Revelation of the Genetic Basis for Convergent Innovative Anal Fin Pigmentation Patterns in Cichlid Fishes

Langyu Gu[*1,2], Chao Zhang[3], Canwei Xia[4]

[1] Zoological Institute, University of Basel, Vesalgasse 1, 4051, Basel, Switzerland.

[2] Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education, Laboratory of Aquatic Science of Chongqing, School of Life Sciences, 400715, Southwest University, Chongqing, China.

[3] Centre for Integrative Bioinformatics Department of Computer Science, VU Amsterdam, De Boelelaan 1081A, 1081HV Amsterdam, the Netherlands.

[4] Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China.

Author for Correspondence:

Langyu Gu,

Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education, Laboratory of Aquatic Science of Chongqing, School of Life Sciences, 400715, Southwest University, Chongqing, China.

Tel: 008618513056804

email address: langyugu@gmail.com

Abstract

Determining whether convergent novelties share a common genetic basis is vital to understanding the extent to which evolution is predictable. The convergent evolution of innovative anal fin pigmentation patterns in cichlid fishes is an ideal model for studying this question. Here, we focused on two patterns: 1) egg-spots, circular pigmentation patterns in the haplochromine lineage with different numbers, sizes and positions; and 2) the blotch, with limited variation and an irregular boundary independently evolved in the ectodine lineage. However, how these two novelties originate and evolve still remains unclear. Based on a thorough comparative transcriptomic and genomic analysis, we observed a common genetic basis (i.e. genes related to pigment cells, signaling pathways and metabolic pathways) with high evolutionary rates and similar expression levels between egg-spots and the blotch. Furthermore, associations of common genes with transcription factors and the integration of advantageous signalling pathway genes with accelerated evolutionary rates were observed for the egg-spots gene network. Thus, we propose that the re-use of the common genetic basis indicates important conservative functions (e.g., toolkit genes) for the origin of these convergent novel phenotypes, whereas independently evolved associations of common genes with transcription factors and the integration of advantageous genes free the evolution of egg-spots to evolve as a key innovation for the adaptive radiation of cichlid fishes. This hypothesis will further illuminate the mechanism of the origin and evolution of novelties in a broad sense.

Key words

convergent evolution, novelty, gene network, egg-spots, the blotch, cichlid fishes

Introduction

How evolutionary novelties evolve remains an open question in evolutionary biology (Annona et al. 2015; Clark-Hachtel & Tomoyasu 2016; Soltis & Soltis 2016; Yong & Yu 2016). Such novelties provide the raw materials for downstream selection, thereby contributing to biological diversification (Pigliucci & Müller 2010). Examples of evolutionary novelties are neural crest cells in vertebrates (Shimeld & Holland 2000), the beaks of birds (Bhullar et al. 2015; Bright et al. 2016), and eyespots on the wings of nymphalid butterflies (Monteiro 2015). Evolutionary novelty is generally defined in the context of character homology, e.g., as "*a structure that is neither homologous to any structure in the ancestral species nor serially homologous to any part of the same organism*" (Müller & Wagner 1991). However, the inference of homology is not always straightforward (Bang et al. 2002; Cracraft 2005; Panchen 2007; Hall 2013; Faunes et al. 2015). Wagner (Wagner 2007) thus proposed that homology should be inferred using information from the gene network underlying a trait. In this case, characters would be homologous if they shared the same underlying core gene network, whereas novelty would involve the evolution of a quasi-independent gene network, which integrates signals (e.g., input signals and effectors) into a gene expression pattern unique to that character (Wagner 2014). To what extend a gene network is "innovative" compared to the ancestral gene network, and how this gene network affects the evolution of a trait are interesting questions that remain unanswered.

In addition, similar traits can independently evolve in two or more lineages, such as the convergent evolution of echolocation in bats and whales (Shen et al. 2012) and anal fin pigmentation patterns in cichlid fishes (see below) (Salzburger et al. 2007; Santos et al. 2016). Whether convergent phenotypes result from the same gene (e.g., *Mc1r* in mammals

(Hoekstra et al. 2006) and birds (San-Jose et al. 2015)) or not (e.g., anti-freezing protein genes (Chen et al. 1997) in Antarctic notothenioid and Arctic cod)), and the extent to which evolution is predictable (Stern 2013) have long been discussed, but no agreement has been reached (Arendt et al. 2008; Stern 2013). Actually, convergent phenotypes can partially share a common genetic basis, for example, independently re-deploy conserved toolkit genes (wing patterns in Heliconius butterfly (Joron et al. 2006) and the repeated emergence of yeasts (Nagy et al. 2014)), or different genes within the same pathway (Berens et al. 2015). Alternatively, these phenotypes can also be derived from deep homology based on ancestral structure (Shubin et al. 2009). Therefore, instead of simply focusing on whether convergent phenotypes are derived from the same genes, it is important to answer the following questions: 1) To what extent does the common genetic basis apply to convergent evolution? 2) Among the genes expressed in independently evolved traits, which genes are conserved? 3) What are the roles of these genes in the evolution of convergent phenotypes? 4) How did these genes evolve in independent lineages, i.e., are they deeply homologized or independently recruited? The answers to these questions at both the individual gene and gene network levels will give a clue about the genetic basis of the origin and evolution of novelty.

East African cichlid fishes, exposed to explosive radiation within millions and even hundreds of thousands of years, are classical evolutionary model species (Kocher 2004; Salzburger 2009). The relatively close genetic background (Salzburger 2009; Brawand et al. 2014) of these fishes and availability of genomics data (Brawand et al. 2014) provide powerful tools to study the relationship between phenotype and genotype. Many convergent phenotypes have been identified in cichlid fishes, such as thick lip (Colombo et al. 2013), lower pharyngeal jaw (Muschick et al. 2012), and the convergent evolution of anal fin pigmentation patterns, a new model to study the origin of evolutionary novelty (Salzburger et

al. 2007; Santos et al. 2016). Two anal fin pigmentation patterns have primarily been described in independently evolved lineages: egg-spots, i.e., conspicuous pigmentation patterns with a circular boundary in the most species-rich lineage, i.e., haplochromine lineage; and the blotch, with an irregular boundary in ectodine lineage (Fige 1). Egg-spots exhibit large varieties (different numbers, sizes, colours and positions) among different species (Salzburger et al. 2007) and have been associated with sexual selection (female attraction (Wickler 1962; Hert 1989)) and male-male competition (Theis et al. 2012, 2015). Although not studied as thoroughly as egg-spots, the blotch has also been associated with female attraction (Wickler 1962).

Previous studies have begun to dissect the genetic underpinnings and evolutionary origin of egg-spots. For example, *csf1ra*, a gene implicated in yellow to reddish pigment cells (xanthophores), is expressed in egg-spots and has undergone adaptive sequence evolution in the ancestral lineage of haplochromines (Salzburger et al. 2007). Subsequently, a *cis*-regulatory change in the form of a transposable element (TE) insertion upstream of *fhl2b* resulting in a gain of expression in iridophores (another type of pigment cells) was causally associated with the origin of egg-spots (Santos et al. 2014). More recently, Santos et al. (Santos et al. 2016) suggested that the blotch and egg-spots do not share a common genetic basis (Santos et al. 2016) with rather distinct expression patterns. However, this study was based on the gene expression profile of 46 out of 1229 (3.7%) candidate genes; and the investigated genes were only egg-spots-related candidate genes that did not reflect the expression pattern of the blotch. Therefore, it is still too early to draw this conclusion.

Therefore, to investigate the genetic basis of the emergence of these two convergent novelties, and characterize the mechanism underlying their evolutionary differences (large

varieties of egg-spots but no variation for the blotch), we conducted a thorough comparative transcriptomic and genomic analysis with respect to the gene network, proposing the origin and evolution of these two convergent novel phenotypes in an evolutionary scope.

Results

DE genes for the blotch and egg-spots

Illumina-based RNA sequencing of anal fin tissues of *C. macrops* generated between 15 to 23 million raw reads per library. The average read quality was 28. Approximately 40,000 to 160,000 reads were trimmed through adaptor trimming (the exact read numbers were provided in Additional File 1). Between eight to ten million reads per library were finally mapped to the Tilapia transcriptome assembly available from the Broad Institute (ftp://ftp.ensembl.org/pub/release-81/fasta/oreochromis_niloticus/cdna/. version 0.71) as a reference. Illumina reads are available from the Sequence Read Archive (SRA) at NCBI under accession number SRA SRP082469.

To obtain blotch-specific genes, the controlling position and sex-determining effects involving the blotch and non-blotch anal fin tissues of males and females were compared. Briefly, we determined the differentially expressed (DE) genes between the blotch and non-blotch tissues in males. This set of genes was associated with the blotch pattern formation, but could also be associated with distal fin position patterning without affecting blotch formation. To control this positional effect, we compared the corresponding anal fin tissues in females, generating a list of potentially sex-specific DE genes related to distal fin pattern formation. However, further comparisons between females and males excluded both

150 positional and sex-specific effects. Using edgeR (Robinson & Smyth 2007, 2008; Robinson
151 et al. 2010; McCarthy et al. 2012; Zhou et al. 2013), we thus identified 263 significantly
152 over-expressed genes and 11 significantly under-expressed genes in the blotch tissue. Re-
153 analysis of the raw data for egg-spots and non-egg-spots from Santos et al. (Santos et al.
154 2016) revealed 351 significantly over-expressed genes and 461 significantly under-expressed
155 genes for egg-spots tissues.

156

157 **Comparative transcriptomic analysis for the blotch and egg-spots**

158

159 Forty-four genes showed similar gene expression patterns in both the blotch and egg-
160 spots. Particularly for the top ten DE blotch-related candidate genes, six genes showed
161 similar expression patterns in egg-spots, including two duplicated iridophore-related genes
162 (*pnp4a* and *pnp5a*), the immunity-related gene *ly6d*, the ligand transporter-related gene
163 apolipoprotein D (*ApoD*), and two genes previously associated with egg-spots formation
164 (*fhl2a* and *fhl2b*) (Santos et al. 2014) (Table 1). Two genes showed high expression patterns
165 in egg-spots, but low expression patterns in the blotch tissues, including the collagen-related
166 gene *col8a1b* (Shellswell et al. 1980) and the immunity-related gene *steap4* (Benard et al.
167 2014) (Table 2). Fifteen genes were over-expressed in the blotch tissues but were down-
168 regulated in egg-spots (Table 2). Clustering analysis based on the expression levels of 44
169 common genes clustered egg-spots and the blotch together, despite their species boundaries
170 (Figure 1).

Table 1 Top ten common blotch DE genes

	gene	Ensembl	logFC the blotch	logFC egg-spots	Ensembl Description
1	apod	ENSONIT00000023475	-3.58	-0.91	Apolipoprotein Da
2	fhl2b	ENSONIT00000017889	-3.56	-1.71	four and a half LIM domains 2b
3	pnp4a	ENSONIT00000000731	-3.52	-1.37	Purine nucleoside phosphorylase 4a
4	ly6d	ENSONIT00000024767	-3.5	-0.89	lymphocyte antigen 6D-like
5	fhl2a	ENSONIT00000015512	-3.46	-1.11	four and a half LIM domains 2a
6	cecr5	ENSONIT00000000334	-3.32	-1.66	cat eye syndrome critical region protein 5-like
7	ifi30	ENSONIT00000016006	-3.29	-1.37	interferon, gamma-inducible protein 30
8	gpnmb	ENSONIT00000005686	-3.21	-1.1	Glycoprotein (transmembrane) nmb
9	TM4SF4	ENSONIT00000005261	-2.88	-0.66	TM4SF4
10	plin5	ENSONIT00000002684	-2.87	-1.11	perilipin-5-like

Table 2 DE genes with contrasting expression patterns between the blotch and egg-spots

	gene	Ensembl	logFC the blotch	logFC egg-spots	Ensembl Description
1	pnp5a	ENSONIT00000016474	-3.55	0.77	Purine nucleoside phosphorylase 5a
2	BDH1	ENSONIT00000016804	-3.18	1.06	D-beta-hydroxybutyrate dehydrogenase, mitochondrial-like
3	keratin	ENSONIT00000023698	-2.96	1.19	keratin, type I cytoskeletal 20-like
4	egfl6	ENSONIT00000020235	-1.96	0.89	EGF-like-domain, multiple 6
5	KIF5B	ENSONIT00000004756	-1.93	0.82	Kinesin family member 5B
6	hmx4	ENSONIT00000004437	-1.83	1.67	H6 family homeobox 4
7	calb2a	ENSONIT00000014036	-1.8	0.74	Calbindin 2a
8	dhrsx	ENSONIT00000009275	-1.58	0.87	Dehydrogenase/reductase (SDR family) X-linked
9	tgm1l2	ENSONIT00000003616	-1.35	0.6	Transglutaminase 1 like 2
10	KIF5B	ENSONIT00000004755	-1.25	0.79	Kinesin family member 5B
11	cx43	ENSONIT00000026288	-1.24	0.77	Connexin 43
12	KIAA2026	ENSONIT00000017688	-0.94	0.78	KIAA2026
13	oca2	ENSONIT00000006160	-0.94	1.1	Oculocutaneous albinism II
14	fosb	ENSONIT00000004670	-0.84	1.29	FBJ murine osteosarcoma viral oncogene homolog B
15	hcst	ENSONIT00000003510	-0.84	0.56	Hematopoietic cell signal transducer
16	steap4	ENSONIT00000007508	0.61	-0.53	STEAP family member 4
17	col8a1b	ENSONIT00000023612	2.17	-0.94	Collagen, type XII, alpha 1b

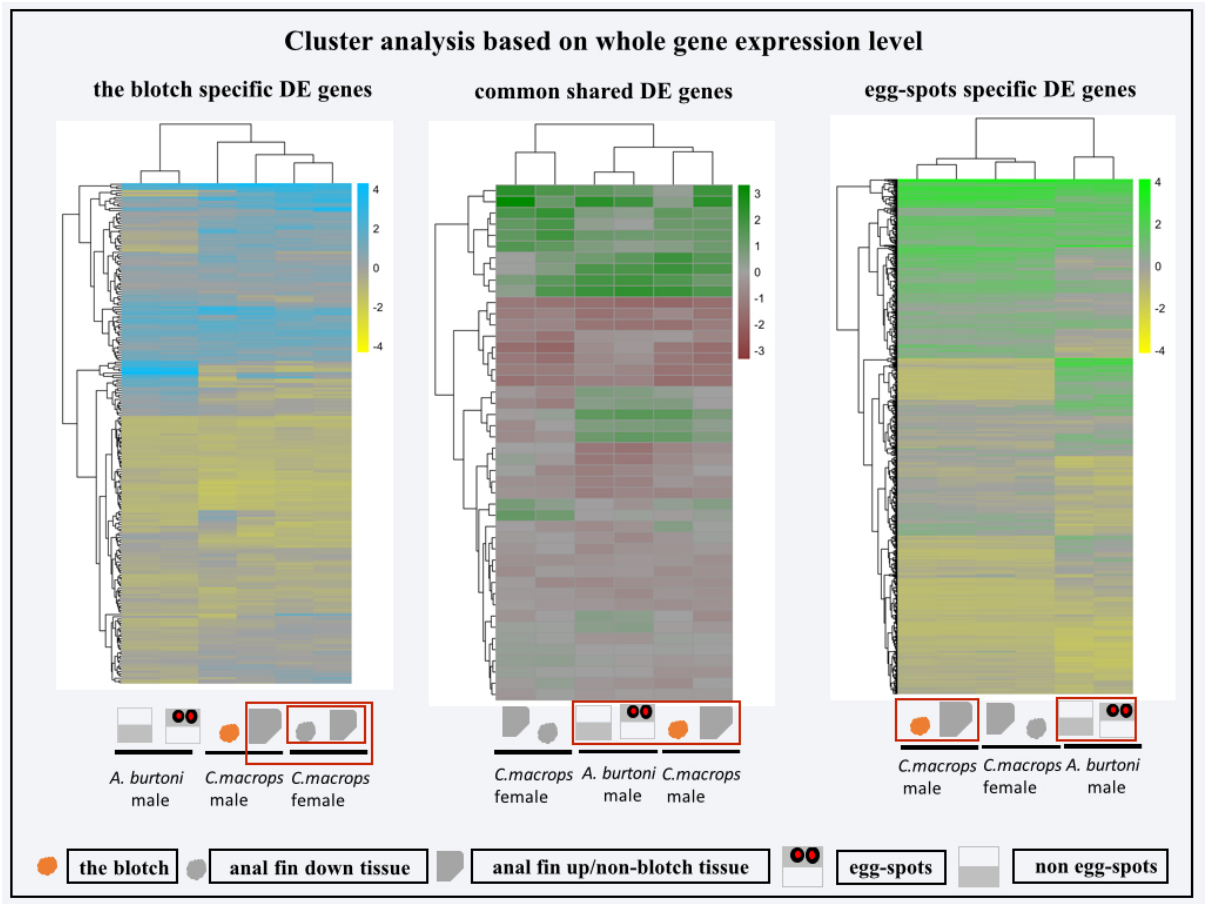

Figure 1 Clustering analysis based on the gene expression level of common differentially expressed (DE) genes and pattern-specific DE genes of anal fin pigmentation patterns. The cluster analysis was conducted with the blotch and non-blotch tissues in male fish, the corresponding tissues in the female cichlid fish *C. macrops*, and egg-spots and non-egg-spots tissues in the male cichlid fish *A. burtoni*.

More than 50% GO terms (Biological Process, Molecular Evolution and Cellular Component) and pathways were shared between the blotch and egg-spots (Figure 2a). GO enrichment analysis (FDR<0.1) showed that the common shared enriched GO terms were primarily related to pigmentation, including pigmentation cell differentiation (GO: 0050931), pigment granule organization (GO: 0048753), melano-related terms (GO:0030318, GO: 0032438, GO: 0004962), iridophore-related GO term (GO: 0050935), and nucleobase-containing small molecule metabolic process (GO: 0055086) (Figure 2b).

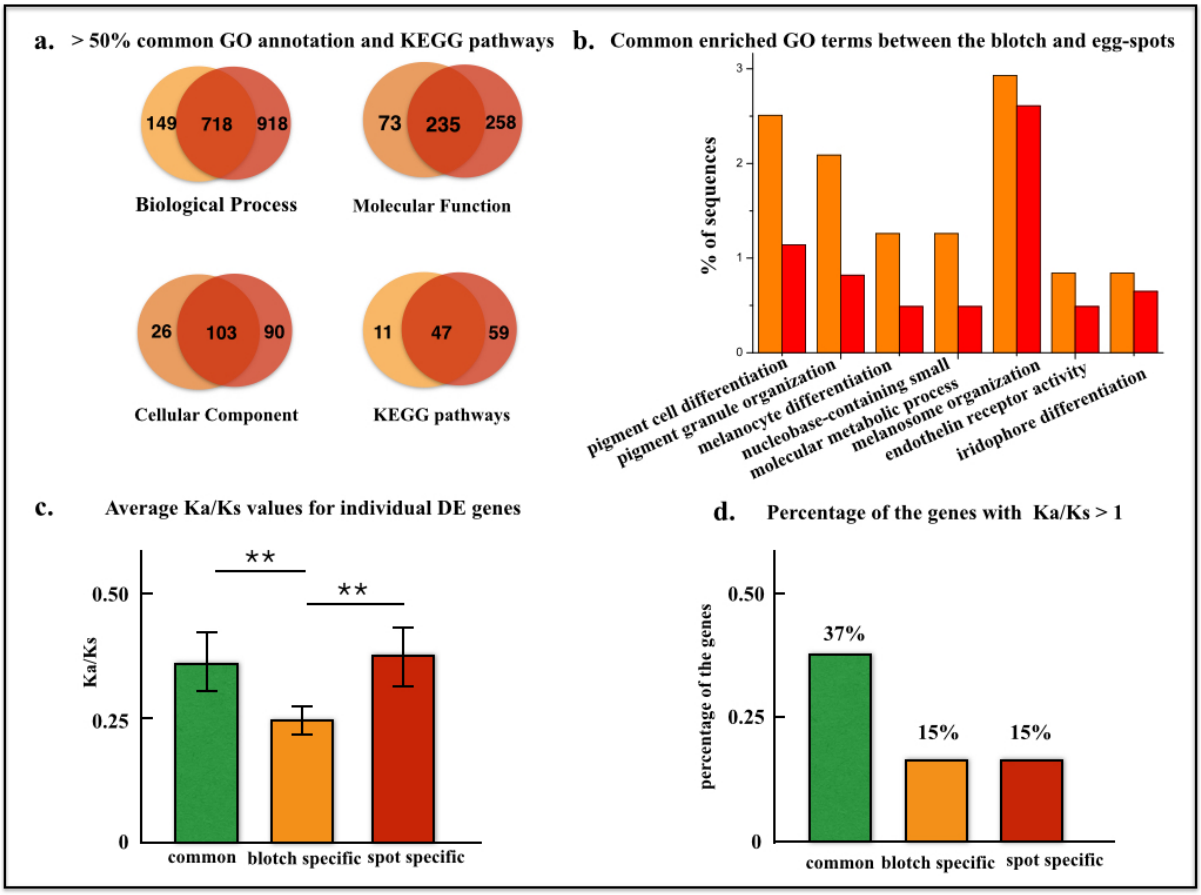

Figure 2 Common genetic basis was shared between the blotch and egg-spots with accelerated evolutionary rates. (a) > 50% common shared Gene annotation (GO) terms and pathways between the blotch and egg-spots. (b) Common enriched GO terms between the blotch and egg-spots (FDR<0.01). (c) Average evolutionary rates (Ka/Ks) of common shared differentially expressed (DE) genes, the blotch-specific DE genes and egg-spots-specific DE genes. ** represents p<0.01. (d) Statistics of the genes with Ka/Ks values larger than one in common genes, the blotch-specific DE genes and egg-spots-specific DE genes, respectively.

Evolutionary rate detection for DE genes and related pathways

To compare the evolutionary rates of common shared and pattern-specific DE genes between the blotch and egg-spots, we calculated the average pairwise Ka/Ks values of the DE genes based on the data from five available cichlid fishes and *C. macrops* in the present study. The average Ka/Ks value of common genes showed a significantly higher rate than the value of the blotch-specific DE genes, but not higher than the egg-spots-specific DE genes

(Figure 2c, Additional file 2). Common genes also showed a higher percentage of DE genes with Ka/Ks values larger than 1 (Figure 2d, Additional file 2), explaining the high average Ka/Ks values of the common genes.

To determine the evolutionary rates of common and pattern-specific DE genes in the same and different pathways, we calculated the average Ka/Ks and dN/dS values using a free-ratio model. Both methods showed consistent results (Figure 3, Additional file 2): 1) in general, in the common shared pathways, egg-spots DE genes showed higher evolutionary rates than the blotch DE genes, consistent with the average Ka/Ks values calculated above (Figure 2c); 2) common DE genes showed higher evolutionary rates than the blotch DE genes in the common shared pathways, and occasionally showed even higher rates than egg-spots DE genes, confirming the previous results (Figure 2c); 3) egg-spots DE genes showed higher evolutionary rates primarily in signalling pathways and metabolism pathways, in which genes related to egg-spots-specific pathways play an important role; and 4) the blotch DE genes primarily showed higher evolutionary rates in other pathways (Figure 3a and 3b), although several individual pathways showed high average Ka/Ks rates in signalling pathways (Figure 3a), but not in the free-ratio model. This finding primarily reflects a novel gene (ENSONIG00000013347) with an average Ka/Ks value larger than 1. This difference could reflect the fact that the dN/dS value was calculated in a lineage-specific manner using the free-ratio model instead of averaging across different species. This result did not affect the consistent general observation that egg-spots DE genes primarily showed higher evolutionary rates than the blotch DE genes, particularly in signalling pathways and metabolic pathways.

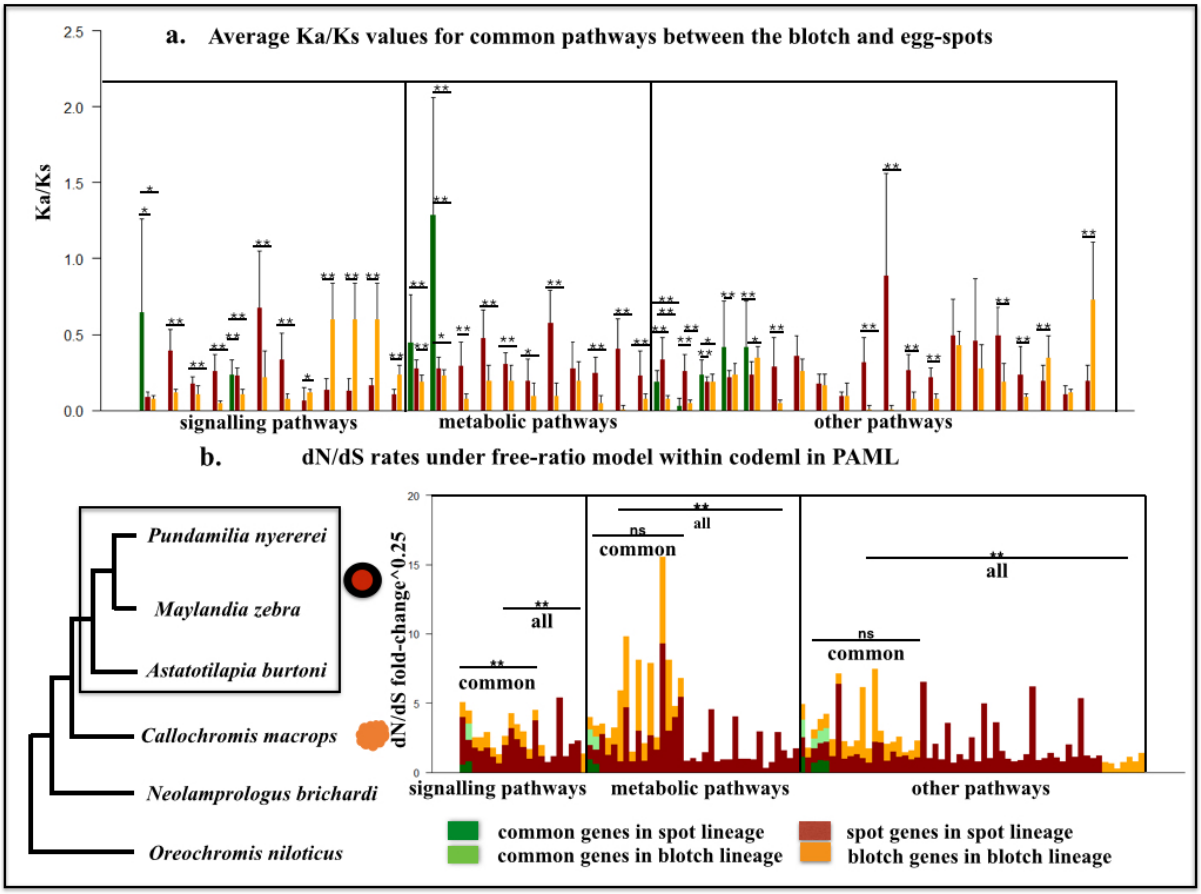

Figure 3 Evolutionary rates calculation for anal fin pigmentation differentially expressed (DE) genes related pathways. (a) Average evolutionary rates (Ka/Ks) values of common shared signalling pathways between the blotch and egg/spots. *represents p<0.05, **represents p<0.01. (b) Rates of evolution (dN/dS) for common shared and pattern-specific pathways using the free-ratio model in a lineage-specific manner within codeml in PAML. Fold-change ratios of the genes in the corresponding lineage were calculated and compared. **represents p<0.01, ns represents no significance.

First common neighbour gene networks (FCN) of the blotch and egg-spots DE genes

To characterize interactions of these DE genes, we examined their protein-protein interaction networks. Network enrichment analysis in both the blotch and egg-spots showed that these genes had more interactions among themselves than what would be expected from the genome (p<0.01), suggesting that these genes are at least partially biologically related as a group (https://string-db.org/), which further confirmed the robustness of the experimental

241 design. To predict the roles of these common genes in the network and determine whether
242 they have similar functions in both the blotch and egg-spots, we compared their first
243 neighbour gene network (FCN) (Figure 4), as direct interactions among genes can indicate
244 their roles to a maximum extent. These FCNs showed significantly higher interaction levels
245 than the total DE genes (Table 3), suggesting that the FCN of common genes belongs to a
246 core gene network.

247

Table 3 Statistics of degree of first common neighbour gene network (FCN)

comparison	average	W	p-value	significance
blotch FCN vs. blotch total	6.08+-4.24 vs. 4.11+-3.85	5135.5	<0.01	**
spot FCN vs. spot total	9.27+-9,69 vs. 5.75+-6.31	34364	<0.01	**
blotch CCN vs. spot CCN	5.81+-4.13 vs. 11.33+-10.66	186.5	0.04	*
blotch SCN vs. spot SCN	6.17+-4.36 vs. 8.47+-9.29	2097.5	0.56	ns

248 method: Mann-Whitney U test

249

250 Two types of FCNs were formed: 1) direct interactions among the common genes,
251 comprising pigmentation (melanophore and xanthophore)-related genes (Figure 4) (core
252 common network, CCN). Notably, unlike in the blotch, the CCN in egg-spots integrated more
253 transcription factors (TFs) (Figure 4), largely contributing to its higher interaction degree
254 (Table 3). 2) The second type of FCN involved interactions among common genes and
255 pattern-specific DE genes (specific common network, SCN). Unlike the CCN, the common
256 genes in SCN were only loosely associated (Figure 4). For example, the SCN of the blotch
257 could clearly be divided into sub-networks, connected only by two common genes (*myo5aa*
258 and *pde7b*) (Figure 4). However, the common genes of the SCN in egg-spots (8.47±9.29)
259 participated in more interactions than in the blotch (6.17±4.36), although no significance was
260 detected, which likely reflects the limited sample size (Table 3). In addition, genes related to
261 purine metabolism pathways were detected in the SCN of the blotch, but more signalling
262 pathways were involved in the SCN of egg-spots (Figure 5).

The evolutionary rates of common genes were significantly higher than those of pattern--specific DE genes in both the CCN and SCN of the blotch, but not those of the egg-spots (Figure 4), and egg-spots DE genes showed significantly higher evolutionary rates than the blotch DE genes in both the CCN and SCN (Figure 4). These findings are also consistent with previous results, showing that the average Ka/Ks value of common DE genes and egg-spots DE genes were significantly higher than those of the blotch-specific DE genes (Figure 2).

In addition, several novel genes were involved in first neighbour gene networks (Figure 4). Since the interactions among genes indicate their similar functions, the network presented here enabled the prediction of new gene functions. For example, in the CCN of the blotch, the novel gene ENSONIG00000017631 interacted with *mtifb*, *mc1r* and *sox10*, suggesting a similar role related to melanin (Hou et al. 2006; San-Jose et al. 2015). The blotch-specific novel gene ENSONIG00000021415 interacted with *mitfb*, *pax7a* and *pax7b* (Minchin & Hughes 2008; Curran et al. 2010), suggesting a xanthophore-related role.

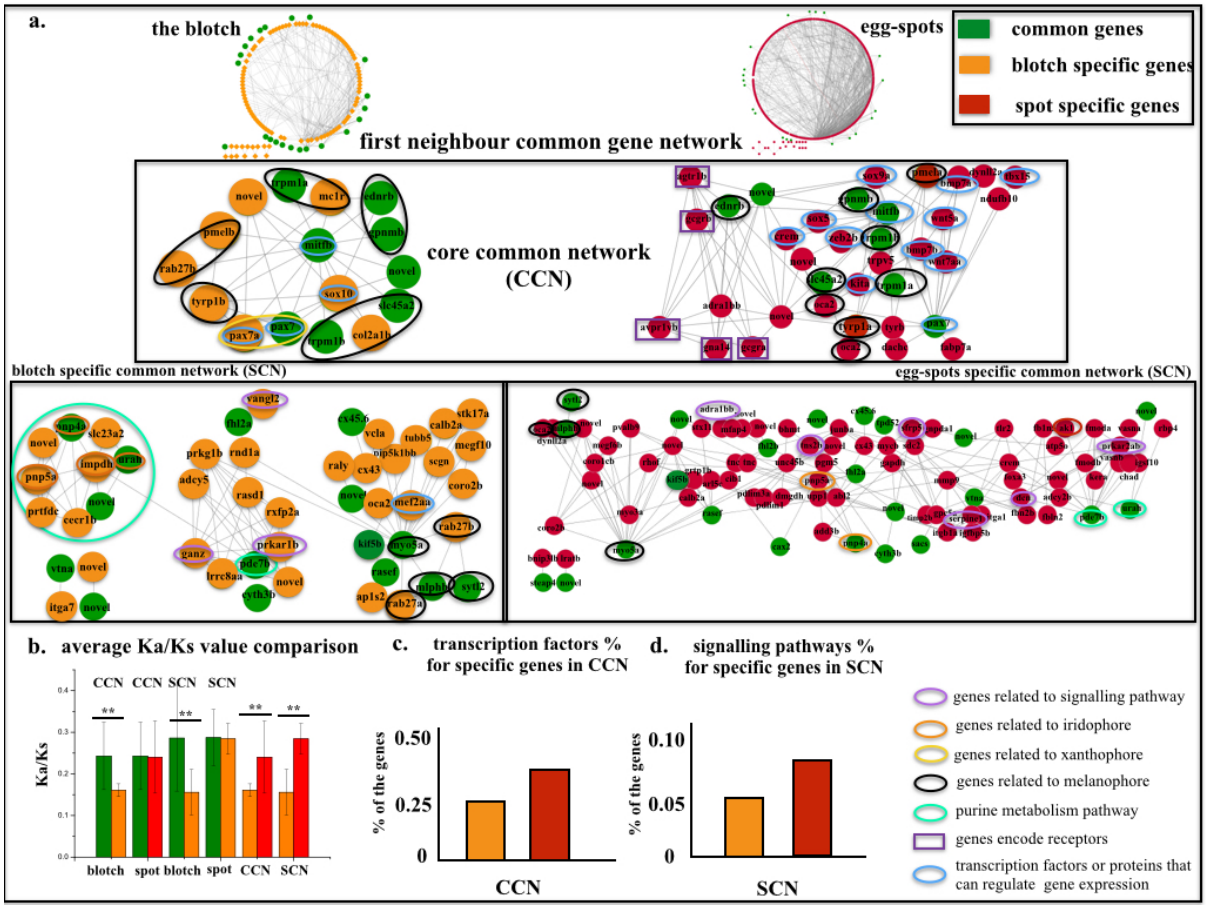

Figure 4 First common neighbour gene interaction network analysis for the blotch and egg-spots DE genes. (a) First common neighbour gene network between the blotch and egg-spots, including core common gene network (CCN), primarily comprising direct interactions among pigment cells (above), and the specific core gene network (SCN), comprising interactions among common genes and pattern-specific DE genes (below). (b) Evolutionary rates (Ka/Ks) of different genes from the CCN and SCN between the blotch and egg-spots. **represents p<0.01. (c) Percentage of transcription factors (TFs) in the corresponding gene network between the blotch and egg-spots. (d) Percentage of signalling pathway genes in the corresponding gene network between the blotch and egg-spots.

Discussion

The thorough comparative transcriptomic and genomic data analysis conducted herein revealed a common genetic basis with higher evolutionary rates between the blotch and egg-spots. Focusing on the whole expression level and expression patterns of both novelties

294 prevents potential biases that lead to different and even opposite conclusions conclusions
295 from Santos et al. (Santos et al. 2016). Here, we proposed that the common genetic basis is
296 essential for the emergence of these novel phenotypes, and independently evolved
297 connections between common genes and TFs and the integration of advantageous genes
298 involved in the signalling and metabolic pathways are crucial for the emergence and
299 evolution of egg-spots.
300
301 **Common genetic basis is important for the evolution of convergent novel anal fin**
302 **pigmentation patterns**
303
304 Similar gene expression patterns in convergent taxa are generally associated with
305 conserved and important roles (such as toolkit genes) in the origin of phenotype (Khaitovich
306 et al. 2006; Pankey et al. 2014). In the present study, clustering analysis based on gene
307 expression level of common genes clustered egg-spots and the blotch together in spite of
308 species boundary, suggesting that convergent evolution of anal fin pigmentation patterns is in
309 parallel with gene expression level, at least on the common genetic basis level. Furthermore,
310 the gene *fhl2b*, which was associated with egg-spots formation (Santos et al. 2014) was also
311 observed ranking at the top of the blotch candidate DE gene list in the present study (Table
312 1). Re-use of the same gene further attached their importance of common gene to the
313 formation of both egg-spots and the blotch. These common genes also showed higher
314 accelerated evolutionary rates than the blotch-specific DE genes, evincing their advantageous
315 roles during the evolution. The average Ka/Ks value of egg-spots DE genes was higher than
316 the previous results from Santos et al. (Santos et al. 2016), which could be explained by the
317 fact that we integrated into the present data analysis the recent duplicated genes, which are
318 too important for the evolution in fish to ignore (Glasauer & Neuhauss 2014). All of these

319 highlighted the importance of the common genes we observed for the emergence of these
320 novel sexually related traits.
321
322 Then what is the common genetic basis and its roles in the convergent evolution of
323 anal fin pigmentation patterns? Both GO enrichment and gene network construction showed
324 that the common genes primarily comprise pigmentation-related genes. Interactions among
325 pigment cells (melanophore, iridophore and xanthophore) are important for pigmentation
326 pattern formation (Parichy 2009; Parichy & Spiewak 2015; Singh & Nüsslein-Volhard 2015;
327 Watanabe & Kondo 2015). For example, the mutation of melanophore-related genes (*mitfb*,
328 *gpnmb*, *trpm1a*, *ednrb*, *trpm1b* and *sox10*), the iridophore-related gene (*pnp4a*), the
329 xanthophore-related gene (*pax7*), and the connexion-related gene *cx45.6* could induce the
330 irregular stripe pattern formation in zebrafish (Shin et al. 1999; Hou et al. 2006; Minchin &
331 Hughes 2008; Curran et al. 2010; Zhang et al. 2012; Reissmann & Ludwig 2013; Irion et al.
332 2014), suggesting important roles for these genes in the corresponding core gene network.
333 The genes *fhl2b* (iridophore-related gene), *pax7* (xanthophore-related gene) and *melanocortin*
334 were also associated with pigmentation pattern formation in cichlid fishes (Salzburger et al.
335 2007; Santos et al. 2014; Dijkstra et al. 2017). These genes were detected among the common
336 candidate gene list and FCN networks in the present study. Here, these genes could have
337 similar roles belonging to the core gene network of anal fin pigmentation pattern formation.
338 This idea was further confirmed by the finding that the gene network comprising these genes
339 showed a significantly higher interaction degree than the average interaction degree of total
340 pattern DE genes (Table 3) and accelerated evolutionary rates (Figure 4).
341
342 In addition, metabolism related pathways and signalling pathways occupied a large
343 portion of the shared pathways between the blotch and egg-spots. Actually, it is unsurprising

that metabolism pathways were involved in these sexually related traits, considering the importance of trade-off of the energy allocation between the survival and reproduction (Bleu et al. 2016). Signalling pathways such as Wnt signalling pathway, ErbB signalling pathway and MAPK signalling pathway related to the pattern formation and morphogenesis (Budi et al. 2008; Martin et al. 2012; Schneider et al. 2012; Plestant & Anton 2013; Martin & Reed 2014) can have similar roles here and are advantageous to the emergence of these novel anal fin pigmentation patterns**,** considering their accelerated evolutionary rates. This could be explained by the reasoning that the trait itself is under selection so that the related genes showed accelerated evolutionary rates, or alternatively that the formation of anal fin patterns integrated genes with accelerated evolutionary rates in the pathways during evolution, which conserved the phenotype during evolution. Since common genes showed significantly higher evolutionary rates than pattern specific genes, we prefer the later possibility. In spite of this, they both reconfirmed the conclusion that common genetic basis is important for the emergence and evolution of these convergent anal fin pigmentation patterns.

Independently evolved connections among common genes and pattern specific genes are crucial for the evolution and evolution of egg-spots

Although both the blotch and egg-spots are novelties, their patterns are quite different (Figure 1). The blotch is a reddish pattern with an irregular boundary, whereas egg-spots are pigmentation patterns with circular transparent boundaries. In addition, these two novelties exhibit different evolutions. The blotch only appears in the ectodine lineage, with almost no variables among species. However, egg-spots exhibit significant diversity, including different numbers, sizes, positions and colours, in the species-rich haplochromine lineage (Salzburger et al. 2005; Santos & Salzburger 2012). The emergence of egg-spots is associated with the

369 radiation of cichlid fishes (Salzburger 2009). Therefore, it is important to understand the
370 mechanism underlying their different evolution pattern.
371
372 Genes are not isolated from each other, but rather remain connected. Although
373 common genes were shared between the blotch and egg-spots, more direct interactions were
374 observed between common genes and egg-spots-specific DE genes than the blotch-specific
375 DE genes (Table 3, Figure 5). This finding largely reflects additional connections with TFs
376 and signalling pathways in the egg-spots gene network (Figure 5). New connections could be
377 established by the insertion of transposable elements or a single nucleotide polymorphism
378 mutation in the upstream region of target genes (e.g., *fhl2b* and *pax7a*), which can change
379 binding opportunities to TFs to affect pigmentation pattern formation (Santos et al. 2014;
380 Roberts et al. 2017). Gene network integrating pattern-related TFs and signalling pathways
381 can become relatively independent to integrate signals into a gene expression pattern unique
382 to that character, representing novelty (Wagner 2014), similar to the case of butterfly wing
383 spot (Shirai et al. 2012). However, if the gene network of the novelty is still tightly linked to
384 the ancestral gene network, such as the "Christmas tree model" for the wing spot
385 development in fly, then the evolution of the phenotype will be limited (Wagner & Lynch
386 2008). Therefore, the extent of independence of a gene network can be causally linked to the
387 evolution of a trait (Wagner 2014), likely explaining the differences in evolution between
388 egg-spots and the blotch: the relatively independent network comprising TFs and signalling
389 pathways can free the evolution of egg-spots from the ancestral anal fin network to relatively
390 freely evolve to a higher variety compared to the blotch (intrinsic factor, developmental
391 constraint).
392

In addition, egg-spots showed significantly higher evolutionary rates in both the CCN and SCN compared to the blotch, which primarily reflected the associated TFs and signalling pathways, which occupied a large portion of the network. Higher evolutionary rates of signalling pathways and metabolic pathways of egg-spots DE genes further suggested that evolutionary advantageous genes recruited in the egg-spots gene network could help conserve the phenotype and subsequently evolve during evolution (extrinsic factor, selection) and causally link egg-spots to the adaptive radiation of cichlid fishes.

Hypothesis about the genetic basis of the origin and evolution of convergent novel anal fin pigmentation patterns in cichlid fishes

The findings of the present study drive a reassessment with respect to whether convergent pigmentation pattern formation in fish is as the same as that in mammals and birds, whose pigmentation pattern formation is controlled through a few genes (*MC1R* and *Agouti*) (Hoekstra et al. 2006; Uy et al. 2016), or should be viewed with respect to a gene network. For example, iridophores were observed as the prerequisites for stripe formation in zebrafish (*Danio rerio*) (Singh & Nüsslein-Volhard 2015), but not in its closely related species *D. albolineatus*, in which xanthophores play a crucial role (Patterson et al. 2014). Similar cases were also observed for skin (melanocortin-related genes and the xanthophore-related gene *pax7a*) (Dijkstra et al. 2017; Roberts et al. 2017) and anal fin (the xanthophore-related gene *csf1ra* and the iridophore-related gene *fhl2b*) (Salzburger et al. 2007; Santos et al. 2014) pigmentation pattern formation. This difference might reflect interactions among rather diversified pigment cells in fish, resulting in more complex pigmentation patterns than in mammals and birds.

Therefore, we propose one hypothesis with respect to a gene network that can explain the origin and evolution of convergent novel anal fin pigmentation patterns in cichlid fishes (Figure 5). On the one hand, common genes, comprising pigment cells, signalling pathways and metabolic pathways, are important for the emergence of these novel traits (intrinsic factor, developmental constraint). These genes could be derived from deep homology (Shubin et al. 2009; McCune & Schimenti 2012), i.e., pigment cells already present in the anal fin before the origination of these novel convergent pigmentation patterns (new structures can evolve by deploying pre-existing regulatory circuits (Shubin et al. 2009), or evolve independently. Accelerated evolutionary rates of common genes and pathways help conserve these novelties during evolution (extrinsic factor, selection). On the other hand, independently evolved connections with TFs and signalling pathways during evolution are important for the origin and evolution of egg-spots. Theses connections form a relatively independent gene network that frees egg-spots to evolve into diversified phenotypes (different numbers, positions, sizes and colours) (intrinsic factor, developmental constraint). Simultaneously, the integration of advantageous genes with accelerated evolutionary rates into the corresponding gene network can causally link egg-spots to the adaptive radiation in cichlid fishes as a key innovation (Salzburger et al. 2005, 2007; Salzburger 2009). In this case, egg-spots should be considered as “an innovation out of innovation”. The hypothesis here will further illuminate the mechanism of origin and evolution of novelties in a broad sense. Further thorough comparative transcriptomic and genomic analysis across the phylogeny, including more diversified anal fin pigmentation patterns,, will provide additional evidence for this evidence for this hypothesis.

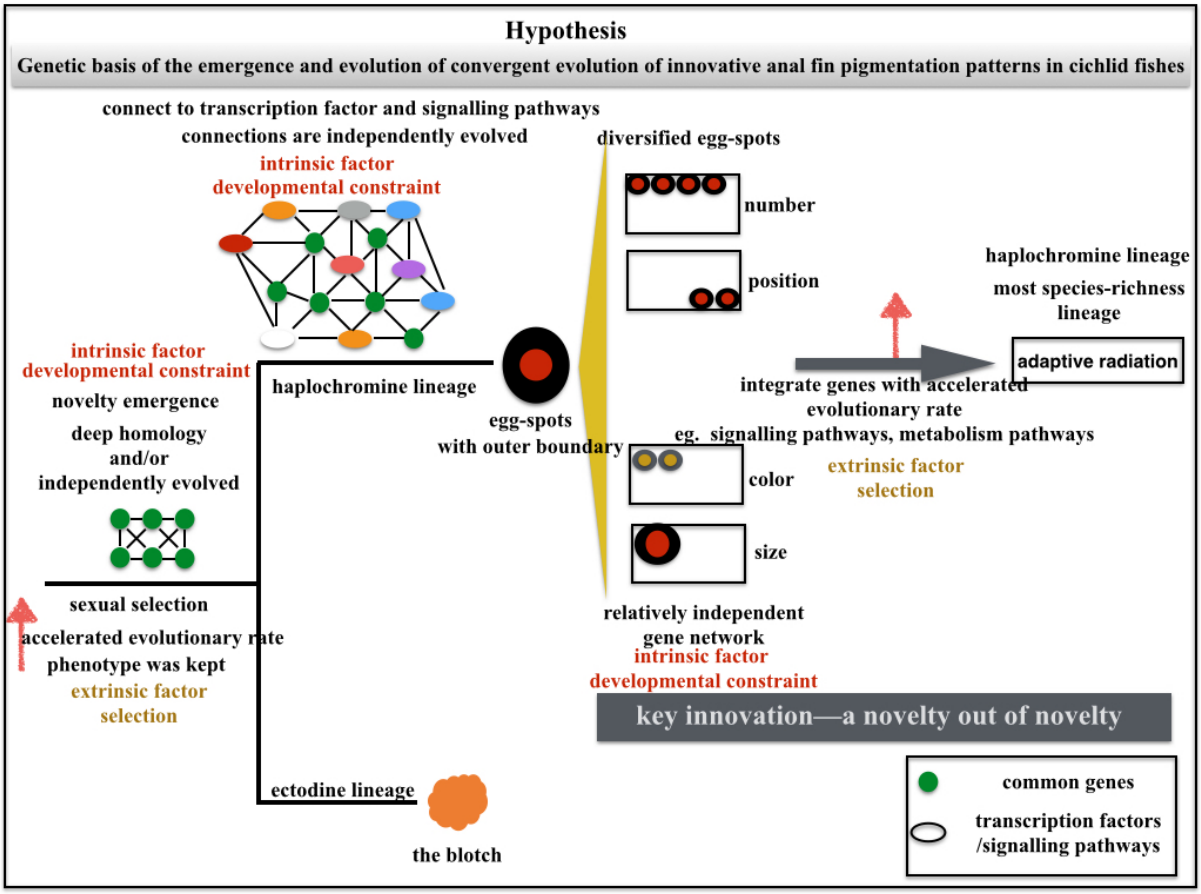

Figure 5 Hypothesis concerning the genetic basis for convergent evolution of the blotch and egg-spots. Common genes derived from deep homology or independently evolved are important for the emergence of convergent novel anal fin pigmentation patterns (intrinsic factor, developmental constraint). The advantageous roles of these genes during evolution helped conserve the novel traits during evolution (extrinsic factor, selection). Independently evolved connections to transcription factors (TFs) and signalling pathways formed a relatively independent gene network that could free the evolution of egg-spots from ancestral anal fin to evolve into diversified phenotypes (different numbers, sizes, positions and colours) (intrinsic factor, developmental constraint). Simultaneously, the integration of advantageous genes related to signalling pathways and metabolic pathways with accelerated evolutionary rates in the corresponding gene network linked egg-spots to the adaptive radiation in cichlid fishes as the key innovation (extrinsic factor, selection). In this case, egg-spots should be considered “an innovation out of innovation”.

Materials and Methods

Illumina-based RNAseq

Laboratory strains of *C. macrops* were maintained at the University of Basel (Switzerland) under standard conditions (12-h light/12-h dark; 26°C, pH=7). Prior to tissue dissection, the specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following approved procedures (permit nr. 2317 issued by the Cantonal Veterinary Office Veterinary Office). For RNAseq of *C. macrops*, we dissected the anal fins from three adult males and three adult females. In the male fins, we separated the blotch area from the remaining fin tissue; in the females, which do not possess the blotch, we separated the fin in the same way, i.e., the areas corresponding to the blotch and non-blotch tissue in males, resulting in a total of 12 samples for *C. macrops*. RNA extraction was performed using was performed using TRIzol® reagent (Invitrogen, USA). Sample clean up and DNase treatments treatments were performed using the using the RNA clean&Concentrator™-5 (Zymo Research Corporation, USA). RNA quality and quantity was determined using theusing the Nanodrop 1000 spectrophotometer (Thermo Scientific, USA) and a Bioanalyser 2100 (Agilent Technologies, Germany). Libraries were generated using the Illumina TruSeq RNA Sample Preparation Preparation Kit (low-throughput protocol) according to the manufacturer's instructions. Per tissue, 330 ng of RNA was subjected to mRNA selection. Pooling and sequencing were performed at the Department of Biosystems Science and Engineering (D-BSSE), University of Basel and ETH-Zurich. Single-end sequencing of these pooled 12 samples was performed in one lane of an Illumina Genome Analyser IIx (maximum read length was 50 bp). The existing transcriptomic data for anal fin with egg-spots and without egg-spots of *A. burtoni* was retrieved from Santos et al. (Santos et al. 2016).

Differential gene expression analysis for the blotch and egg-spots

481 Quality assessment of the sequence reads for the blotch was conducted using Fastqc
482 0.10.1 (www.bioinformatics.babraham.ac.uk/projects/). Contaminated Illumina adapter were
483 removed using cutadapt 1.3 (Martin 2011). The reads were subsequently aligned to the
484 Tilapia transcriptome assembly available from Broad Institute
485 (ftp://ftp.ensembl.org/pub/release-81/fasta/oreochromis_niloticus/cdna/. version 0.71).
486 NOVOINDEX (www.novocraft.com/) was used for indexing the reference and
487 NOVOALIGN (www.novocraft.com/) was used for mapping the reads against the reference.
488 The output SAM files of read mapping were subsequently transformed into the BAM format
489 using SAMtools (Li et al. 2009). The count files were concatenated into count tables and
490 analysed using the Bionconductor R package (Gentleman et al. 2004; Huber et al. 2015) and
491 edgeR (Robinson & Smyth 2007, 2008; Robinson et al. 2010; McCarthy et al. 2012; Zhou et
492 al. 2013). To compare blotch and non-blotch tissues, we used the individual as the block
493 factor, since these tissues were paired. In analyses with edgeR, genes that achieved at least
494 one count per million (cpm) in at least three samples were maintained. To reduce false
495 positive numbers, differentially expressed transcripts were maintained if the false discovery
496 rate (FDR) was smaller than 0.01. The existing transcriptomic data of anal fin with egg-spots
497 and without egg-spots from three males of *A. burtoni* in Santos et al. (Santos et al. 2016)
498 were re-analysed using the same parameters.
499
500 **Comparative transcriptomics between the blotch and egg-spots**
501
502 Gene expression clustering was visualized as a heatmap using the *pheatmap* R
503 package v1.0.8 (https://cran.r-project.org/web/packages/pheatmap/index.html). Gene
504 ontology (GO) annotation of the differential expressed transcripts was conducted with
505 Blast2GO version 2.5.0 (Conesa et al. 2005). BLASTx searches were achieved using

506 BLASTx (threshold: e-6) and a number of hits of 10. KEGG (Kyoto Encyclopaedia of Genes
507 and Genomes) pathway annotation was conducted with KAAS (KEGG Automatic
508 Annotation Server) http://www.genome.jp/tools/kaas/ (Moriya et al. 2007) using zebrafish as
509 the reference with a threshold e-value of e^{-10}. Enrichment analysis of the GO terms was
510 performed using DAVID (https://david.ncifcrf.gov/). Protein-protein interaction network
511 analysis was conducted with the string database (http://string-db.org/). The network was
512 visualized using Cytoscape v3.0 (Shannon et al. 2003).
513

514 **Evolutionary rates detection of DE genes and related pathways**

515

516 To detect the evolutionary rate of the coding sequences of DE genes, we extracted the
517 protein-coding sequences for these DE genes by extracting the corresponding transcripts of
518 tilapia retrieved from the Ensembl database (http://www.ensembl.org/index.html). These
519 transcripts were used as references after removing codon sequences using Biopython
520 (http://biopython.org/wiki/Biopython). The raw reads of transcriptomic data from the other
521 four available cichlid species (*Pundamila nyererei, Neolamprologus brichardi, Maylandia*
522 *zebra* and *A. burtoni*) were retrieved from the NCBI database
523 (https://www.ncbi.nlm.nih.gov/). These data, together with transcriptome data of egg-spots
524 and the blotch, were mapped to individual transcripts (without stop codon) trimmed to the
525 reference using Geneious http://www.geneious.com (Kearse et al. 2012). To retrieve recently
526 duplicated genes, we differentiated the variable single nucleotide polymorphisms deriving
527 from duplicates by mapping these anomalies against the reference. We concatenated all the
528 transcripts for each species using Geneious, followed by visual assessment. Subsequently, we
529 constructed two rounds of consensus sequences through pairwise alignment between the
530 target species and tilapia using MAFFT, which removes the variable single nucleotide

polymorphism according to the reference. The resulting concatenated sequences were translated into amino acids and re-mapped using existing transcriptome data, followed by visual assessment. To obtain individual transcripts, we annotated each transcript in concatenated sequences and subsequently extracted the transcripts using Geneious.

To detect the evolutionary rate of these DE genes, we aligned the sequences using T_Coffee (Notredame et al. 2000) based on the codon sequence. Subsequently, pairwise Ka/Ks values were calculated using yn00 in PAML (Yang 1997, 2007). The pipeline was written in BioPerl http://bioperl.org/. To determine whether the same pattern also occurred in a lineage-specific manner, we detected the dN/dS values for individual DE genes using the free-ratio model within codeml in PAML.

Additional File 1: Illumina sequencing reads for the ectodine blotch in *C. macrops.*

Additional File 2: Average evolutionary rates (Ka/Ks) and comparison of the DE genes and related pathways between the blotch and egg-spots.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LG designed the experiment and wrote the manuscript with the help from CZ. LG performed the RNAseq library construction and data analysis with the help from CZ and CX. All authors read and approved the manuscript.

Acknowledgements

We particularly thank Prof. Walter Salzburger for the guidance and support for the project. We also particularly thank Dr. Emilia M. Santos for providing the raw transcriptomic data of egg-spots. Many thanks to Shengkai Pan, Yongjian Zhang, Lei Luo, Yongsheng Li, Jie Zhang, Tingting Zhou, De Chen for the discussion and suggestion. We thank Philippe Demougin and Ina Nissen for the assistance with Illumina sequencing. We also thank Attila Rüegg for the assistance in the aquarium room, Nicolas Boileau for the assistance in the lab and Brigitte Aeschbach for the laboratory apparatus organization. Illumina reads from this study are available at NCBI under the accession number SRP082469. Alignment sequences are available upon request from the corresponding author. This work was supported by the PhD grant from University of Basel, Switzerland and Postdoc funding from Southwest University, China.

References

Annona G et al. 2015. Evolution of the notochord. Evodevo. 6:30. doi: 10.1186/s13227-015-0025-3.

Arendt J et al. 2008. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol. Evol. 23:26–32. doi: 10.1016/j.tree.2007.09.011.

Bang R, Schultz TR, DeSalle R. 2002. Development, homology and systematics. EXS. 175–86. http://www.ncbi.nlm.nih.gov/pubmed/11924496 (Accessed October 15, 2016).

Benard EL, Roobol SJ, Spaink HP, Meijer AH. 2014. Phagocytosis of mycobacteria by zebrafish macrophages is dependent on the scavenger receptor Marco, a key control factor of pro-inflammatory signalling. Dev. Comp. Immunol. 47:223–33. doi: 10.1016/j.dci.2014.07.022.

581 Berens AJ, Hunt JH, Toth AL. 2015. Comparative Transcriptomics of Convergent Evolution:
582 Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of
583 Eusocial Insects. Mol. Biol. Evol. 32:690–703. doi: 10.1093/molbev/msu330.
584 Bhullar B-AS et al. 2015. A molecular mechanism for the origin of a key evolutionary
585 innovation, the bird beak and palate, revealed by an integrative approach to major transitions
586 in vertebrate history. Evolution. 69:1665–77. doi: 10.1111/evo.12684.
587 Bleu J, Gamelon M, Sæther B-E. 2016. Reproductive costs in terrestrial male vertebrates:
588 insights from bird studies. Proc. R. Soc. B Biol. Sci. 283:20152600. doi:
589 10.1098/rspb.2015.2600.
590 Brawand D et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish.
591 Nature. 513:375–81. doi: 10.1038/nature13726.
592 Bright JA, Marugán-Lobón J, Cobb SN, Rayfield EJ. 2016. The shapes of bird beaks are
593 highly controlled by nondietary factors. Proc. Natl. Acad. Sci. U. S. A. 113:5352–5357. doi:
594 10.1073/pnas.1602683113.
595 Budi EH, Patterson LB, Parichy DM. 2008. Embryonic requirements for ErbB signaling in
596 neural crest development and adult pigment pattern formation. Development. 135:2603–14.
597 doi: 10.1242/dev.019299.
598 Chen L, DeVries AL, Cheng CH. 1997. Convergent evolution of antifreeze glycoproteins in
599 Antarctic notothenioid fish and Arctic cod. Proc. Natl. Acad. Sci. U. S. A. 94:3817–22.
600 http://www.ncbi.nlm.nih.gov/pubmed/9108061 (Accessed October 15, 2016).
601 Clark-Hachtel CM, Tomoyasu Y. 2016. Exploring the origin of insect wings from an evo-
602 devo perspective. Curr. Opin. Insect Sci. 13:77–85. doi: 10.1016/j.cois.2015.12.005.
603 Colombo M et al. 2013. The ecological and genetic basis of convergent thick-lipped
604 phenotypes in cichlid fishes. Mol. Ecol. 22:670–84. doi: 10.1111/mec.12029.
605 Conesa A et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in

606 functional genomics research. Bioinformatics. 21:3674–6. doi:
607 10.1093/bioinformatics/bti610.
608 Cracraft J. 2005. Phylogeny and evo-devo: Characters, homology, and the historical analysis
609 of the evolution of development. Zoology. 108:345–356. doi: 10.1016/j.zool.2005.09.003.
610 Curran K et al. 2010. Interplay between Foxd3 and Mitf regulates cell fate plasticity in the
611 zebrafish neural crest. Dev. Biol. 344:107–18. doi: 10.1016/j.ydbio.2010.04.023.
612 Dijkstra PD et al. 2017. The melanocortin system regulates body pigmentation and social
613 behaviour in a colour polymorphic cichlid fish. Proc. R. Soc. B Biol. Sci. 284:20162838. doi:
614 10.1098/rspb.2016.2838.
615 Faunes M, Francisco Botelho J, Ahumada Galleguillos P, Mpodozis J. 2015. On the
616 hodological criterion for homology. Front. Neurosci. 9:223. doi: 10.3389/fnins.2015.00223.
617 Gentleman RC et al. 2004. Bioconductor: open software development for computational
618 biology and bioinformatics. Genome Biol. 5:R80. doi: 10.1186/gb-2004-5-10-r80.
619 Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its
620 evolutionary consequences. Mol. Genet. Genomics. 289:1045–60. doi: 10.1007/s00438-014-
621 0889-2.
622 Hall BK. 2013. Homology, homoplasy, novelty, and behavior. Dev. Psychobiol. 55:4–12.
623 doi: 10.1002/dev.21039.
624 Hert E. 1989. The function of egg-spots in an African mouth-brooding cichlid fish. Anim.
625 Behav. 37:726–732. doi: 10.1016/0003-3472(89)90058-4.
626 Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino
627 acid mutation contributes to adaptive beach mouse color pattern. Science. 313:101–4. doi:
628 10.1126/science.1126121.
629 Hou L, Arnheiter H, Pavan WJ. 2006. Interspecies difference in the regulation of melanocyte
630 development by SOX10 and MITF. Proc. Natl. Acad. Sci. U. S. A. 103:9081–5. doi:

631 10.1073/pnas.0603114103.
632 Huber W et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor.
633 Nat. Methods. 12:115–21. doi: 10.1038/nmeth.3252.
634 Irion U et al. 2014. Gap junctions composed of connexins 41.8 and 39.4 are essential for
635 colour pattern formation in zebrafish. Elife. 3:e05125. doi: 10.7554/eLife.05125.
636 Joron M, Jiggins CD, Papanicolaou A, McMillan WO. 2006. Heliconius wing patterns: an
637 evo-devo model for understanding phenotypic diversity. Heredity (Edinb). 97:157–67. doi:
638 10.1038/sj.hdy.6800873.
639 Kearse M et al. 2012. Geneious Basic: an integrated and extendable desktop software
640 platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–9. doi:
641 10.1093/bioinformatics/bts199.
642 Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression.
643 Nat. Rev. Genet. 7:693–702. doi: 10.1038/nrg1940.
644 Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat.
645 Rev. Genet. 5:288–98. doi: 10.1038/nrg1316.
646 Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics.
647 25:2078–9. doi: 10.1093/bioinformatics/btp352.
648 Martin A, Papa R, Nadeau N. 2012. Diversification of complex butterfly wing patterns by
649 repeated regulatory evolution of a Wnt ligand. Proc. ….
650 https://www.pnas.org/content/109/31/12632.full (Accessed April 12, 2016).
651 Martin A, Reed RD. 2014. Wnt signaling underlies evolution and development of the
652 butterfly wing pattern symmetry systems. Dev. Biol. 395:367–78. doi:
653 10.1016/j.ydbio.2014.08.031.
654 Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing
655 reads. EMBnet.journal. 17:10. doi: 10.14806/ej.17.1.200.

656 McCarthy DJ, Chen Y, Smyth GK. 2012. Differential expression analysis of multifactor
657 RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 40:4288–97.
658 doi: 10.1093/nar/gks042.
659 McCune AR, Schimenti JC. 2012. Using genetic networks and homology to understand the
660 evolution of phenotypic traits. Curr. Genomics. 13:74–84. doi:
661 10.2174/138920212799034785.
662 Minchin JEN, Hughes SM. 2008. Sequential actions of Pax3 and Pax7 drive xanthophore
663 development in zebrafish neural crest. Dev. Biol. 317:508–22. doi:
664 10.1016/j.ydbio.2008.02.058.
665 Monteiro A. 2015. Origin, development, and evolution of butterfly eyespots. Annu. Rev.
666 Entomol. 60:253–71. doi: 10.1146/annurev-ento-010814-020942.
667 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic
668 genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–5. doi:
669 10.1093/nar/gkm321.
670 Müller GB, Wagner GP. 1991. Novelty in Evolution: Restructuring the Concept. Annu. Rev.
671 Ecol. Syst. 22:229–256. doi: 10.1146/annurev.es.22.110191.001305.
672 Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive
673 radiation of cichlid fishes. Curr. Biol. 22:2362–8. doi: 10.1016/j.cub.2012.10.048.
674 Nagy LG et al. 2014. Latent homology and convergent regulatory evolution underlies the
675 repeated emergence of yeasts. Nat. Commun. 5:4471. doi: 10.1038/ncomms5471.
676 Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate
677 multiple sequence alignment. J. Mol. Biol. 302:205–217. doi: 10.1006/jmbi.2000.4042.
678 Panchen AL. 2007. Homology - History of a Concept. In: John Wiley & Sons, Ltd. pp. 5–23.
679 doi: 10.1002/9780470515655.ch2.
680 Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014. Predictable

681 transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proc.
682 Natl. Acad. Sci. U. S. A. 111:E4736–42. doi: 10.1073/pnas.1416574111.
683 Parichy DM. 2009. Animal pigment pattern: an integrative model system for studying the
684 development, evolution, and regeneration of form. Semin. Cell Dev. Biol. 20:63–4. doi:
685 10.1016/j.semcdb.2008.12.010.
686 Parichy DM, Spiewak JE. 2015. Origins of adult pigmentation: diversity in pigment stem cell
687 lineages and implications for pattern evolution. Pigment Cell Melanoma Res. 28:31–50. doi:
688 10.1111/pcmr.12332.
689 Patterson LB, Bain EJ, Parichy DM. 2014. Pigment cell interactions and differential
690 xanthophore recruitment underlying zebrafish stripe reiteration and Danio pattern evolution.
691 Nat. Commun. 5:5299. doi: 10.1038/ncomms6299.
692 Pigliucci M, Müller GB. 2010. Evolution: The Extended Synthesis. MIT Press.
693 Plestant C, Anton ES. 2013. Scaling the MAPK Signaling Threshold during CNS Patterning.
694 Dev. Cell. 25:221–222. doi: 10.1016/j.devcel.2013.04.014.
695 Reissmann M, Ludwig A. 2013. Pleiotropic effects of coat colour-associated mutations in
696 humans, mice and other mammals. Semin. Cell Dev. Biol. 24:576–586. doi:
697 10.1016/j.semcdb.2013.03.014.
698 Roberts RB, Moore EC, Kocher TD. 2017. An allelic series at *pax7a* is associated with
699 colour polymorphism diversity in Lake Malawi cichlid fish. Mol. Ecol. 26:2625–2639. doi:
700 10.1111/mec.13975.
701 Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for
702 differential expression analysis of digital gene expression data. Bioinformatics. 26:139–40.
703 doi: 10.1093/bioinformatics/btp616.
704 Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag
705 abundance. Bioinformatics. 23:2881–7. doi: 10.1093/bioinformatics/btm453.

706 Robinson MD, Smyth GK. 2008. Small-sample estimation of negative binomial dispersion,
707 with applications to SAGE data. Biostatistics. 9:321–32. doi: 10.1093/biostatistics/kxm030.
708 Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive
709 radiations of cichlid fishes. Mol. Ecol. 18:169–85. doi: 10.1111/j.1365-294X.2008.03981.x.
710 Salzburger W, Braasch I, Meyer A. 2007. Adaptive sequence evolution in a color gene
711 involved in the formation of the characteristic egg-dummies of male haplochromine cichlid
712 fishes. BMC Biol. 5:51. doi: 10.1186/1741-7007-5-51.
713 Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: genesis, explosive
714 speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC
715 Evol. Biol. 5:17. doi: 10.1186/1471-2148-5-17.
716 San-Jose LM et al. 2015. Effect of the MC1R gene on sexual dimorphism in melanin-based
717 colorations. Mol. Ecol. 24:2794–808. doi: 10.1111/mec.13193.
718 Santos ME et al. 2016. Comparative transcriptomics of anal fin pigmentation patterns in
719 cichlid fishes. BMC Genomics. 17:712. doi: 10.1186/s12864-016-3046-y.
720 Santos ME et al. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory
721 change. Nat. Commun. 5:5149. doi: 10.1038/ncomms6149.
722 Santos ME, Salzburger W. 2012. Evolution. How cichlids diversify. Science. 338:619–21.
723 doi: 10.1126/science.1224818.
724 Schneider PN, Slusarski DC, Houston DW. 2012. Differential role of Axin RGS domain
725 function in Wnt signaling during anteroposterior patterning and maternal axis formation.
726 PLoS One. 7:e44096. doi: 10.1371/journal.pone.0044096.
727 Shannon P et al. 2003. Cytoscape: a software environment for integrated models of
728 biomolecular interaction networks. Genome Res. 13:2498–504. doi: 10.1101/gr.1239303.
729 Shellswell GB, Bailey AJ, Duance VC, Restall DJ. 1980. Has collagen a role in muscle
730 pattern formation in the developing chick wing? 1. An immunofluorescence study. J.

731 Embryol. Exp. Morphol. 60:245–54. http://www.ncbi.nlm.nih.gov/pubmed/7031161
732 (Accessed June 2, 2016).
733 Shen Y-Y, Liang L, Li G-S, Murphy RW, Zhang Y-P. 2012. Parallel evolution of auditory
734 genes for echolocation in bats and toothed whales. PLoS Genet. 8:e1002788. doi:
735 10.1371/journal.pgen.1002788.
736 Shimeld SM, Holland PW. 2000. Vertebrate innovations. Proc. Natl. Acad. Sci. U. S. A.
737 97:4449–52. http://www.ncbi.nlm.nih.gov/pubmed/10781042 (Accessed November 7, 2016).
738 Shin MK, Levorse JM, Ingram RS, Tilghman SM. 1999. The temporal requirement for
739 endothelin receptor-B signalling during neural crest development. Nature. 402:496–501. doi:
740 10.1038/990040.
741 Shirai LT et al. 2012. Evolutionary history of the recruitment of conserved developmental
742 genes in association to the formation and diversification of a novel trait. BMC Evol. Biol.
743 12:21. doi: 10.1186/1471-2148-12-21.
744 Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty.
745 Nature. 457:818–823. doi: 10.1038/nature07891.
746 Singh AP, Nüsslein-Volhard C. 2015. Zebrafish Stripes as a Model for Vertebrate Colour
747 Pattern Formation. Curr. Biol. 25:R81–R92. doi: 10.1016/j.cub.2014.11.013.
748 Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in
749 angiosperms. Curr. Opin. Plant Biol. 30:159–65. doi: 10.1016/j.pbi.2016.03.015.
750 Stern DL. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–64.
751 doi: 10.1038/nrg3483.
752 Theis A, Bosia T, Roth T, Salzburger W, Egger B. 2015. Egg-spot pattern and body size
753 asymmetries influence male aggression in haplochromine cichlid fishes. Behav. Ecol. arv104.
754 doi: 10.1093/beheco/arv104.
755 Theis A, Salzburger W, Egger B. 2012. The function of anal fin egg-spots in the cichlid fish

Astatotilapia burtoni. PLoS One. 7:e29878. doi: 10.1371/journal.pone.0029878.

Uy JAC et al. 2016. Mutations in different pigmentation genes are associated with parallel melanism in island flycatchers. Proc. Biol. Sci. 283. doi: 10.1098/rspb.2016.0731.

Wagner GP. 2014. Homology, genes and evolutionary innovation. Princet. Univ. Press. Princet.

Wagner GP. 2007. The developmental genetics of homology. Nat. Rev. Genet. 8:473–9. doi: 10.1038/nrg2099.

Wagner GP, Lynch VJ. 2008. The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. 23:377–85. doi: 10.1016/j.tree.2008.03.006.

Watanabe M, Kondo S. 2015. Is pigment patterning in fish skin determined by the Turing mechanism? Trends Genet. 31:88–96. doi: 10.1016/j.tig.2014.11.005.

Wickler W. 1962. 'Egg-dummies' as Natural Releasers in Mouth-breeding Cichlids. Nature. 194:1092–1093. doi: 10.1038/1941092a0.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–91. doi: 10.1093/molbev/msm088.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–6. http://www.ncbi.nlm.nih.gov/pubmed/9367129 (Accessed April 16, 2016).

Yong LW, Yu J-K. 2016. Tracing the evolutionary origin of vertebrate skeletal tissues: insights from cephalochordate amphioxus. Curr. Opin. Genet. Dev. 39:55–62. doi: 10.1016/j.gde.2016.05.022.

Zhang P et al. 2012. Silencing of GPNMB by siRNA inhibits the formation of melanosomes in melanocytes in a MITF-independent fashion. PLoS One. 7:e42955. doi: 10.1371/journal.pone.0042955.

Zhou X, Lindsay H, Robinson MD. 2013. Robustly detecting differential expression in RNA

781 sequencing data using observation weights. 18. http://arxiv.org/abs/1312.3382 (Accessed
782 April 12, 2016).