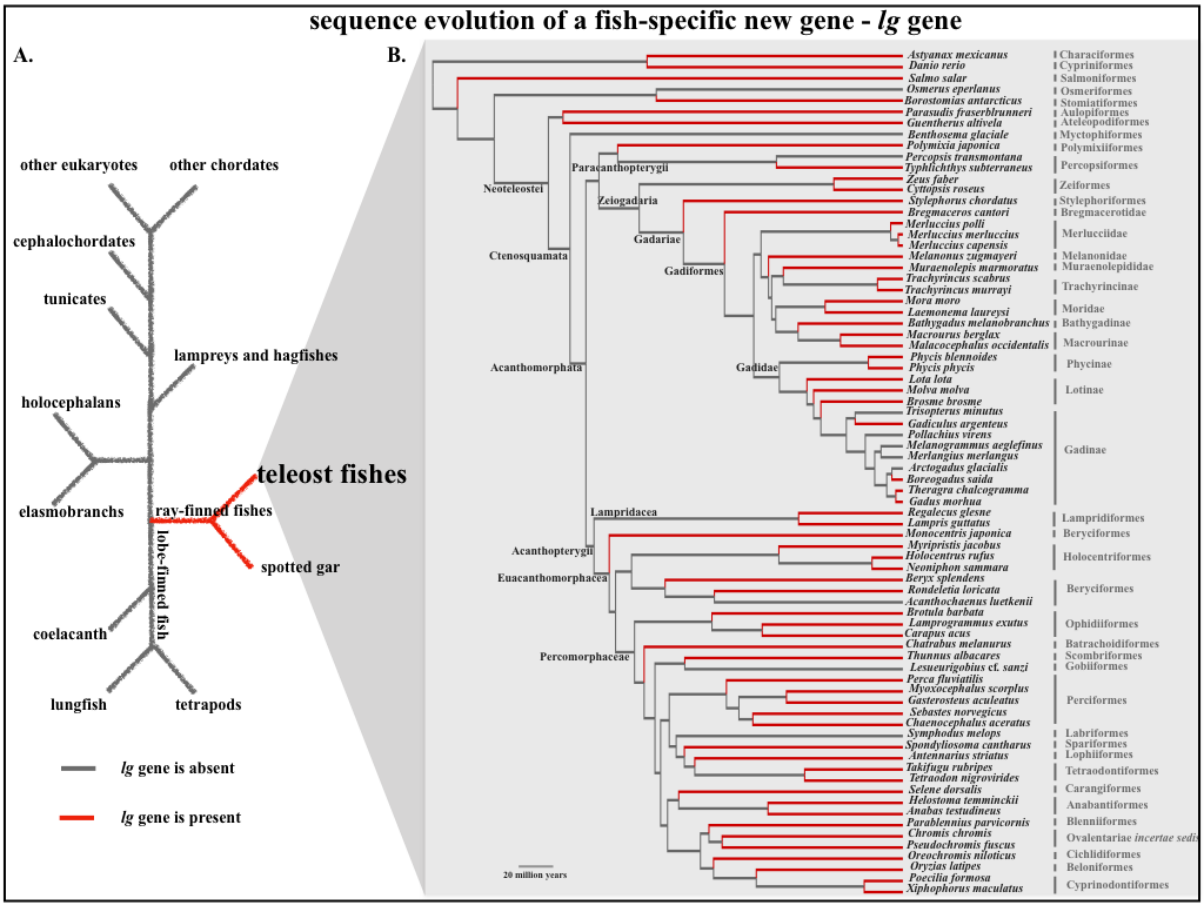

Fig. 1 Sequence evolution of *lg*.

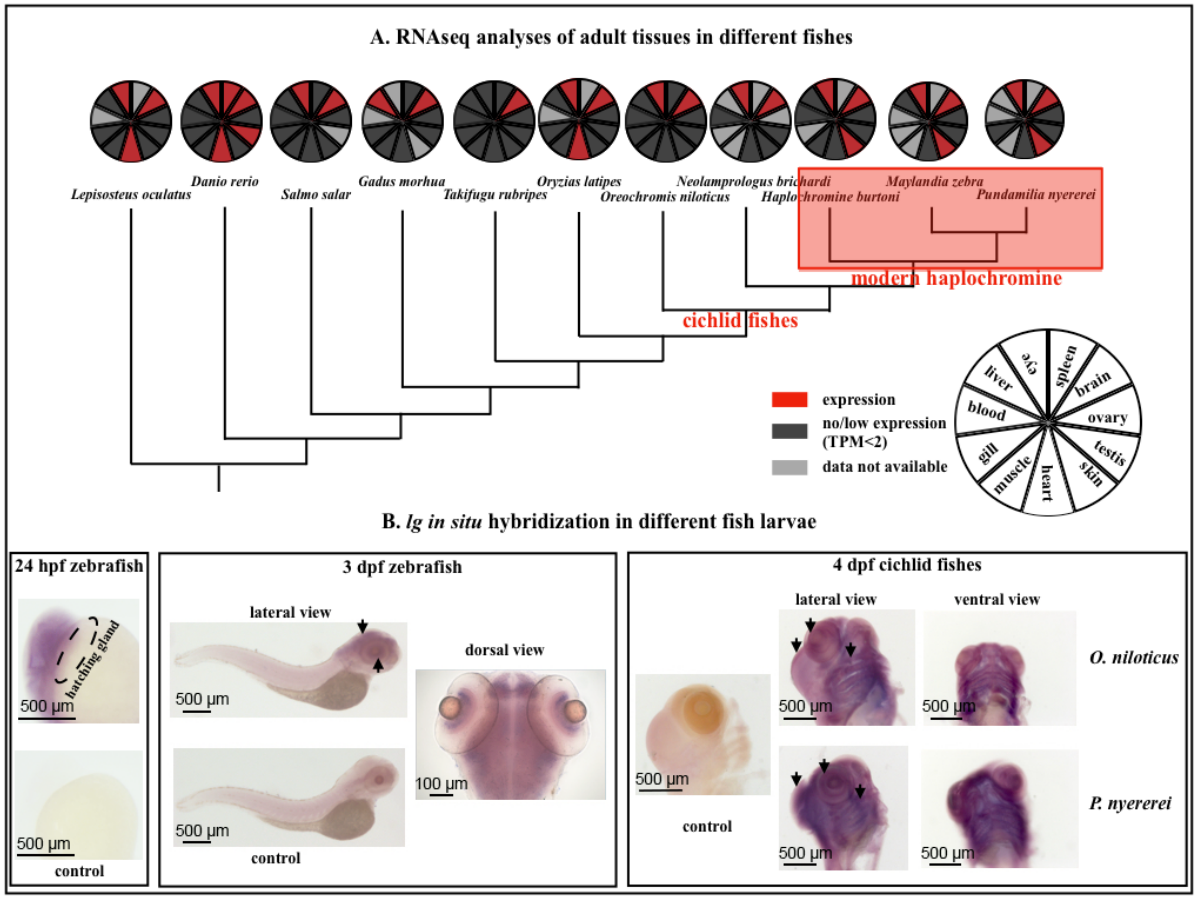

Fig. 2 Expression profiles of *lg* in fishes.

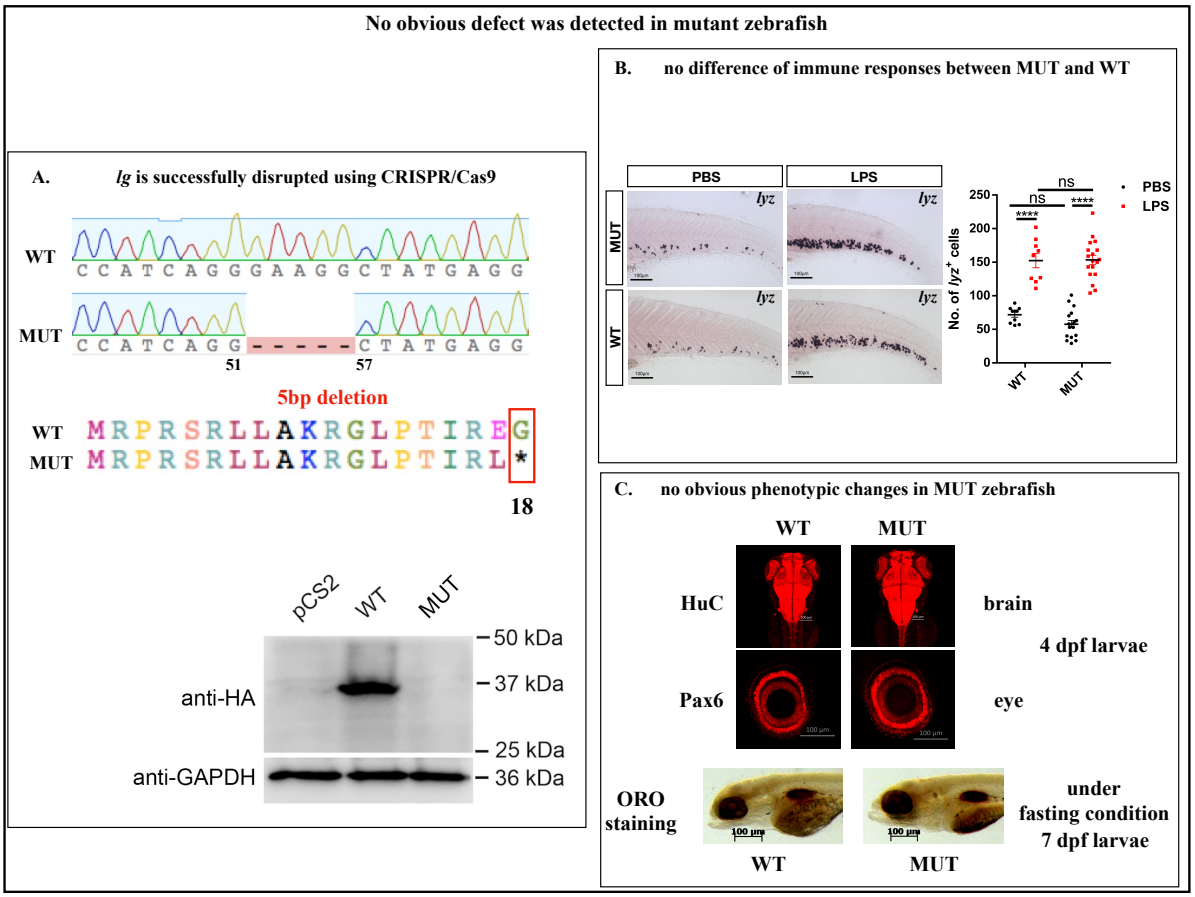

Fig. 3 *lg* gene knockout in zebrafish and comparisons between mutant (MUT) and wild-type (WT) zebrafish.

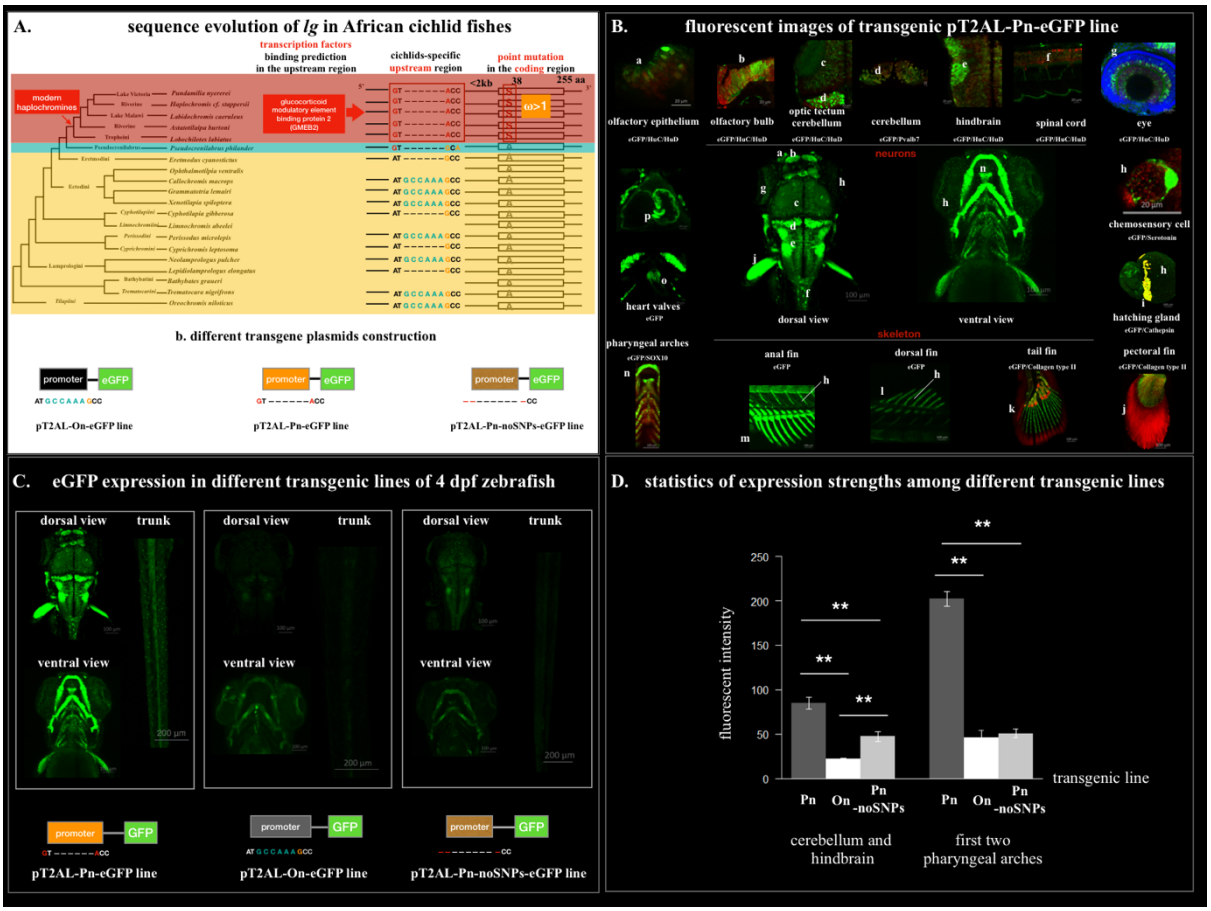

Fig. 4 Sequence evolution of *lg* across East African cichlid fish phylogeny and regulatory function evolution in cichlid fishes.

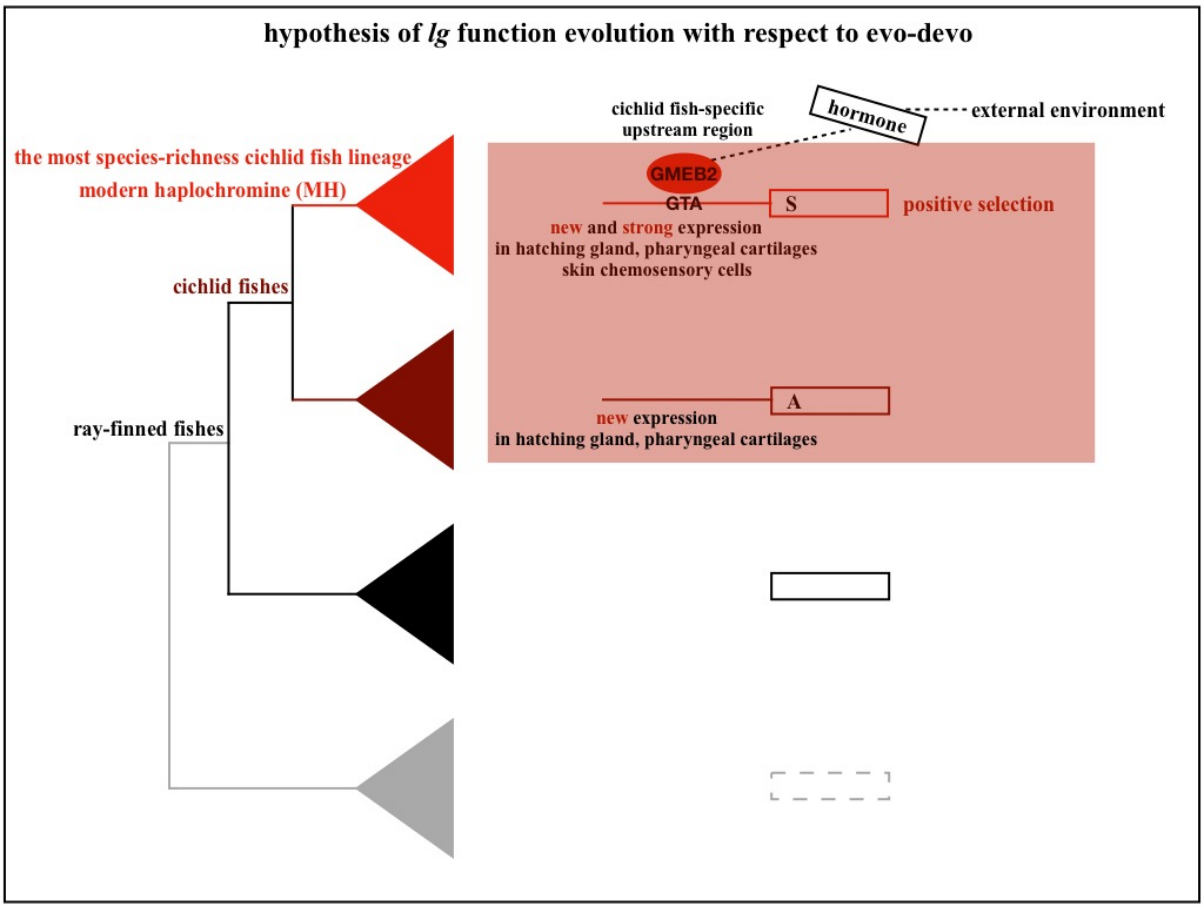

Fig. 5 Hypothesis of *lg* sequence and function evolution with respect to evo-devo.

Supplementary information

Table S1 *lg* appearance in species with available genomes assembly in Ensembl

Table S2 RNAseq analyses of *lg* in different fishes with available data

Table S3 Top differentially expressed transcripts between MUT and WT zebrafish

Table S4 Positive selection detection of *lg* in cichlid fishes within codeml in PAML

Table S5 eGFP fluroscence strength comparisons among different transgenic lines

Table S6 PCR primers and the corresponding annealing temperatures

787 Table S7 The numbers of LPS immune response signal comparisons between
788 WT and MUT zebrafish
789 Supplementary Figures