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

Langyu Gu[*1,2], Canwei Xia[3]

[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. langyugu@gmail.com

[3] Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China. xiacanwei@126.com

*Author for Correspondence

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 with different numbers, sizes and positions; and 2) the blotch, irregular pattern with no variation among species. How these two novelties originate and evolve remains unclear. Based on a thorough comparative transcriptomic and genomic analysis, we observed a common genetic basis with high evolutionary rates and similar expression levels between egg-spots and the blotch. Furthermore, associations of common genes with transcription factors and signalling pathways in the core gene network, as well as the integration of advantageous genes were observed for egg-spots. 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 signalling pathways (intrinsic factor) can free the evolution of egg-spots, and together with the integration of advantageous genes (extrinsic factor) can provide a clue to link egg-spots as a key innovation to the adaptive radiation in 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 (Salzburger et al. 2007; M Emília 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 of these fishes and availability of genomics data provide powerful tools to study the relationship between phenotype and genotype (Salzburger 2009; Brawand et al. 2014). 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). Convergent

evolution of anal fin pigmentation patterns primarily include two patterns: egg-spots, i.e. conspicuous pigmentation patterns with a circular boundary, originate once in the most species-rich lineage, i.e. haplochromine lineage (Salzburger et al. 2005a, 2007); and the blotch, with an irregular boundary that only present in four subspecies within *Callochromis* genus in the ectodine lineage (Brichard 1989) (Figure 1). Egg-spots exhibit large varieties (different numbers, sizes and positions, etc.) both within and among different species (Santos & Salzburger 2012; Theis et al. 2017). Both egg-spots and the blotch have been associated with sexual selection (Wickler 1962; Hert 1989; Theis et al. 2012, 2015).

By far, only a few studies begun to dissect the genetic basis of these anal fin pigmentation patterns. The first one came from Salzburger et al. (Salzburger et al. 2007), which suggested that a xanthophore related gene, *csf1ra*, is expressed in egg-spots and has undergone adaptive sequence evolution. Subsequently, Santos et al. (Santos et al. 2014) found a *cis*-regulatory element insertion in the upstream region of *fhl2b* segregating with egg-spots can drive the expression of green fluorescent protein (GFP) in iridophore in zebrafish, and could be causally linked to the origin of egg-spots. More recently, Santos et al. (2016) partially touched on the genetic basis of the blotch by detecting the gene expression profiles of 46 out of 1229 egg-spots related candidate genes and found few common expression patterns between the blotch and egg-spots; thus, concluded that these two traits do not have the same genetic basis. However, this inference was only based on the results from a small portion of candidate genes (3.7%); and the investigated genes were top egg-spots specific candidate genes that are unlikely to reflect the expression pattern of the blotch. Therefore, it is still too early to draw this conclusion.

To investigate the genetic basis of the origin and evolution of these two convergent novelties, we conducted a thorough comparative transcriptomic and genomic analysis with respect to a gene network, and proposed the origin and evolution of these two convergent novel phenotypes in an evolutionary scope. Since these two novelties are both pigmentation patterns with similar colours, albeit with different patterns, we propose that they share a common genetic basis, but independent associations with different genes and pathways can be responsible for their differences.

Results

Differential expressed (DE) transcripts for the blotch and egg-spots

Illumina-based RNA sequencing of anal fin tissues of *Callochromis macrops* generated between 15 to 23 million raw reads per library with average read quality 28. After adaptor trimming, 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) (Additional File 1). Illumina reads are available from the Sequence Read Archive (SRA) at NCBI under accession number SRA SRP082469. After position and sexual effects controlling, we identified 263 up-regulated DE transcripts and 11 down-regulated DE transcripts for the blotch in *C. macrops*; 286 up-regulated DE transcripts and 429 down-regulated DE transcripts for egg-spots in *Astatotilapia burtoni* (Figure 1E).

Comparative transcriptomic analyses for the blotch and egg-spots

Forty-four transcripts showed similar expression patterns in both the blotch and egg-spots. Particularly for the top ten blotch DE transcripts, six showed similar expression patterns in egg-spots, corresponding to two duplicated iridophore-related genes (*pnp4a* and *pnp5a*), the immunity-related gene *ly6d*, the ligand transporter-related gene (*ApoD*), and the two genes associated with egg-spots formation (*fhl2a* and *fhl2b*) (Santos et al. 2014) (Table 1). Besides, two transcripts were highly expressed in egg-spots, but low expressed in the blotch, corresponding to a novel transcript ENSONIT00000023612 and the immunity-related gene *steap4* (Benard et al. 2014) (Table 2). Fifteen transcripts were over-expressed in the blotch but down-regulated in egg-spots (Table 2).

Clustering analysis based on the expression levels of 44 common transcripts clustered anal fin tissues possessing egg-spots and the blotch together, despite their species boundary (Figure 2B). This was not the case for pattern specific DE transcripts (Figure 2A and 2C). Since sexual effect has already been controlled (Figure 1), sex difference was not the reason here. More than 50% Gene Ontology (GO) terms and pathways were shared between the blotch and egg-spots (Figure 3). Pigmentation related GO terms, 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) were enriched (Figure 3). Noticeably, there were more egg-spots specific pathways compared to the blotch (Figure 3 and Figure 4).

Evolutionary rate detection for DE transcripts and related pathways

The average Ka/Ks values showed significantly higher rates for common transcripts than blotch transcripts, but not higher than egg-spots transcripts (Figure 3). Common transcripts also possessed the largest percentage of transcripts under positive selection (average Ka/Ks>1) (Figure 3). Interestingly, despite the large difference in the numbers of down-regulated transcripts between the blotch (11) and egg-spots (429), they both showed significantly higher evolutionary rates than up-regulated transcripts (Figure 3). This pattern can also be found for the percentage of transcripts under positive selection (Figure 3). Noticeably, more than half of the transcripts under positive selection are unannotated novel transcripts (Table 3).

To detect the evolutionary rates of related pathways, we calculated average Ka/Ks values for common pathways and dN/dS values under the free-ratio model for all the pathways (Figure 4). Consistent patterns were found using these two different methods. 1) In general, common transcripts showed higher evolutionary rates than the blotch transcripts within the same pathway, confirming our previous results. 2) Egg-spots transcripts showed higher evolutionary rates, primarily in signalling pathways, i.e. 7 out of 12 (58.3%) pathways under average Ka/Ks value calculations (Figure 4a), and were statistically significant under the free-ratio model (common**) (Figure 4b). The three pathways (9,10 and 11) in Figure 4a with high average Ka/Ks values for the blotch were caused by only one unannotated transcript, ENSONIT00000016796. However, under the free-ratio model after lineage effect controlling, the effect of this transcript was minimized (Figure 4b), further confirming the higher evolutionary rates of egg-spots transcripts in signalling pathways. 3) There was no large difference in the evolutionary rates of other pathways between egg-spots and the blotch, i.e. less than half showed higher average rates in either pattern (Figure 4a and 4b), and no statistical significance was detected under the free-ratio model; however, more egg-spots

specific pathways, together with common pathways, showed higher evolutionary rates than the blotch (all**) (Figure 4b). Noticeably, different patterns were found for the metabolic pathways in different tests, i.e. higher average Ka/Ks values were found for egg-spots (Figure 4a), but no significance was detected under the free-ratio model (Figure 4b). This finding could be because the later method considers the lineage effect. However, egg-spots possess more specific metabolic pathways than the blotch (Figure 4b, Additional file 2).

First common neighbour gene networks (FCNs) of the blotch and egg-spots

To characterize the interactions of DE transcripts, we examined their protein-protein interaction networks. Network enrichment analysis in both the blotch and egg-spots showed that the corresponding 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, further confirming the robustness of our experimental design. To predict the roles of the common genes in the network, we analysed their FCNs (Figure 5), as direct interactions among genes can indicate their roles to the largest extant. These FCNs showed significantly higher interaction degrees than the total gene network (Table 4), reflecting their core roles in the network.

Two types of FCNs were formed: 1) direct interactions among the common genes, comprising pigmentation (melanophore and xanthophore)-related genes (core common network, CCN) (Figure 5). Noticeably, unlike in the blotch, the CCN in egg-spots integrated more patterning-related transcription factors (TFs) and genes belonging to patterning-related signalling pathways. 2) The second type of FCN involved interactions among common genes and pattern specific DE genes (specific common network, SCN). Unlike the CCN, the

common genes in SCN were only loosely associated (Figure 5). For example, the SCN of the blotch could clearly be divided into sub-networks, connected only by two common genes (*myo5aa* and *pde7b*) (Figure 5). In addition, genes related to purine metabolism pathways were detected in the SCN of the blotch (Figure 5).

The evolutionary rates of common genes were significantly higher than those of pattern specific genes in both CCN and SCN of the blotch, but not those of egg-spots (Figure 5), confirming our previous results (Figure 3). In addition, several novel genes were involved in FCNs (Figure 5). 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*, suggesting a xanthophore-related role (Minchin & Hughes 2008; Curran et al. 2010).

Discussion

The thorough comparative transcriptomic and genomic data analysis conducted herein revealed a common genetic basis between the blotch and egg-spots. Focusing on the whole transcriptomic expression level instead of limited numbers of candidate genes can explain the opposite conclusion between our study and the one from Santos et al. (M Emília Santos et al. 2016). Here, we proposed that the common genetic basis is essential for the emergence of these novel phenotypes, and independently evolved associations among common genes with TFs and signalling pathways, as well as the integration of advantageous genes are crucial for the evolution of egg-spots.

Common genetic basis is important for the origin of the blotch and egg-spots

Similar gene expression patterns in convergent taxa are generally associated with conserved and important roles (such as toolkit genes) in the origin of phenotype (Khaitovich et al. 2006; Pankey et al. 2014). In the present study, clustering analysis based on expression level revealed that convergent evolution of anal fin pigmentation patterns is in parallel with gene expression level, at least on the common genetic basis level. Genes *fhl2a* and *fhl2b*, which were suggested to be associated with egg-spots formation (Santos et al. 2014), were also observed ranking at the top of the blotch candidate gene list, further confirming our hypothesis. Common transcripts showed higher accelerated evolutionary rates in total and in the shared pathways, as well as possessed the highest percentage of transcripts under positive selection, evincing their advantageous roles during evolution. The higher average Ka/Ks value of egg-spots transcripts than the result from Santos et al. (M Emília Santos et al. 2016) could be explained by the integration of recent duplicated genes into our analysis, which is too important to ignore (Glasauer & Neuhauss 2014). High evolutionary rates can be explained by the reasoning that the trait itself is under selection, or alternatively that the formation of the trait integrated advantageous genes in the network, which conserved the phenotype during evolution. Since common transcripts also showed significantly higher evolutionary rates than pattern specific transcripts, we prefer the later scenario. In any case, they both suggest that the common genetic basis is important for the origin of these convergent anal fin pigmentation patterns.

Then what is the common genetic basis and its role in the convergent evolution of anal fin pigmentation patterns? Both GO enrichment and gene network construction showed that

the common genes primarily comprise pigmentation-related genes. Indeed, interactions among pigment cells are important for pigmentation pattern formation (Parichy 2009; Parichy & Spiewak 2015; Watanabe & Kondo 2015; Kondo 2017). For example, the mutation of melanophore-related genes (*mitfb*, *gpnmb*, *trpm1a*, *ednrb*, *trpm1b* and *sox10*), the iridophore-related gene (*pnp4a*), the xanthophore-related gene (*pax7*) and the connexion-related gene *cx45.6* can induce irregular stripe pattern formation in zebrafish (Shin et al. 1999; Hou et al. 2006; Minchin & Hughes 2008; Curran et al. 2010; Zhang et al. 2012; Reissmann & Ludwig 2013; Irion et al. 2014), suggesting their important roles in the corresponding core gene network. The genes *fhl2b* (iridophore-related gene), *pax7a* (xanthophore-related gene) and *melanocortin* were also reported to be associated with pigmentation pattern formation either on the anal fin (*fhl2b*) or in skin (*pax7a* and *melanocortin*) in cichlid fishes (Salzburger et al. 2007; Santos et al. 2014; Dijkstra et al. 2017). These genes were in the common candidate gene list in the present study, which can have similar roles in the core gene network of anal fin pigmentation pattern formation. This was further confirmed by the finding that the FCNs showed a significantly higher interaction degree than the average degree of total gene network.

In addition, signalling pathways and metabolism pathways occupied a large portion of the shared pathways between the blotch and egg-spots. It is unsurprising that metabolism pathways were involved in these sexually related traits, considering the important trade-off of energy allocation between the survival and reproduction (Bleu et al. 2016). For example, both patterns contain carotenoid-based reddish colours, a limited resource used in the immune system, which in fish, can only be obtained through diet; it is important to trade off its allocation between immune response and sexual signalling (Theis et al. 2017). In addition, many patterning related signalling pathways were shared between the blotch and egg-spots,

such as the Wnt signalling pathway, ErbB signalling pathway and MAPK signalling pathway (Budi et al. 2008; Martin et al. 2012; Schneider et al. 2012; Plestant & Anton 2013; Martin & Reed 2014), which can have similar roles here and are advantageous for the emergence of these novel patterns**.** The re-use of the same large portion of the metabolic and signalling pathways again indicated their importance in the formation of both egg-spots and the blotch.

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

Although both the blotch and egg-spots are novelties, their patterns are quite different. The blotch is a reddish pattern with an irregular boundary, exhibiting no varieties among species (Brichard 1989); whereas egg-spots are with conspicuous outer rings and exhibit high evolvability including different numbers, sizes and positions within and among species (Santos et al. 2014; Theis et al. 2017). Egg-spots originate once in haplochromine lineage, the most species-richness lineage, and was suggested to be one of the key innovations associated with the adaptive radiation of cichlid fishes based on phylogeny and character reconstruction (Salzburger et al. 2005a, 2007; Salzburger 2009). It is important to understand the mechanism underlying the different patterns and evolvability between the blotch and egg-spots .

Genes are not isolated from each other, but rather remain connected. Although common genes were shared, more specific pathways were found in egg-spots, especially for signalling pathways and metabolic pathways. Additionally, more direct interactions were observed among common genes for egg-spots genes than those of the blotch in FCNs, especially for the connections with patterning related TFs (*crem*, *sox5* and *zeb2b*), and genes belonging to patterning related signalling pathways (*bmp7a*, *wnt7aa* and *kita*) in the CCN of

egg-spots, compared to the only two pigmentation related TFs (*pax7a* and *sox10*) of the blotch (Figure 5) (https://www.ncbi.nlm.nih.gov/gene/). Gene network integrating patterning related TFs and signalling pathways can become relatively independent to integrate signals into a gene expression pattern unique to that character, representing novelty (Wagner 2014), similar to the case of butterfly eyespot (Shirai et al. 2012). The extent of independence of a gene network is important for the evolution of a trait (Wagner 2014): the relatively independent network can free the evolution of egg-spots from the ancestral anal fin network compared to the blotch (intrinsic factor, developmental constraint). In addition, the DE transcripts of egg-spots showed higher evolutionary rates than those of the blotch, both in total and in FCNs, especially in the signalling pathways. Additionally, there are larger percentages of transcripts under positive selection in egg-spots than in the blotch. These findings further suggest that evolutionary advantageous genes can be recruited in the egg-spots gene network, which can be causally linked to the high evolvability of egg-spots (extrinsic factor, selection).

Noticeably, there was a large difference in the numbers of down-regulated DE transcripts between the blotch (11) and egg-spots (429). This difference could be due to the effect of lineage-specific position related transcripts that were not ruled out based on our strategy. However, this explanation is unlikely, considering the very closely related genetic background in cichlid fishes (Baldo et al. 2011; Brawand et al. 2014) and the relatively conservative fin developmental constraint (Maxwell & Wilson 2013; Paço & Freitas 2017). In addition, after position effect controlling, we observed similar numbers of up-regulated DE transcripts between egg-spots and the blotch, suggesting that the comparison strategy was not the issue. Furthermore, down-regulated transcripts of both the blotch and egg-spots showed the same results (clustering analyses and evolutionary rates detection), suggesting that

position effect was not the issue, since the blotch transcripts have already controlled position effect.

One explanation could be the importance of down-regulated transcripts in patterning, especially for egg-spots, although this issue is largely ignored. Indeed, down-regulated transcripts showed significantly higher evolutionary rates than up-regulated transcripts in both the blotch and egg-spots, and possessed a large percentage of transcripts under positive selection, suggesting their advantageous roles. Especially, most of which were novel transcripts, reflecting the novelty. Considering the more complicated pattern of egg-spots compared to the blotch, more down-regulated genes could be essential for its formation. Noticeably, unlike the blotch, there was only a small number of position effect genes in egg-spots (Figure 1E), reflecting the relatively homogenous genetic background of anal fin tissue possessing egg-spots in *A. burtoni*, which could also be one of the factors prompting the diversification of egg-spots on the anal fin.

Conclusions

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

Our findings 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. Although anal fin pigmentation patterns in cichlid fishes are emerging models, stripe pattern formation in

zebrafish can provide some clues. Unlike in mammals, pigment cells in fish are quite diversified, and different pathways are involved in their differentiations (see review from (Parichy 2009)). Stripe pattern formation is driven by the interactions among different pigment cells, and mathematical models have been developed to explain the interactions (Kondo & Miura 2010; Watanabe & Kondo 2015; Kondo 2017). Different pigment cells also have different roles in driving pattern formation in different species (Singh & Nüsslein-Volhard 2015; Patterson et al. 2014). Genes related to endocrine system, such as thyroid hormone, can also affect stripe pattern formation by affecting the development of pigment cells (McMenamin et al. 2014). The mutations of many genes can induce irregular stripe pattern formation in zebrafish, and different genes are involved in pigmentation pattern formation in cichlid fishes, as we mentioned above. These findings reflect that pigmentation pattern formation in fishes is polygenic and involved in different pathways.

Therefore, we propose one hypothesis with respect to a gene network to explain the origin and evolution of convergent novel anal fin pigmentation patterns in cichlid fishes (Figure 6). On one hand, the common genetic basis, comprising pigment cells, signalling pathways and metabolic pathways, is 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. pre-existing pigment cells on the anal fin (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). The similar effects of the initial mutations affecting the common gene network (e.g., pathways and cascade of changes) inspired the emergence of these convergent novel traits. On the other hand, independently evolved associations with patterning related TFs and signalling pathways

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). The relatively homogenous anal fin background could also prompt this diversification. Simultaneously, the integration of advantageous genes, especially those related to signalling pathways into the corresponding gene network can conserve egg-spots and prompt its evolution under selection (extrinsic factor, selection). This hypothesis can provide a clue to identify egg-spots as one of the key innovations to the adaptive radiation in cichlid fishes, which was proposed previously based on phylogeny and character reconstruction (Salzburger et al. 2005a, 2007; Salzburger 2009). In this case, although both are novelties, egg-spots can be viewed as "a novelty out of novelty" compared to the blotch. Further thorough study including more samples across the phylogeny in cichlid fishes will be helpful to test this hypothesis. This hypothesis here can further illuminate the mechanism of the origin and evolution of novelties in a broad sense.

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 with similar body sizes. 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* (Figure 1A). RNA extraction 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 Nanodrop 1000 spectrophotometer (Thermo Scientific, USA) and 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 raw transcriptomic data for anal fin with egg-spots and without egg-spots of *A. burtoni* was retrieved from Santos et al. (M Emília Santos et al. 2016).

Differential gene expression analysis for the blotch and egg-spots

Quality assessment of the sequence reads was conducted using Fastqc 0.10.1 (www.bioinformatics.babraham.ac.uk/projects/). Contaminated Illumina adapter were removed using cutadapt 1.3 (Martin 2011). The reads were subsequently aligned to the Tilapia transcriptome assembly available from Broad Institute (ftp://ftp.ensembl.org/pub/release-81/fasta/oreochromis_niloticus/cdna/. version 0.71). NOVOINDEX (www.novocraft.com/) was used for indexing the reference and NOVOALIGN (www.novocraft.com/) was used for mapping the reads against the reference.

The output SAM files of read mapping were subsequently transformed into the BAM format using SAMtools (Li et al. 2009). The count files were concatenated into count tables and analysed using the Bionconductor R package (Gentleman et al. 2004; Huber et al. 2015) and edgeR (Robinson & Smyth 2007; Robinson et al. 2010; Zhou et al. 2013). For paired tissues, we used the individual as the block factor. In the analyses with edgeR, genes that achieved at least one count per million (cpm) in at least three samples were maintained. Differentially expressed transcripts were maintained if the false discovery rate (FDR) was smaller than 0.01.

To obtain blotch candidate genes to the largest extant, the position and sexual effects were controlled (for details see Figure 1A). We did not use the same strategy to find egg-spots candidate genes, since the females of *A. burtoni* also possess egg-spots, although they are less conspicuous than those in males (Theis et al. 2017). Therefore, we corrected the position effect for egg-spots using position genes retrieved from *C. macrops*. Considering the very closely related genetic background in cichlid fishes (Baldo et al. 2011; Brawand et al. 2014) and the conservative fin development constraint (Maxwell & Wilson 2013; Paço & Freitas 2017), the effect of species-specific position genes should be small, and thus, we can get the egg-spots candidate genes to the largest extant. In addition, for finding common genes between the blotch and egg-spots, the comparison itself has already perfectly corrected the position effect considering their different locations on the anal fin (Figure 1).

Comparative transcriptomics between the blotch and egg-spots

Gene expression clustering was visualized as a heatmap using the *pheatmap* R package v1.0.8 (https://cran.r-project.org/web/packages/pheatmap/index.html). GO annotation of the differential expressed transcripts was conducted with Blast2GO version 2.5.0 (Conesa

et al. 2005). BLASTx searches were achieved using BLASTx (threshold: e-6) against zebrafish (*Danio rerio*) protein database and a number of hits of 10. KEGG (Kyoto Encyclopaedia of Genes and Genomes) pathway annotation was conducted with KAAS (KEGG Automatic Annotation Server) http://www.genome.jp/tools/kaas/ (Moriya et al. 2007) using zebrafish as the reference with a threshold e-value of e^{-10}. Enrichment analysis of the GO terms was performed using DAVID (https://david.ncifcrf.gov/) with tilapia genome as the background. Protein-protein interaction network analysis was conducted with the string database (http://string-db.org/) with tilapia genome as the reference. The FCNs were extracted and visualized with Cytoscape v3.0 (Shannon et al. 2003). The bootstrap method was used to detect the significance of interaction degrees. Briefly, we randomly extracted the same numbers of the genes as we would like to detect in the network for 10,000 times, and calculated the corresponding average degree every time to get the distribution first. Then, we tested whether the average degrees of the target genes was located outside the 95% confidence interval.

Evolutionary rates detection of DE transcripts and related pathways

To detect the evolutionary rate of the DE transcripts, we extracted the corresponding transcripts of tilapia retrieved from the Ensembl database (http://www.ensembl.org/index.html) first. These transcripts were used as references after removing stop codon with Biopython (http://biopython.org/wiki/Biopython). The raw reads of transcriptomic data from the other four available cichlid fishes (*Pundamila nyererei*, *Neolamprologus brichardi*, *Maylandia zebra* and *A. burtoni*) were retrieved from the NCBI database (https://www.ncbi.nlm.nih.gov/). These data, together with transcriptomic data of egg-spots and the blotch, were mapped to individual transcripts trimmed to the reference

using Geneious (Kearse et al. 2012). To retrieve recent duplicated genes, we differentiated the variable single nucleotide polymorphisms deriving from duplicates by mapping these anomalies against the reference. To do this, we concatenated all the transcripts for each species using Geneious followed by visual assessment. Subsequently, we constructed two rounds of consensus sequences through pairwise alignment between the target species and tilapia using MAFFT, which removes the variable single nucleotide polymorphism according to the reference. The resulting concatenated sequences were re-mapped using existing transcriptome data and translated into amino acids followed by visual assessment. Afterwards, individual transcripts were retrieved with Geneious.

To detect the evolutionary rate of these DE transcripts, we aligned the sequences using T_Coffee (Notredame et al. 2000) based on the codon. Subsequently, pairwise Ka/Ks values were calculated using yn00 in Ka/Ks calculator (Wang et al. 2010). To keep the comparison unbiased, we removed the gene pairs when any NA (not applicable or infinite) value occurred (Wang et al. 2011). To determine whether the same pattern also occurred in a lineage-specific manner, we detected the dN/dS values for individual DE transcripts under the free-ratio model within codeml in PAML (Yang 1997, 2007).

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LG designed the experiment, performed the RNAseq library construction and did data analysis. LG constructed the figures and did the analysis of interaction degree of gene networks with the help of CX. Both authors read and approved the manuscript.

Acknowledgements

We particularly thank Prof. Walter Salzburger for the support. We also thank Dr. Emilia M. Santos for providing the raw transcriptomic data of egg-spots. Many thanks to Chao Zhang, Shengkai Pan, Yongjian Zhang, Lei Luo, Yongsheng Li, Jie Zhang, Tingting Zhou, De Chen and Guang Yang 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 and suggestions. Many thanks to Nicolas Boileau for the assistance in the lab and Brigitte Aeschbach for the laboratory apparatus organization. Pictures were provided by Heinz Büscher, Anya Theis and Adrian Indermau. This work was supported by the PhD grant from University of Basel, Switzerland and Postdoc funding from Southwest University, China. Illumina reads from this study are available at NCBI under the accession number SRP082469. Alignment sequences are available upon request from the corresponding author.

Figures legends

Figure 1 Comparative transcriptomic analysis to find candidate genes for the blotch and egg-spots. (A) a. Four kinds of tissues (blotch (b) and non-blotch (n) in males; corresponding position in females, up (u) and down (d)). b. Different effects can be involved in the candidate genes resulting from different comparisons. c. Different comparisons to find candidate genes for the blotch. (B) a. Two kinds of tissues (egg-spots and non-egg-spots) were used in the

males of *A. burtoni*, and since the females also possess egg-spots, they cannot be used as the control. b. Comparative transcriptomic analyses to find candidate genes of egg-spots with position effect controlling. (C) Comparisons to find common and pattern specific candidates for the blotch and egg-spots. (D) Convergent evolution of egg-spots (originating once in the most species-rich lineage, haplochromine lineage) and the blotch (ectodine lineage). The area of the triangle represents the species richness based on the study from (Salzburger et al. 2005b) and https://en.wikipedia.org. The schematic molecular phylogeny was based on the combined evidences from previous studies (Salzburger et al. 2005b; Meyer et al. 2015; Takahashi & Sota 2016). (E) The numbers of candidate transcripts.

Figure 2 Clustering analysis based on the expression levels of common differentially expressed (DE) transcripts and pattern specific DE transcripts of anal fin pigmentation patterns. The cluster analysis was conducted with the blotch and non-blotch tissues in the male and the corresponding tissues (up tissue and down tissue) in the female of cichlid fish *Callochromis macrops*, and egg-spots and non-egg-spots tissues in the male of cichlid fish *Astatotilapia burtoni*. Orange rectangles showed the clustered anal fin tissues.

Figure 3 A common genetic basis was shared between the blotch and egg-spots with accelerated evolutionary rates. (A) > 50% common shared gene ontology (GO) terms and pathways between the blotch and egg-spots. (B) Common enriched GO terms between the blotch and egg-spots (FDR<0.1). (C) Average evolutionary rates (Ka/Ks) of common shared differentially expressed (DE) transcripts, the blotch-specific DE transcripts and egg-spots-specific DE transcripts. ** represents $p<0.01$. (D) Statistics of the transcripts under positive selection with average Ka/Ks values larger than 1. Total numbers were given in the boxes.

Figure 4 Evolutionary rates calculations for anal fin pigmentation differentially expressed (DE) transcripts related pathways. (A) Average evolutionary rates (Ka/Ks) values of common shared pathways between the blotch and egg-spots. To keep the comparison unbiased, gene pairs with any NA (not applicable or infinite) value occurred were removed (Wang et al. 2011), and thus, 43 common pathways were compared in total. For details see Additional File 2. * represents $p<0.05$; and ** represents $p<0.01$. (B) Rates of evolution (dN/dS) for all pathways under free-ratio model in a lineage-specific manner within codeml in PAML. Fold-change ratios of dN/dS in the corresponding lineage were calculated and compared. For details see Additional File 2. **represents $p<0.01$; **ns** represents no significance; **common** represents the common shared pathways; and **all** represents all the related pathways. Species *Pundamilia nyererei*, *Maylandia zebra*, *Astatotilapia burtoni* possess egg-spots, species *Callochromis macrops* possess the blotch, species *Neolamprologus brichardi*, *Oreochromis niloticus* do not possess either the blotch or egg-spots. The phylogeny was based on studies from (Meyer et al. 2015; Brawand et al. 2014; Salzburger et al. 2005a, 2007).

Figure 5 First common neighbour gene interaction network (FCN) analysis for the blotch and egg-spots. (A) FCNs between the blotch and egg-spots, including the core common gene network (CCN) primarily comprising the direct interactions among pigment cells (above), and the specific core gene network (SCN) comprising the interactions among common genes and pattern specific genes (below). (B) Average evolutionary rates (Ka/Ks) comparison of different genes of the CCN and SCN in the blotch and egg-spots. **represents $p<0.01$. (C) Comparison of the percentage of pattern specific transcription factors (TFs) and genes belonging to the patterning related signalling pathways in CCNs between the blotch and egg-spots.

Figure 6 Hypothesis with respect to a gene network to explain the origin and evolution of convergent novel anal fin pigmentation patterns in cichlid fishes. On one hand, the common genetic basis is important for the emergence of these novel traits (intrinsic factor, developmental constraint). These genes could be derived from deep homology or evolve independently. Accelerated evolutionary rates of the common genetic basis help conserve these novelties during evolution (extrinsic factor, selection). Similar effects of the initial mutations affecting the common gene network (pathways and cascade of changes) inspired the emergence of these convergent novelties. On the other hand, independently evolved associations with patterning related TFs and signalling pathways 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 (intrinsic factor, developmental constraint). The relatively homogenous anal fin background could also prompt this diversification. Simultaneously, the integration of advantageous genes, especially those related to signalling pathways and egg-spots specific pathways, into the corresponding gene network can conserve egg-spots and prompt its evolution under selection (extrinsic factor, selection). This hypothesis provide a clue to identify egg-spots as one of the key innovations to the adaptive radiation in cichlid fishes (Salzburger et al. 2005a, 2007; Salzburger 2009). In this case, although both are novelties, egg-spots can be viewed as “a novelty out of novelty” compared to the blotch.

Table 1 Top ten common blotch DE transcripts and 11 down-regulated blotch DE transcripts

	name	Ensembl	logFC the blotch	Ensembl Description
1	apod	ENSONIT00000023475	-3.58	Apolipoprotein Da
2	fhl2b	ENSONIT00000017889	-3.56	four and a half LIM domains 2b
3	pnp4a	ENSONIT00000000731	-3.52	Purine nucleoside phosphorylase 4a
4	ly6d	ENSONIT00000024767	-3.5	lymphocyte antigen 6D-like
5	fhl2a	ENSONIT00000015512	-3.46	four and a half LIM domains 2a
6	cecr5	ENSONIT00000000334	-3.32	cat eye syndrome critical region protein 5-like
7	ifi30	ENSONIT00000016006	-3.29	interferon, gamma-inducible protein 30
8	gpnmb	ENSONIT00000005686	-3.21	Glycoprotein (transmembrane) nmb
9	novel	ENSONIT00000005261	-2.88	novel transcript
10	plin5	ENSONIT00000002684	-2.87	perilipin-5-like
11	novel	ENSONIT00000011328	0.55	novel transcript
12	steap4	ENSONIT00000007508	0.61	STEAP family member 4
13	novel	ENSONIT00000018333	0.82	novel transcript
14	novel	ENSONIT00000000470	0.95	novel transcript
15	novel	ENSONIT00000002272	0.97	novel transcript
16	col2a1b	ENSONIT00000006473	1.1	collagen, type II, alpha 1b
17	novel	ENSONIT00000004672	1.49	novel transcript
18	novel	ENSONIT00000023704	1.96	novel transcript
19	novel	ENSONIT00000023612	2.17	novel transcript
20	novel	ENSONIT00000012084	2.47	novel transcript
21	clec3a	ENSONIT00000007521	2.75	C-type lectin domain family 3 member A

Table 2 Differential expressed transcripts 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	novel	ENSONIT00000023612	2.17	-0.94	novel transcript

Table 3 Transcripts under positive selection with average Ka/Ks larger than 1

Name	ID	Ka/Ks	expression	tissues	Ensembl Description
map7a	ENSONIT00000000906	1.37	up-regulated	both	Microtubule-associated protein 7a
uraha	ENSONIT00000007140	1.29	up-regulated	both	Urate (5-hydroxyiso-) hydrolase a
FAM180A	ENSONIT00000013027	1.004	up-regulated	both	Family with sequence similarity 180 member A
bscl2l	ENSONIT00000006721	1.23	up-regulated	the blotch	Bernardinelli-Seip congenital lipodystrophy 2, like
novel	ENSONIT00000000470	1.89	down-regulated	the blotch	unannotated
novel	ENSONIT00000012084	1.12	down-regulated	the blotch	unannotated
novel	ENSONIT00000002070	1.65	up-regulated	egg-spots	unannotated
novel	ENSONIT00000010127	1.14	up-regulated	egg-spots	unannotated
novel	ENSONIT00000022442	1.12	up-regulated	egg-spots	unannotated
novel	ENSONIT00000025818	1.1	up-regulated	egg-spots	unannotated
spam1	ENSONIT00000005048	1.02	up-regulated	egg-spots	sperm adhesion molecule 1
novel	ENSONIT00000025871	3.61	down-regulated	egg-spots	unannotated
novel	ENSONIT00000026449	3.13	down-regulated	egg-spots	unannotated
trappc2	ENSONIT00000020233	2.32	down-regulated	egg-spots	Trafficking protein particle complex 2
novel	ENSONIT00000016236	2.13	down-regulated	egg-spots	unannotated
novel	ENSONIT00000017692	1.94	down-regulated	egg-spots	unannotated
novel	ENSONIT00000006770	1.66	down-regulated	egg-spots	unannotated
novel	ENSONIT00000000523	1.48	down-regulated	egg-spots	unannotated
novel	ENSONIT00000025911	1.47	down-regulated	egg-spots	unannotated
mgst3a	ENSONIT00000023478	1.4	down-regulated	egg-spots	Microsomal glutathione S-transferase 3a
novel	ENSONIT00000021326	1.4	down-regulated	egg-spots	unannotated
novel	ENSONIT00000005252	1.15	down-regulated	egg-spots	unannotated
novel	ENSONIT00000004956	1.14	down-regulated	egg-spots	unannotated
crema	ENSONIT00000018860	1.11	down-regulated	egg-spots	CAMP responsive element modulator a
isg15	ENSONIT00000026211	1.09	down-regulated	egg-spots	ISG15 ubiquitin-like modifier
igfbp6a	ENSONIT00000024154	1.08	down-regulated	egg-spots	Insulin-like growth factor binding protein 6a
timp2b	ENSONIT00000006684	1.07	down-regulated	egg-spots	TIMP metallopeptidase inhibitor 2b
OLA1	ENSONIT00000010893	1.06	down-regulated	egg-spots	Obg like ATPase 1
stard14	ENSONIT00000003484	1.06	down-regulated	egg-spots	START domain containing 14
novel	ENSONIT00000001384	1.04	down-regulated	egg-spots	unannotated
novel	ENSONIT00000009645	1.03	down-regulated	egg-spots	unannotated

Table 4 Statistics of interaction degree of first common neighbour gene networks

comparison	average	p-value	significance
egg-spots fcn vs. egg-spots total	8.62 vs. 5.66	0	**
egg-spots ccn vs. egg-spots total	8.74 vs. 5.66	0.02	*
egg-spots scn vs. egg-spots total	8.58 vs. 5.66	0	**
the blotch fcn vs the blotch total	6.28 vs. 4.11	0	**
the blotch ccn vs. the blotch total	6.63 vs. 4.11	0.01	*
the blotch scn vs. the blotch total	6.17 vs. 4.11	0	**

fcn, first neighbour gene network; ccn, core common network; scn, specific common network.

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

Additional File 2: Evolutionary rates detection of pathways for the blotch and egg-spots.

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.

Baldo L, Santos ME, Salzburger W. 2011. Comparative transcriptomics of Eastern African cichlid fishes shows signs of positive selection and a large contribution of untranslated regions to genetic diversity. Genome Biol. Evol. 3:443–55. doi: 10.1093/gbe/evr047.

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.

Berens AJ, Hunt JH, Toth AL. 2015. Comparative Transcriptomics of Convergent Evolution: Different Genes but Conserved Pathways Underlie Caste Phenotypes across Lineages of Eusocial Insects. Mol. Biol. Evol. 32:690–703. doi: 10.1093/molbev/msu330.

Bhullar B-AS et al. 2015. A molecular mechanism for the origin of a key evolutionary innovation, the bird beak and palate, revealed by an integrative approach to major transitions in vertebrate history. Evolution. 69:1665–77. doi: 10.1111/evo.12684.

Bleu J, Gamelon M, Sæther B-E. 2016. Reproductive costs in terrestrial male vertebrates: insights from bird studies. Proc. R. Soc. B Biol. Sci. 283:20152600. doi: 10.1098/rspb.2015.2600.

Brawand D et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513:375–81. doi: 10.1038/nature13726.

Brichard P. 1989. *Pierre Brichard's Book of Cichlids and All the Other Fishes of Lake Tanganyika*. TFH Publications.

Bright JA, Marugán-Lobón J, Cobb SN, Rayfield EJ. 2016. The shapes of bird beaks are highly controlled by nondietary factors. Proc. Natl. Acad. Sci. U. S. A. 113:5352–5357. doi: 10.1073/pnas.1602683113.

Budi EH, Patterson LB, Parichy DM. 2008. Embryonic requirements for ErbB signaling in neural crest development and adult pigment pattern formation. Development. 135:2603–14. doi: 10.1242/dev.019299.

Chen L, DeVries AL, Cheng CH. 1997. Convergent evolution of antifreeze glycoproteins in Antarctic notothenioid fish and Arctic cod. Proc. Natl. Acad. Sci. U. S. A. 94:3817–22. http://www.ncbi.nlm.nih.gov/pubmed/9108061 (Accessed October 15, 2016).

Clark-Hachtel CM, Tomoyasu Y. 2016. Exploring the origin of insect wings from an evo-devo perspective. Curr. Opin. Insect Sci. 13:77–85. doi: 10.1016/j.cois.2015.12.005.

Colombo M et al. 2013. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol. Ecol. 22:670–84. doi: 10.1111/mec.12029.

Conesa A et al. 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 21:3674–6. doi: 10.1093/bioinformatics/bti610.

Cracraft J. 2005. Phylogeny and evo-devo: Characters, homology, and the historical analysis of the evolution of development. Zoology. 108:345–356. doi: 10.1016/j.zool.2005.09.003.

Curran K et al. 2010. Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafish neural crest. Dev. Biol. 344:107–18. doi: 10.1016/j.ydbio.2010.04.023.

Dijkstra PD et al. 2017. The melanocortin system regulates body pigmentation and social behaviour in a colour polymorphic cichlid fish. Proc. R. Soc. B Biol. Sci. 284:20162838. doi:

10.1098/rspb.2016.2838.

Faunes M, Francisco Botelho J, Ahumada Galleguillos P, Mpodozis J. 2015. On the hodological criterion for homology. Front. Neurosci. 9:223. doi: 10.3389/fnins.2015.00223.

Gentleman RC et al. 2004. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 5:R80. doi: 10.1186/gb-2004-5-10-r80.

Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics. 289:1045–60. doi: 10.1007/s00438-014-0889-2.

Hall BK. 2013. Homology, homoplasy, novelty, and behavior. Dev. Psychobiol. 55:4–12. doi: 10.1002/dev.21039.

Hert E. 1989. The function of egg-spots in an African mouth-brooding cichlid fish. Anim. Behav. 37:726–732. doi: 10.1016/0003-3472(89)90058-4.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 313:101–4. doi: 10.1126/science.1126121.

Hou L, Arnheiter H, Pavan WJ. 2006. Interspecies difference in the regulation of melanocyte development by SOX10 and MITF. Proc. Natl. Acad. Sci. U. S. A. 103:9081–5. doi: 10.1073/pnas.0603114103.

Huber W et al. 2015. Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods. 12:115–21. doi: 10.1038/nmeth.3252.

Irion U et al. 2014. Gap junctions composed of connexins 41.8 and 39.4 are essential for colour pattern formation in zebrafish. Elife. 3:e05125. doi: 10.7554/eLife.05125.

Joron M, Jiggins CD, Papanicolaou A, McMillan WO. 2006. Heliconius wing patterns: an evo-devo model for understanding phenotypic diversity. Heredity (Edinb). 97:157–67. doi: 10.1038/sj.hdy.6800873.

Kearse M et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–9. doi: 10.1093/bioinformatics/bts199.

Khaitovich P, Enard W, Lachmann M, Pääbo S. 2006. Evolution of primate gene expression. Nat. Rev. Genet. 7:693–702. doi: 10.1038/nrg1940.

Kocher TD. 2004. Adaptive evolution and explosive speciation: the cichlid fish model. Nat. Rev. Genet. 5:288–98. doi: 10.1038/nrg1316.

Kondo S. 2017. An updated kernel-based Turing model for studying the mechanisms of biological pattern formation. J. Theor. Biol. 414:120–127. doi: 10.1016/j.jtbi.2016.11.003.

Kondo S, Miura T. 2010. Reaction-diffusion model as a framework for understanding biological pattern formation. Science. 329:1616–20. doi: 10.1126/science.1179047.

Li H et al. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 25:2078–9. doi: 10.1093/bioinformatics/btp352.

Martin A et al. 2012. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc. Natl. Acad. Sci. U. S. A. 109:12632–7. doi: 10.1073/pnas.1204800109.

Martin A, Reed RD. 2014. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. Dev. Biol. 395:367–78. doi: 10.1016/j.ydbio.2014.08.031.

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 17:10. doi: 10.14806/ej.17.1.200.

Maxwell EE, Wilson LA. 2013. Regionalization of the axial skeleton in the ‘ambush predator’ guild – are there developmental rules underlying body shape evolution in ray-finned fishes? BMC Evol. Biol. 13:265. doi: 10.1186/1471-2148-13-265.

McCune AR, Schimenti JC. 2012. Using genetic networks and homology to understand the

evolution of phenotypic traits. Curr. Genomics. 13:74–84. doi: 10.2174/138920212799034785.

McMenamin SK et al. 2014. Thyroid hormone-dependent adult pigment cell lineage and pattern in zebrafish. Science. 345:1358–61. doi: 10.1126/science.1256251.

Meyer BS, Matschiner M, Salzburger W. 2015. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol. Phylogenet. Evol. 83:56–71. doi: 10.1016/j.ympev.2014.10.009.

Minchin JEN, Hughes SM. 2008. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. Dev. Biol. 317:508–22. doi: 10.1016/j.ydbio.2008.02.058.

Monteiro A. 2015. Origin, development, and evolution of butterfly eyespots. Annu. Rev. Entomol. 60:253–71. doi: 10.1146/annurev-ento-010814-020942.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 35:W182–5. doi: 10.1093/nar/gkm321.

Müller GB, Wagner GP. 1991. Novelty in Evolution: Restructuring the Concept. Annu. Rev. Ecol. Syst. 22:229–256. doi: 10.1146/annurev.es.22.110191.001305.

Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22:2362–8. doi: 10.1016/j.cub.2012.10.048.

Nagy LG et al. 2014. Latent homology and convergent regulatory evolution underlies the repeated emergence of yeasts. Nat. Commun. 5:4471. doi: 10.1038/ncomms5471.

Notredame C, Higgins DG, Heringa J. 2000. T-coffee: a novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. 302:205–217. doi: 10.1006/jmbi.2000.4042.

Paço A, Freitas R. 2017. Hox D genes and the *fin-to-limb* transition: Insights from fish studies. genesis. doi: 10.1002/dvg.23069.

Panchen AL. 2007. Homology - History of a Concept. In: John Wiley & Sons, Ltd. pp. 5–23. doi: 10.1002/9780470515655.ch2.

Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. 2014. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proc. Natl. Acad. Sci. U. S. A. 111:E4736–42. doi: 10.1073/pnas.1416574111.

Parichy DM. 2009. Animal pigment pattern: an integrative model system for studying the development, evolution, and regeneration of form. Semin. Cell Dev. Biol. 20:63–4. doi: 10.1016/j.semcdb.2008.12.010.

Parichy DM, Spiewak JE. 2015. Origins of adult pigmentation: diversity in pigment stem cell lineages and implications for pattern evolution. Pigment Cell Melanoma Res. 28:31–50. doi: 10.1111/pcmr.12332.

Patterson LB, Bain EJ, Parichy DM. 2014. Pigment cell interactions and differential xanthophore recruitment underlying zebrafish stripe reiteration and Danio pattern evolution. Nat. Commun. 5:5299. doi: 10.1038/ncomms6299.

Pigliucci M, Müller GB. 2010. Evolution: The Extended Synthesis. MIT Press.

Plestant C, Anton ES. 2013. Scaling the MAPK Signaling Threshold during CNS Patterning. Dev. Cell. 25:221–222. doi: 10.1016/j.devcel.2013.04.014.

Reissmann M, Ludwig A. 2013. Pleiotropic effects of coat colour-associated mutations in humans, mice and other mammals. Semin. Cell Dev. Biol. 24:576–586. doi: 10.1016/j.semcdb.2013.03.014.

Robinson MD, McCarthy DJ, Smyth GK. 2010. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 26:139–40. doi: 10.1093/bioinformatics/btp616.

Robinson MD, Smyth GK. 2007. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 23:2881–7. doi: 10.1093/bioinformatics/btm453.

Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18:169–85. doi: 10.1111/j.1365-294X.2008.03981.x.

Salzburger W, Braasch I, Meyer A. 2007. Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biol. 5:51. doi: 10.1186/1741-7007-5-51.

Salzburger W, Mack T, Verheyen E, Meyer A. 2005a. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5:17. doi: 10.1186/1471-2148-5-17.

Salzburger W, Mack T, Verheyen E, Meyer A. 2005b. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5:17. doi: 10.1186/1471-2148-5-17.

San-Jose LM et al. 2015. Effect of the MC1R gene on sexual dimorphism in melanin-based colorations. Mol. Ecol. 24:2794–808. doi: 10.1111/mec.13193.

Santos ME et al. 2016. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. BMC Genomics. 17:712. doi: 10.1186/s12864-016-3046-y.

Santos ME et al. 2016. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. BMC Genomics. 17:712. doi: 10.1186/s12864-016-3046-y.

Santos ME et al. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:5149. doi: 10.1038/ncomms6149.

Santos ME, Salzburger W. 2012. Evolution. How cichlids diversify. Science. 338:619–21. doi: 10.1126/science.1224818.

Schneider PN, Slusarski DC, Houston DW. 2012. Differential role of Axin RGS domain function in Wnt signaling during anteroposterior patterning and maternal axis formation. PLoS One. 7:e44096. doi: 10.1371/journal.pone.0044096.

Shannon P et al. 2003. Cytoscape: a software environment for integrated models of

biomolecular interaction networks. Genome Res. 13:2498–504. doi: 10.1101/gr.1239303.

Shen Y-Y, Liang L, Li G-S, Murphy RW, Zhang Y-P. 2012. Parallel evolution of auditory genes for echolocation in bats and toothed whales. PLoS Genet. 8:e1002788. doi: 10.1371/journal.pgen.1002788.

Shimeld SM, Holland PWH. 2000. Vertebrate innovations. Proc. Natl. Acad. Sci. 97:4449–4452. doi: 10.1073/pnas.97.9.4449.

Shin MK, Levorse JM, Ingram RS, Tilghman SM. 1999. The temporal requirement for endothelin receptor-B signalling during neural crest development. Nature. 402:496–501. doi: 10.1038/990040.

Shirai LT et al. 2012. Evolutionary history of the recruitment of conserved developmental genes in association to the formation and diversification of a novel trait. BMC Evol. Biol. 12:21. doi: 10.1186/1471-2148-12-21.

Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. Nature. 457:818–23. doi: 10.1038/nature07891.

Singh AP, Nüsslein-Volhard C. 2015. Zebrafish Stripes as a Model for Vertebrate Colour Pattern Formation. Curr. Biol. 25:R81–R92. doi: 10.1016/j.cub.2014.11.013.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol. 30:159–65. doi: 10.1016/j.pbi.2016.03.015.

Stern DL. 2013. The genetic causes of convergent evolution. Nat. Rev. Genet. 14:751–64. doi: 10.1038/nrg3483.

Takahashi T, Sota T. 2016. A robust phylogeny among major lineages of the East African cichlids. Mol. Phylogenet. Evol. 100:234–242. doi: 10.1016/j.ympev.2016.04.012.

Theis A et al. 2017. Variation of anal fin egg-spots along an environmental gradient in a haplochromine cichlid fish. Evolution (N. Y). doi: 10.1111/evo.13166.

Theis A, Bosia T, Roth T, Salzburger W, Egger B. 2015. Egg-spot pattern and body size

asymmetries influence male aggression in haplochromine cichlid fishes. Behav. Ecol. arv104. doi: 10.1093/beheco/arv104.

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.

Wang D, Liu F, Wang L, Huang S, Yu J. 2011. Nonsynonymous substitution rate (Ka) is a relatively consistent parameter for defining fast-evolving and slow-evolving protein-coding genes. Biol. Direct. 6:13. doi: 10.1186/1745-6150-6-13.

Wang D, Zhang Y, Zhang Z, Zhu J, Yu J. 2010. KaKs_Calculator 2.0: a toolkit incorporating gamma-series methods and sliding window strategies. Genomics. Proteomics Bioinformatics. 8:77–80. doi: 10.1016/S1672-0229(10)60008-3.

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).

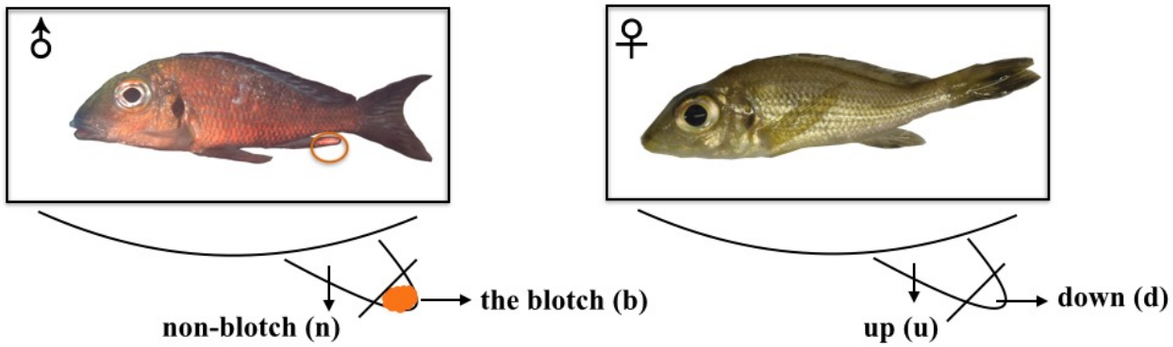
Yong LW, Yu J-K. 2016. Tracing the evolutionary origin of vertebrate skeletal tissues: insights from cephalochordate amphioxus. Curr. Opin. Genet. Dev. 39:55–62. doi: 10.1016/j.gde.2016.05.022.

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.

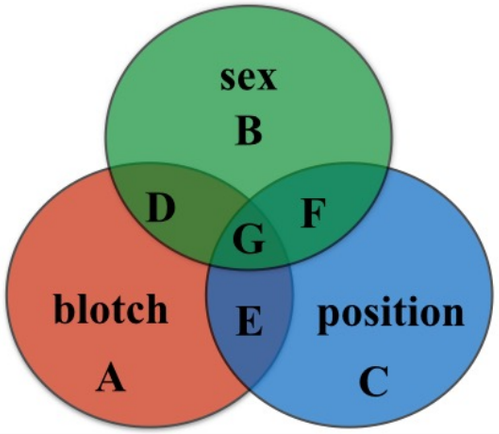
Zhou X, Lindsay H, Robinson MD. 2013. Robustly detecting differential expression in RNA sequencing data using observation weights. 18. http://arxiv.org/abs/1312.3382 (Accessed April 12, 2016).

A. Comparative transcriptomic analysis to find the blotch related candidate genes

a. anal fin tissues from *Callochromis macrops*



b. candidate genes related to different effects



c. comparative transcriptomic analysis

bn=blotch+position=A+D+E+G+F+C

bd=blotch +sex=A+D+E+G+F+B

bn&bd=A+D+E+G+F=A+D+E=blotch related genes (because G+F=0)

bn-(bn&bd)=C=position specific genes

du=position*=E+G+F+C

bn&bd&du=E+G+F=E=(position* +blotch) genes (because G+F=0)

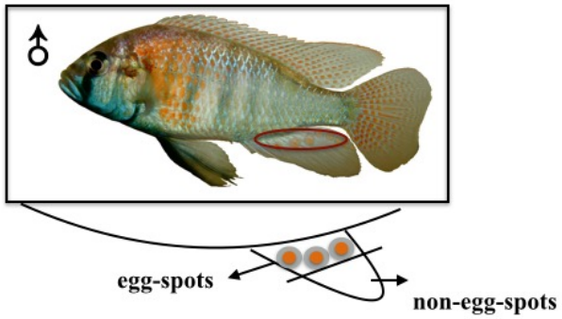
nu=sex=F+G+D+B

bn&bd&nu=D+G+F=D=(sex + blotch) genes (because G+F=0)

(bn&bd&du)&(bn&bd&nu)=G+F=0 genes

B. Comparative transcriptomics to find egg-spots related candidate genes

a. anal fin tissues from *Astatotilapia burtoni*



b. comparative transcriptomic analysis

egg-spots vs. non-egg-spots

↓

egg-spots + position

correct position effect with bn-(bn&bd) in close related species C. macrops

↓

egg-spots related genes

C. Comparisons of candidate genes between the blotch and egg-spots

blotch related genes A+D+E — anal fin with the blotch vs. anal fin with egg-spots — egg-spots related genes

common expressed genes

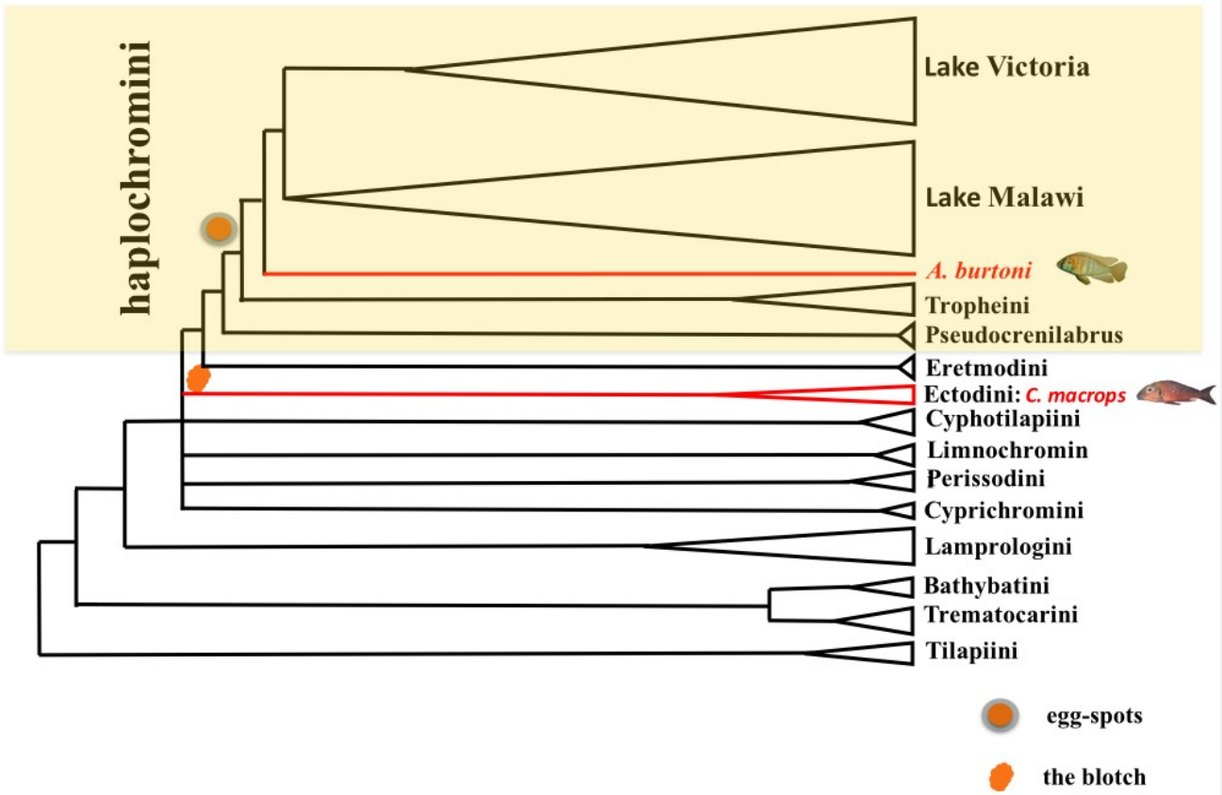
correct position effect perfectly by the comparison itself

related to convergent anal fin pigmentation patterns

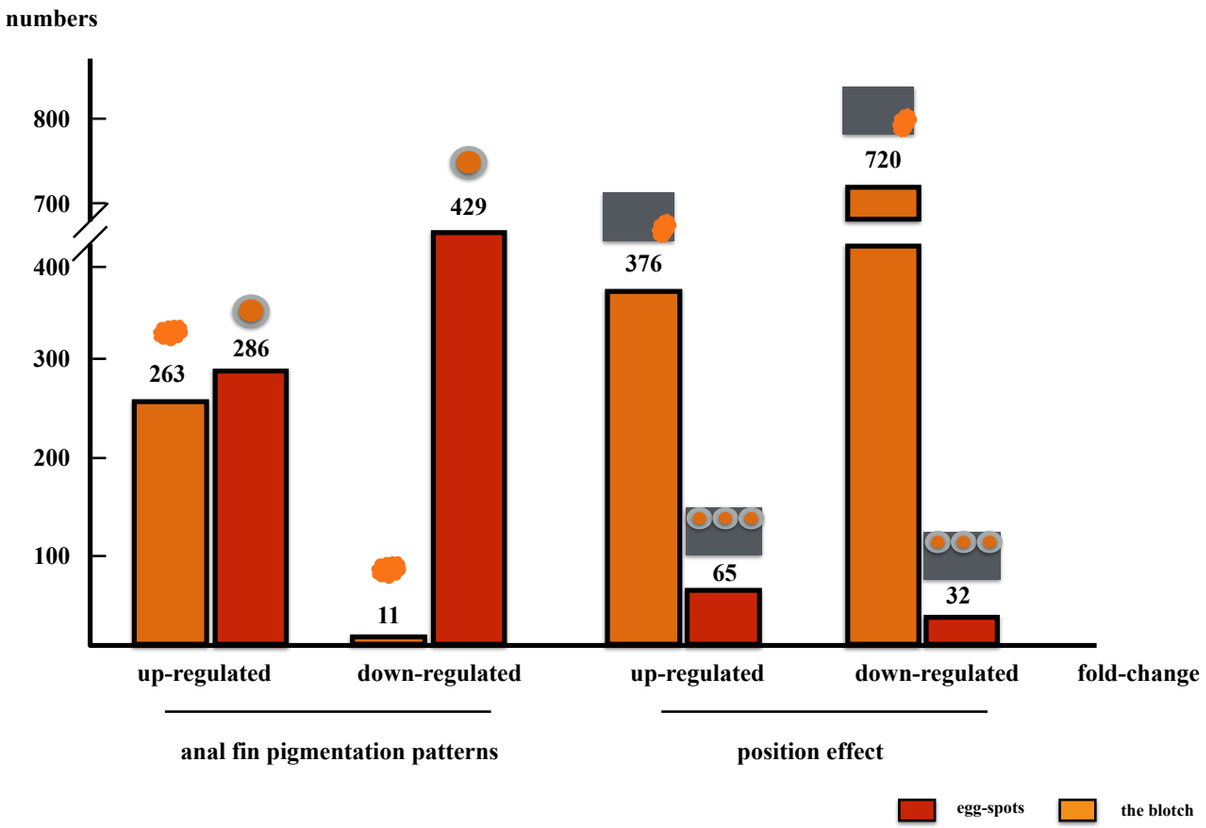
differential expressed genes

related to pattern-specific phenotypes

D. Convergent evolution of egg-spots and the blotch in East African cichlid fishes



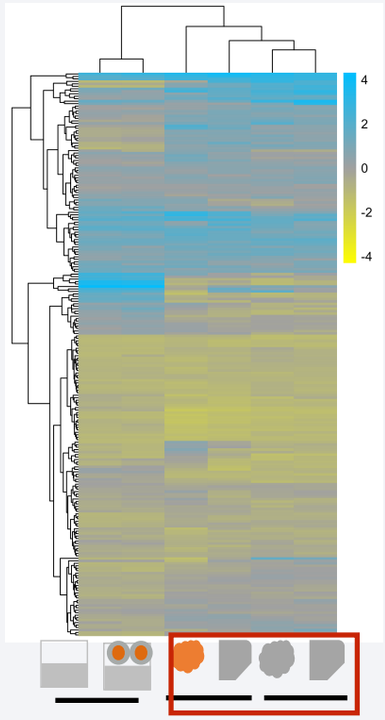
E. Comparison of the numbers of differential expressed transcripts



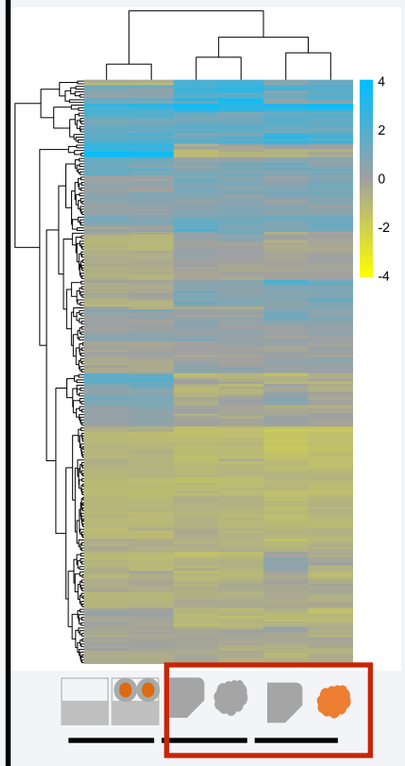
Cluster analysis based on whole gene expression level

A the blotch specific DE transcripts

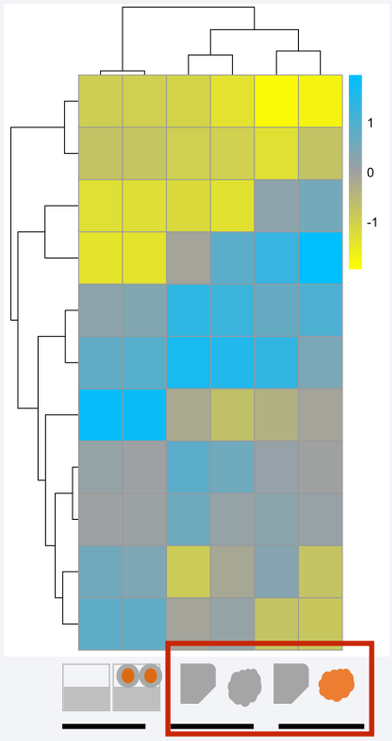
total



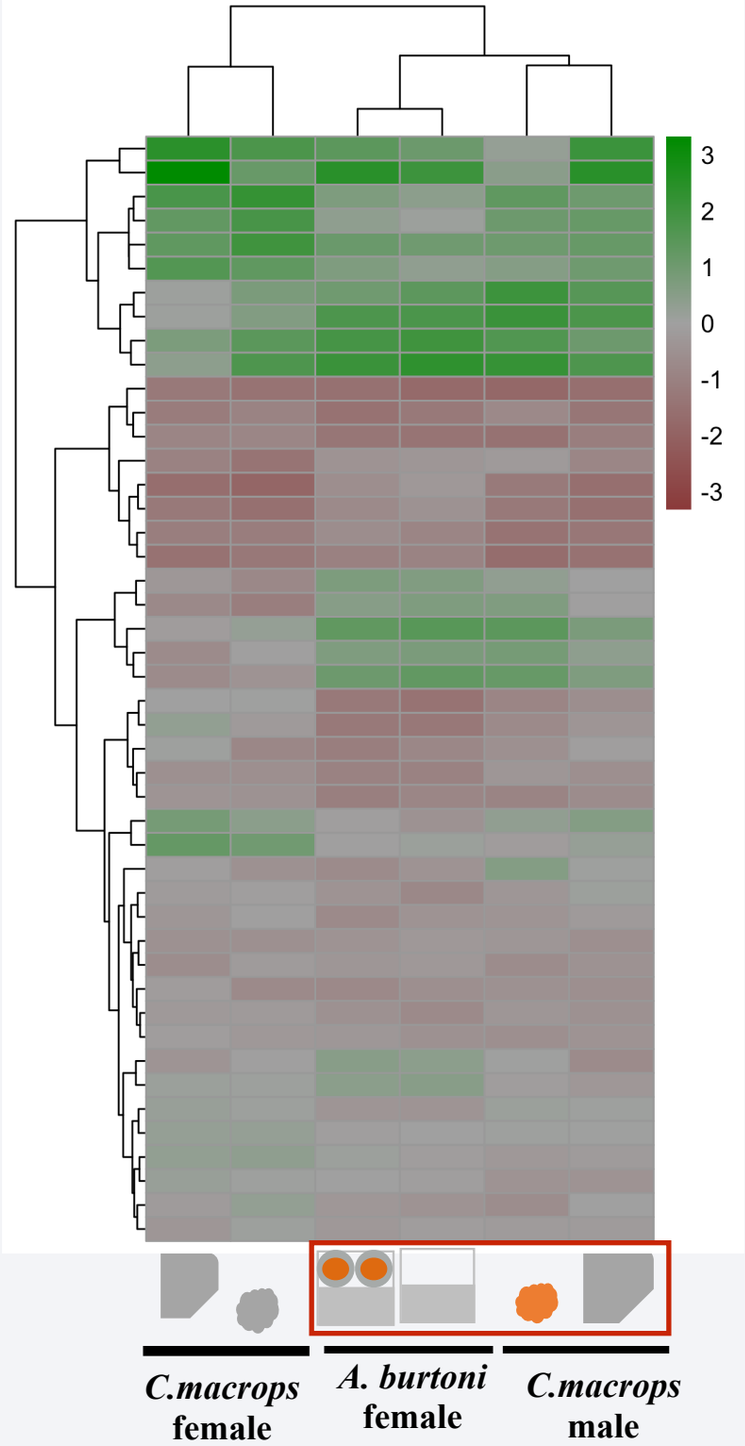
Up-regulated



Down-regulated

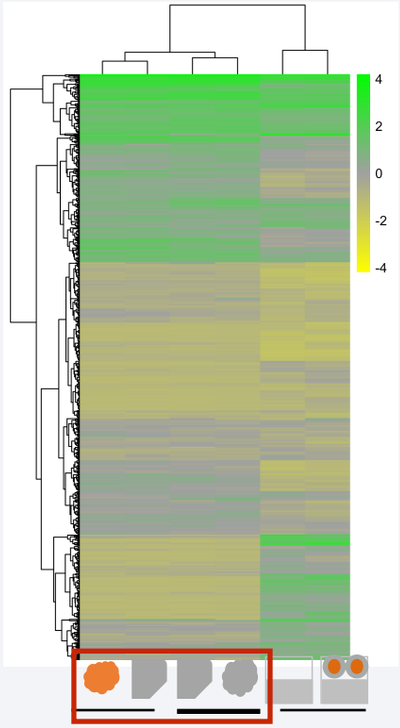


B common shared DE transcripts

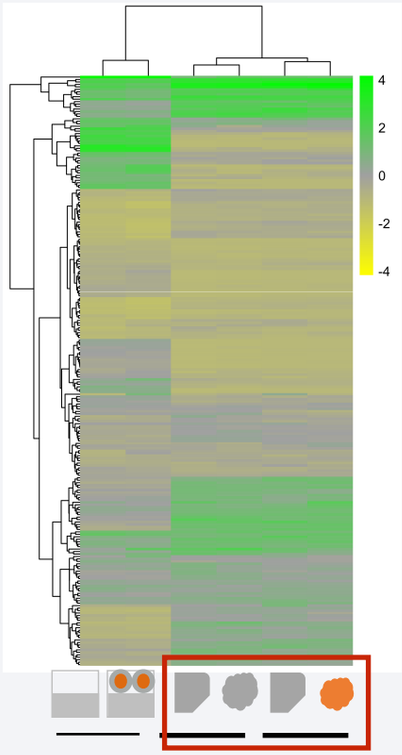


C egg-spots specific DE transcripts

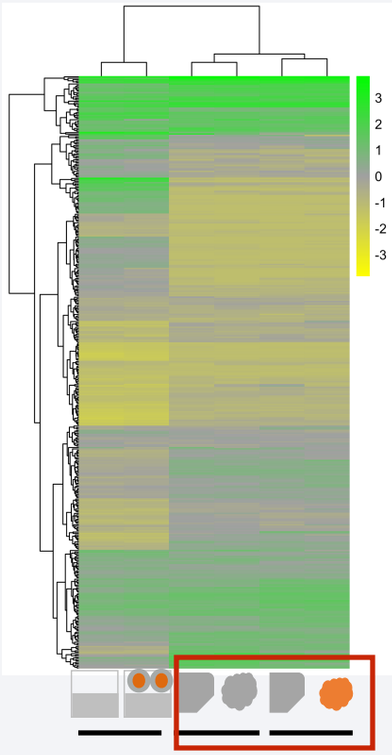
total



Up-regulated



Down-regulated

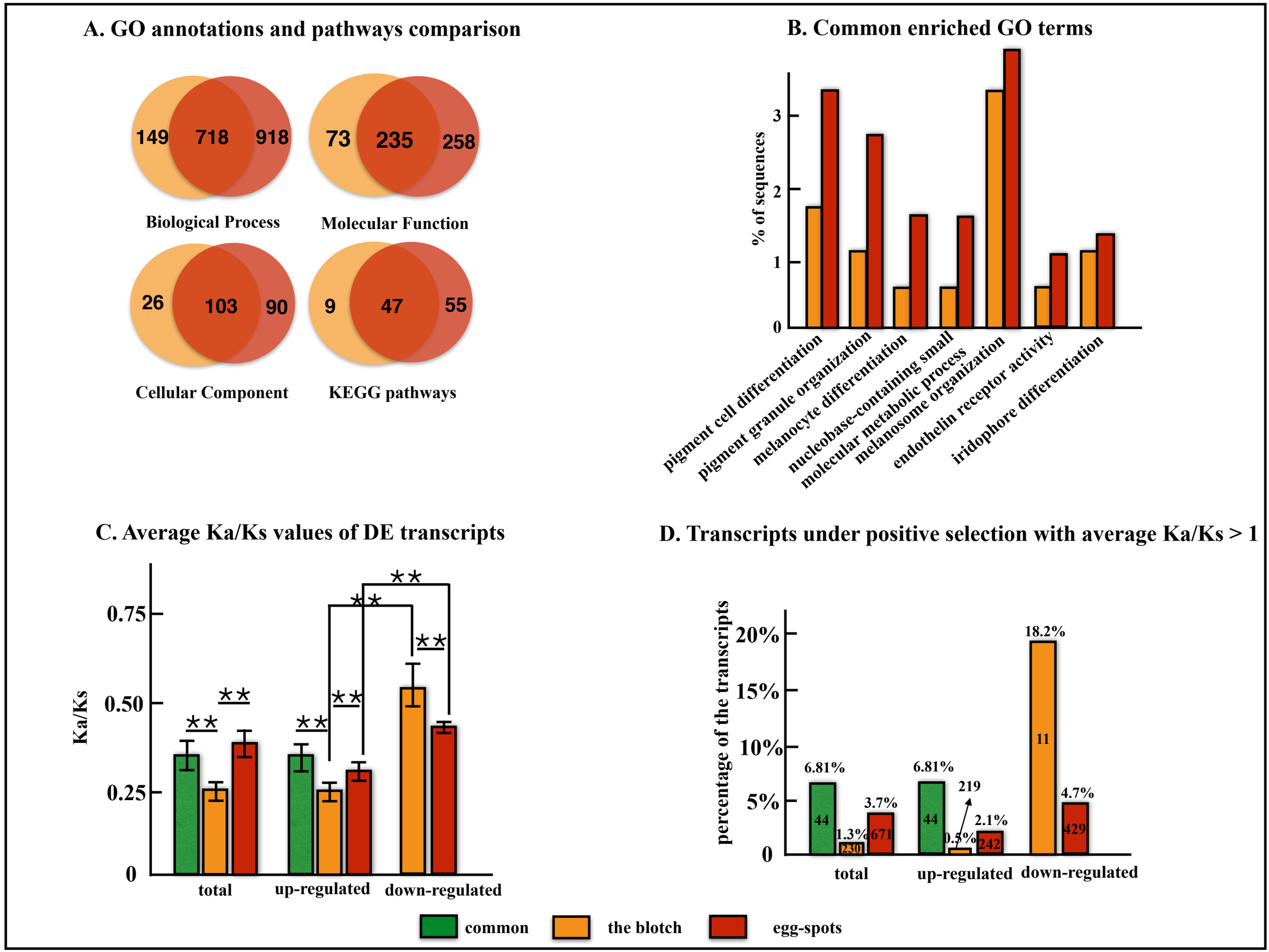


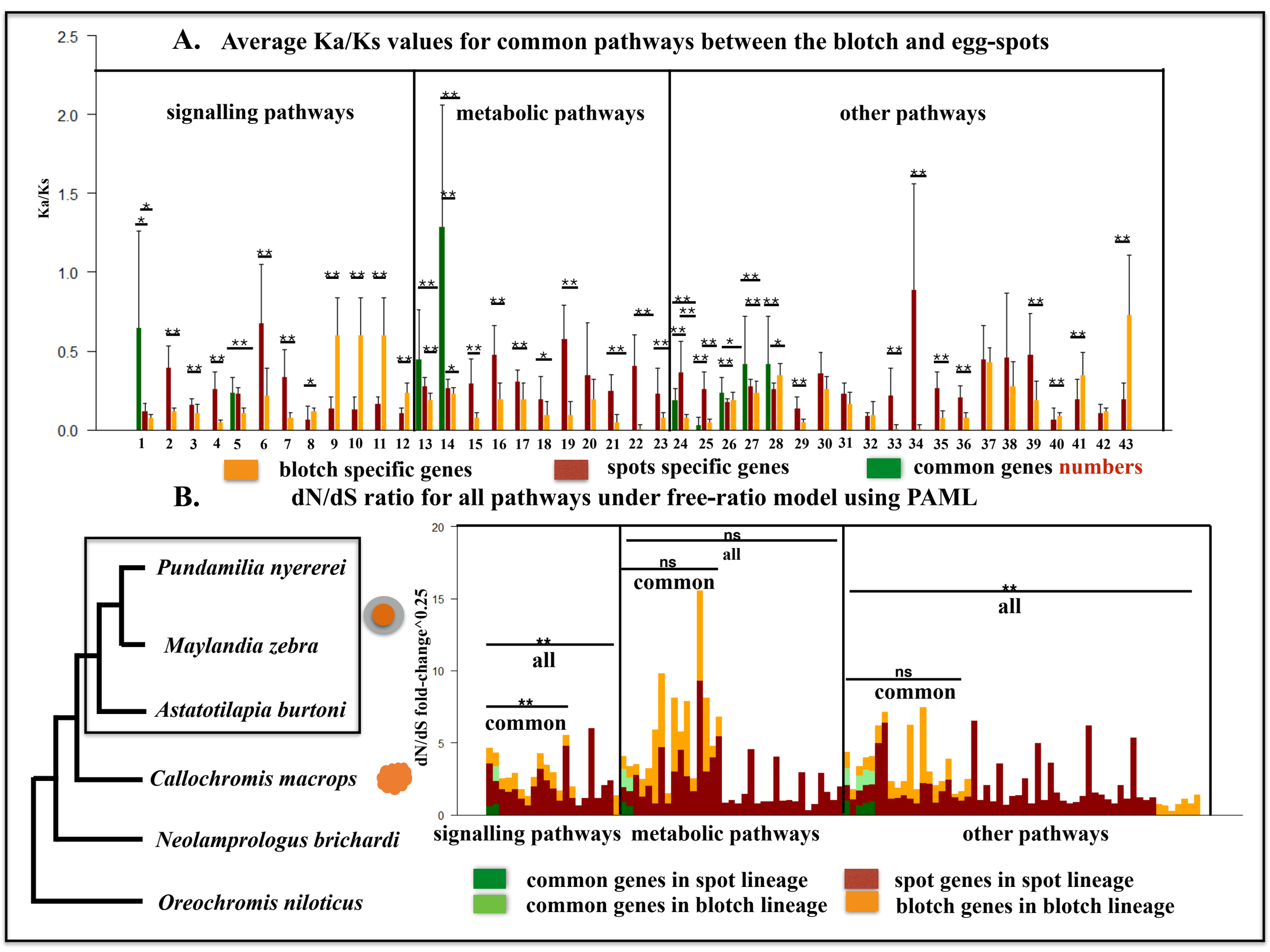
the blotch in male | anal fin down tissue in female | anal fin up tissue in female/non-blotch tissue in male

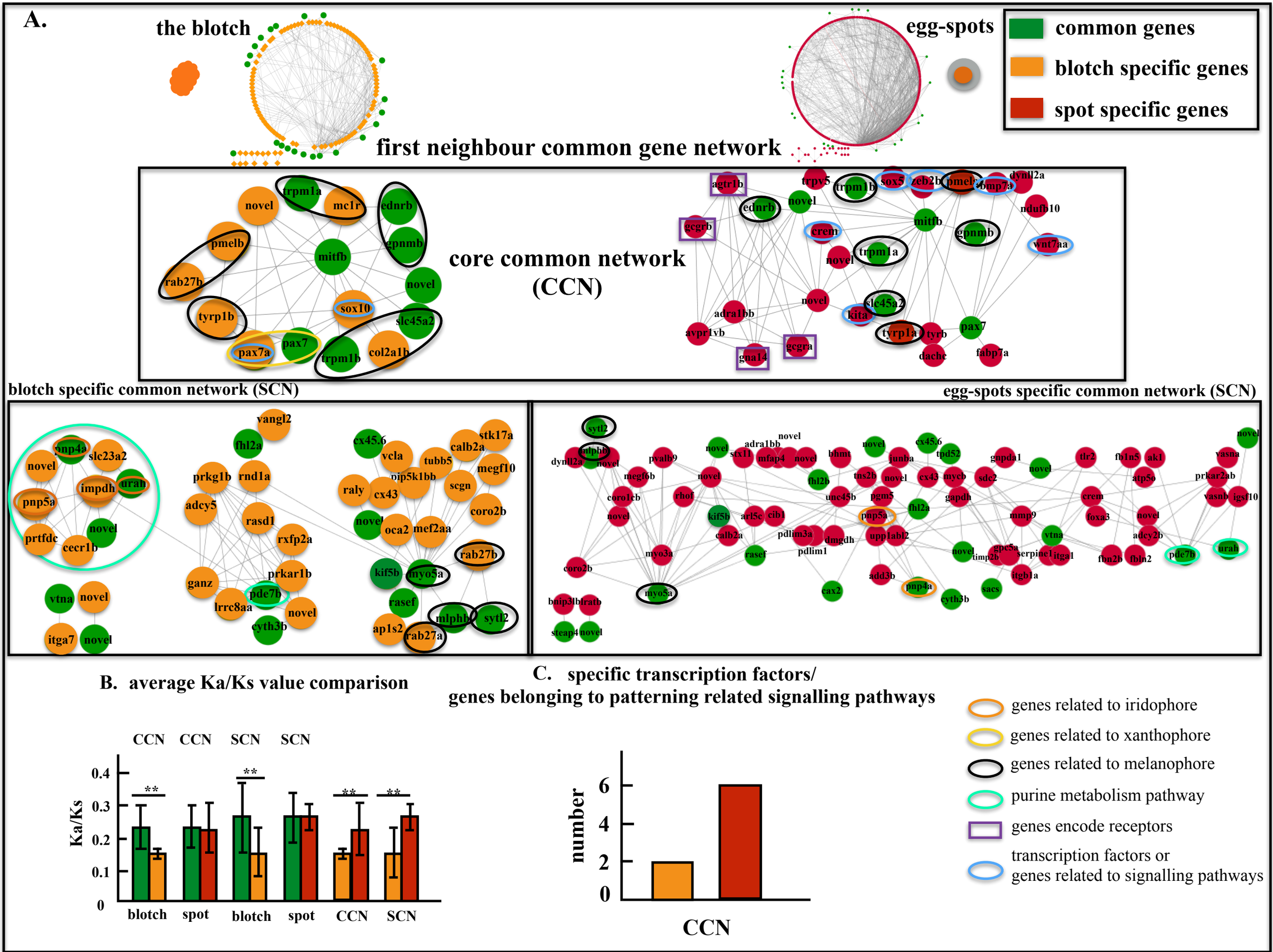
Callochromis macrops

egg-spots in male | non egg-spots in male

Astatotilapia burtoni


A. GO annotations and pathways comparison
149
718
918
Biological Process
73
235
258
Molecular Function
26
103
90
Cellular Component
9
47
55
KEGG pathways
B. Common enriched GO terms
% of sequences
0
1
2
3
pigment cell differentiation
pigment granule organization
melanocyte differentiation
nucleobase-containing small molecular metabolic process
melanosome organization
endothelin receptor activity
iridophore differentiation
C. Average Ka/Ks values of DE transcripts
Ka/Ks
0
0.25
0.50
0.75
**
total
up-regulated
down-regulated
D. Transcripts under positive selection with average Ka/Ks > 1
percentage of the transcripts
0
5%
10%
15%
20%
6.81%
44
1.3%
230
3.7%
671
6.81%
44
0.5%
219
2.1%
242
18.2%
11
4.7%
429
total
up-regulated
down-regulated
common
the blotch
egg-spots


A. Average Ka/Ks values for common pathways between the blotch and egg-spots
Ka/Ks
signalling pathways
metabolic pathways
other pathways
blotch specific genes
spots specific genes
common genes numbers
B. dN/dS ratio for all pathways under free-ratio model using PAML
Pundamilia nyererei
Maylandia zebra
Astatotilapia burtoni
Callochromis macrops
Neolamprologus brichardi
Oreochromis niloticus
dN/dS fold-change^0.25
all
common
ns
signalling pathways
metabolic pathways
other pathways
common genes in spot lineage
common genes in blotch lineage
spot genes in spot lineage
blotch genes in blotch lineage


A.
the blotch
egg-spots
common genes
blotch specific genes
spot specific genes
first neighbour common gene network
core common network
(CCN)
blotch specific common network (SCN)
egg-spots specific common network (SCN)
B. average Ka/Ks value comparison
Ka/Ks
CCN CCN SCN SCN
blotch spot blotch spot CCN SCN
C. specific transcription factors/
genes belonging to patterning related signalling pathways
number
CCN
genes related to iridophore
genes related to xanthophore
genes related to melanophore
purine metabolism pathway
genes encode receptors
transcription factors or
genes related to signalling pathways

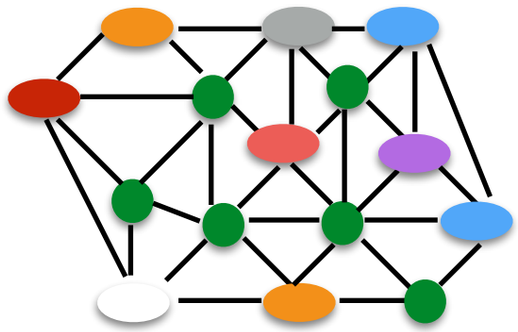
Hypothesis

Genetic basis of the origin and evolution of convergent evolution of innovative anal fin pigmentation patterns in cichlid fishes

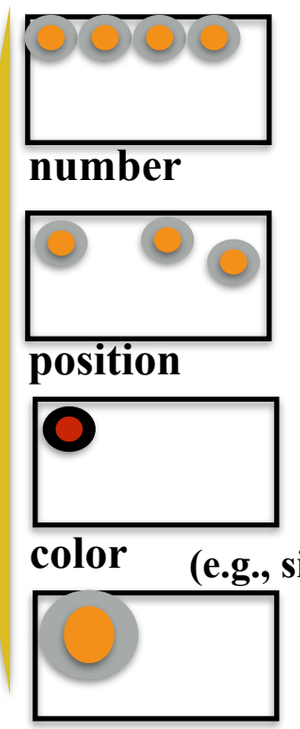
independent connections of core gene network to transcriptions factors and signalling pathways

intrinsic factor

developmental constraint



diversified egg-spots



egg-spots

haplochromine lineage

most-species richness

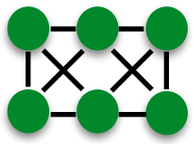
intrinsic factor

developmental constraint

deep homology

and/or

independently evolved



sexual selection

accelerated evolutionary rate

advantageous genes

extrinsic factor

selection

similar effects on pathways and cascade changes (common gene network) inspired the emergence of the convergent novelties

ectodine lineage

the blotch

integrate advantageous genes

higher evolutionary rates

(e.g., signalling pathways and egg-spots-specific pathways)

extrinsic factor

selection

provide a clue

egg-spots——key innovation

adaptive radiation

relatively independent egg-spots gene network

and relatively homogeneous anal fin background

intrinsic factor

developmental constraint

key innovation—a novelty out of novelty

common genes/pathways

transcription factors /signalling pathways