Genetic basis of innovative anal fin pigmentation patterns in cichlid fish

Inauguraldissertation

Zur

Erlangung der Würde eines Doktors der Philosophie

Vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Langyu Gu

von China

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von

Prof. Dr. Walter Salzburger, Prof. Dr. Alistair McGregor

Basel, **June, 21st, 2016**

Prof. Dr. Jörg Schibler
The Dean of Faculty

Table of Contents

Abstract

The origination of novelty is one of the most fascinating questions in evolutionary biology. The repeated evolution of innovative pigmentation patterns on the anal fin in East African cichlid fish is an ideal model to study this question. One pattern is eggspots, the circular pigmentation pattern with a transparent outer ring that emerged once in the most species rich cichlid lineage, the haplochromines, exhibiting large varieties with different numbers, sizes and colours. Eggspots have been suggested to be involved in female attraction, male-male competition and species recognition. While ancestral haplochromine species feature another fin pigment trait in form of blotch, which is reddish with ill-defined boundary. Anal fin pigmentation pattern was also independently evolved in the ectodine lineage, which possesses similar blotch pattern as the haplochromine blotch. The ectodine blotch pattern was also suggested to be involved in female attraction, although less investigated. Unlike haplochromine eggspots, the ectodine blotch shows almost no variation among species. Here, by applying next generation sequencing technology (RNAseq and Ion Torrent sequencing) followed by a series data analysis, we found that haplochromine eggspots and the ectodine blotch share at least parts of a common gene network. Further sequencing data showed that many of the anal fin pigmentation related candidate genes have eggspots specific segregating patterns. While species with the blotch showed similar sequence patterns with species without anal fin pigmentation patterns. This might suggest that eggspots, but not the ectodine blotch, might have a much more independent gene network, which might explain its higher evolvability. Besides, we also described the evolutionary history of apolipoprotein D (ApoD) gene family in teleosts, whose expansion is via gene duplication and are located in two clusters in teleost fish. One member of this gene family was found to be highly expressed in the ectodine blotch. Interestingly, although most genes showed conserved homologous expression pattern in distant related teleosts, duplicated genes with new functions evolved in a lineage specific manner, especially in cichlid fish, and were expressed in two novelties, lower pharyngeal jaw and anal fin pigmentation. By investigating the genetic basis of the innovative anal fin pigmentation patterns in cichlid fish, this doctoral work gives clues about the relationship among evo-devo, novelty and biological diversity.

Chapter 1
Introduction

Chapter 1
Introduction

The origination of novelty is one of the most fascinating questions in evolutionary biology; especially since it is challenging in the light of Darwin's theory with its natural selection centralism (Darwin 1859; Pigliucci and Müller 2010; Laland et al. 2015). Evolutionary novelties, defined as "*is a structure that neither homologous to any structure in the ancestral species nor serially homologous to any part of the same organism"* (Müller and Wagner 1991) provide the raw material for downstream selection and adaptation. That way, evolutionary novelties may contribute to biodiversity, as examplified by the beak of birds (Bhullar et al. 2015; Bright et al. 2016), eyespots in the family nymphalidae in butterfly (Monteiro 2015), or the neural crest in vertebrates (Green et al. 2015). In spite of being a hot topic for a long period in evolutionary biology, several basic questions regarding the origination of novelty are still unclear (Wagner 2014). For example, what is the mechanism of the origination and evolution of morphological novelty? Why do some novelties result in much higher evolvability, while others are less variable? Recent re-burning of evolutionary developmental biology (evo-devo) theory combining with next generation sequencing technology provide an unprecedented opportunity to answer these questions (Lynch et al. 2011; Wagner 2012; Roux et al. 2015). With this background, therefore, in my PhD thesis, I addressed questions about the origination and evolution of novelty by focusing on the genetic basis of two convergent innovative anal fin pigmentation patterns in East African cichlid fish, haplochromine eggspots and ectodine botches in East African cichlid fish (Fig.1).

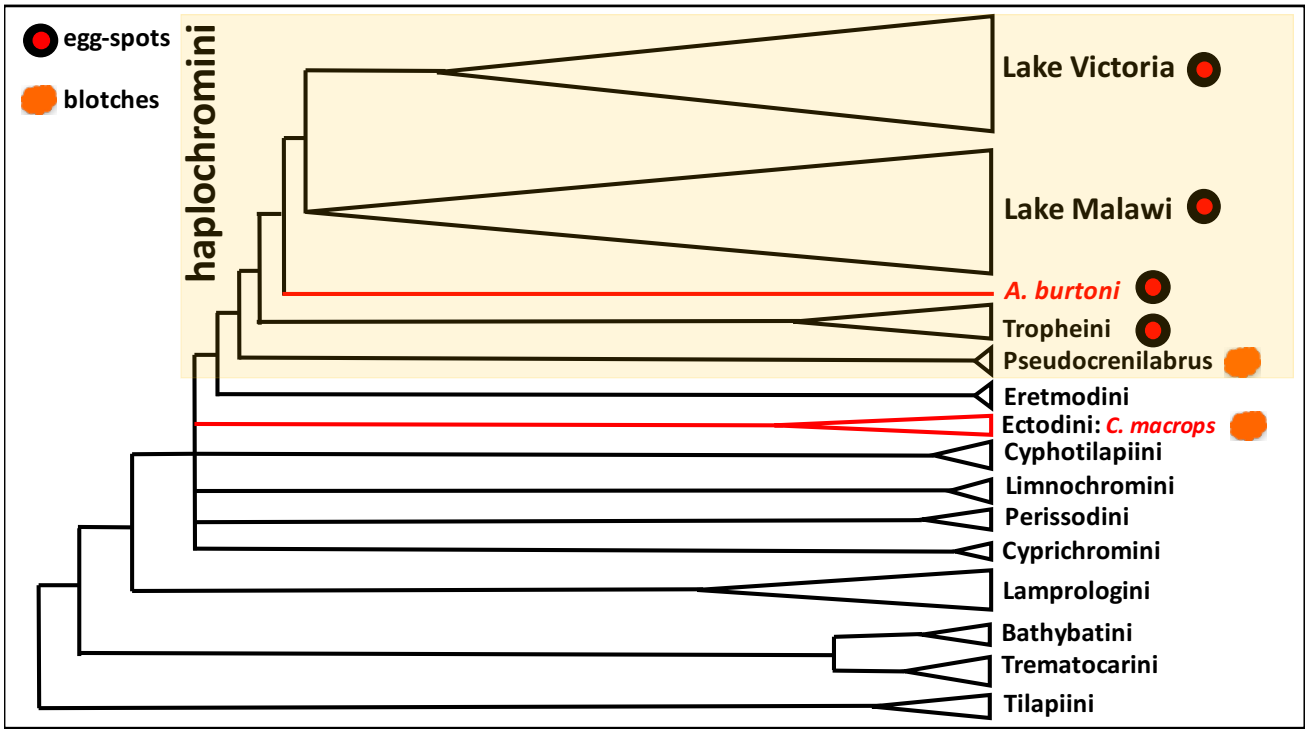

Fig. 1 Convergent evolution of anal fin pigmentation patterns in East African cichlid fish. Schematic molecular phylogeny of the East African cichlid fishes based on combined evidence from Salzburger et al., 2005; Meyer et al., 2015; Takahashi and Sota, 2016 . Triangle symbol represented species richness based on studies from Salzburger et al., 2005 and https://en.wikipedia.org. Names on the right side indicated rivers or tribes.

What is a morphological novelty

Previously, novelty was defined as "*is a structure that neither homologous to any structure in the ancestral species nor serially homologous to any part of the same organism*" (Müller and Wagner 1991). Therefore, understanding the concept of homology is the pre-request to understand what is novelty. However, the definition of homology itself is inconclusive (Wake 1994). For example, inference of homology in digit identity in birds and skinks is based on conflicting evidences from character anatomy, phylogenetic distribution and embryological position (Wagner 2005). Based on developmental experimental data, conserved developmental regulatory genes underlying the maintenance of character identity were found across taxa, such as the Hox gene clusters (Carroll 1995). However, most of times, homologous characters could also exhibit a huge diversity across taxa. It seems that many characters themselves have some sort of modularity, and the corresponding gene-gene interactions can be hierarchically structured for the phenotype with the core gene network being conserved among homologous characters (Wagner et al. 2007; Pigliucci and Müller 2010). Wagner (Wagner 2007) proposed that conserved Character Identity Networks (CHINs) maintained by transcription factors (TFs) is the basis of character homology, while the activation of the expression of CHINs, or the effectors that ultimately express CHINs can be flexible, which can explain the divergence of homologous characters. For example, epistatic

interactions between TFs *FoxD3, SoxE, Snai1/2* and *Pax3/7* constitute a very conserved module of neural crest specification (Green et al. 2015); or 5'-HoxD, which plays a ubiquitous role in digit development (Andrey et al. 2013). With this background in mind, novelty is the evolution of a quasi-independent CHINs to integrate signals into a gene expression pattern unique to that organ, so that it can obtain individuality from ancestral character (Wagner 2014). In this case, the study of evolutionary novelties is to explain the origin of the core gene regulatory network, which executes organ-specific gene expression patterns. Whether these networks are modifications from ancestral gene regulatory networks, or are assembled *de novo* is an open question.

Why do we study morphological novelty

The origination of novelty is one of the most challenging question of Darwin's theory, which is centered around natural selection (Darwin 1859; Pigliucci and Müller 2010; Laland et al. 2015). However, in the case of the evolution of novel traits, natural selection may only be the downstream force and follows the evolution of "raw phenotypic materials", while the intrinsic developmental factors themselves should not be ignored. The recently growing field of evolutionary development biology (evo-devo) makes a contribution to understand how development influences evolution by shaping the phenotypes (Müller 2007). For example, a recent study has found that the breaking of the developmental constraints between beak and braincase might be responsible for the adaptive radiation of passerine birds (Bright et al. 2016). Therefore, studying novelty provides a way to disentangle the relationships between developmental constraints (intrinsic factor) and natural selection (extrinsic factor) in speciation and adaptation. Besides, previously, biology mainly focused on individual genes and single gene-to-phenotype interactions. However, as mentioned above, the phenotype is influenced by gene-gene interactions and how are the gene network rewired are the center to study the evolution of phenotypic novelty. Therefore, by studying novelty, we can focus on the gene network level instead of individual genes (Lynch et al. 2011). Finally, unlike population genetics, which mainly focuse on the description of the results of population dynamics, such as gene flow, reproductive isolation, etc. (Fisher 1930; Wright 1931; Haldane 1932), studying of the emergence of novelty can help find the causation for driving the dynamics of speciation and adaptation. Therefore, it is time to shift from focusing on natural selection centralism to intrinsic development inputs (evo-devo), from individual genes to gene regulatory networks and from explanation to causation (Pigliucci and Müller 2010; Laland et al. 2015). Therefore, studying morphological novelties provides good models to extend the existing biological theory.

Existing mechanisms of the evolution of morphological novelty

How the phenotypes are patterned from identical DNA code depends on different gene activities governed by gene regulatory networks. Hence, to understand the origin of morphological novelty, it is necessary to study how the corresponding gene networks originate and are wired. Many studies have shown that the co-option of existing networks plays a role in the evolution of novel morphological structures (Shubin et al. 2009; Glassford et al. 2015). One efficient way is by modifying the existing *cis*-regulatory element to affect target gene expression, which might explain the repeated evolution of similar morphological patterns, such as wing pigmentation patterns in flies (Gompel et al. 2005). However, it is questionable whether only small-scale changes can explain the origin of complex novelties if they are involved in the recruitment of hundreds of genes (Wagner and Lynch 2010). Besides, evolutionary innovations often require rapid changes of the regulatory network. In this case, transposable element (TE) insertions, which can bring ready-to-use TF binding sites, might rewire regulatory links efficiently and contribute to the establishment of lineage-specific novelties (Warren et al. 2015). For example, TF insertion in the upstream region of a duplicated sex-determine gene, *dmrt1bY*, provided binding sites to the paralogous *dmrt1* to establish new hierarchy which enables the self- and cross-regulation (Herpin et al. 2010); or the evolution of pregnancy in mammals is involved in the transposon-mediated gene regulatory networks rewiring (Lynch et al. 2011).

In addition, protein changes can be also important for the origination of novelty, especially for mutations in binding domain of TFs, such as their role in changing metazoan developmental gene regulatory networks (Cheatle Jarvela and Hinman 2015). By the co-option with other TFs, non-coding RNA or *cis*-regulatory elements, TFs complex might affect the target gene expression. In this case, amino acid substitutions in the TF within the complex might affect this interactions. For example, the highly conserved HOX motifs are differentially required in *Drosophila* for the reduction in the number of sex combs (NEAGS motif) and the reduction of salivary gland nuclei (DYTQL motif) (Sivanantharajah and Percival-Smith 2015). By new protein-protein interactions, new protein domain origination may have a major role in the evolution of direct protein networks and innovation, such as different motif gaining or lost of TF *Ftz* is segregating with different phenotypes gain or lost across insect orders (Sivanantharajah and Percival-Smith 2015).

Except the mechanisms of "old genes playing new tricks" by co-option of pre-existing genes that mentioned above, it remains largely unknown whether genetic novelties (new genes) can contribute to phenotypic novelties directly (but see (Zhang and Long 2014)). It has been suggested

that duplicated genes might be correlated with the origin of evolutionary novelties (Pougach et al. 2014; Ogino et al. 2016; Soltis and Soltis 2016). For example, duplication of *dachshund (dac)* in arachnids (spiders and allies) is linked with the evolution of a novel leg segement, the patella (Turetzek et al. 2016). A teleost-specific extracellular matrix gene, *elastin b*, was acquired leading to new functions contributing to the bulbs arteriosus, an organ of evolutionary novelty seen in the teleost (Moriyama et al. 2016). And a specific but interesting example has shown that both the TF and its targets were duplicated, followed by two point mutations in the promoter regions of the target genes, combined with two single-nucleotide mutations in the DNA-binding domain of the TF paralogs are sufficient to promote the emergence of two independent regulatory circuits, and the emergence of a new regulatory module on the regulation of the *MAL* genes in yeast (Pougach et al. 2014). However, since the chance of beneficial mutation is relatively low compare to deleterious mutation or neutral mutation, the examples about new genes responsible for novelties are still scarce.

An ideal system to study morphological novelty—anal fin pigmentation patterns in cichlid fish

Evo-devo is to study how evolution shapes development (the developmental toolkit) to explain how development influences evolution (developmental constraints) (Pantalacci and Sémon 2015). In this case, it means that evo-devo should be studied in a phylogenetic point of view. Besides, it is better to compare gene network to find the differences of CHINs among closely related species to minimize genetic background and historical noises. In this case, the origin of evolutionary innovative pigment patterns on the anal fin in East African cichlid fish is an ideal model, considering that this system evolved relatively recently (Salzburger et al. 2005; Santos et al. 2014; Meyer et al. 2015), so that the noise of individual evolutionary history and stochasticity effects can be reduced (Faria et al. 2014); and the anal fin pigmentation patterns have evolved repeatedly in several East African cichlid lineages but with different evolvability. Perhaps the most famous examples are the so-called ‘eggspots’ on the anal fins of the haplochromines, the most species-rich lineage of cichlids in East Africa (Salzburger 2009; Santos et al. 2014). Eggspots are circular pigmentation patterns with a transparent outer ring (Santos et al. 2014) (Fig. 1). Diversified phenotypes are found in haplochromine eggspots, with different colours, sizes and numbers in different species. Eggspots was suggested to be involved in female attraction (Wickler 1962), male-male competition (Theis et al. 2012; Theis et al. 2015) and species recognition (Couldridge 2002). While ancestral haplochromine species feature another fin pigmentation trait in form of blotches, which is reddish with ill-defined boundary (Fig. 1). Anal fin pigmentation patterns have independently evolved in another lineage, ectodines; these are similar to blotches in the ancestral haplochromies (Fig. 1). Ectodine blotches were also suggested to be related to female attraction

(Fryer and Iles 1972), although these are less investigated. Noticeably, unlike haplochromine eggspots, the blotches of different species of *Callochromis* show almost no variation among species.

Several studies have focused on the genetic basis of haplochromine eggspots. For example, Santos et al. (2014) have shown that a *cis*-regulatory change in the form of a TE insertion in the upstream region of the gene *fhl2b* appears to be causally related to the eggspots phenotype. Experiments with transgenic zebrafish suggested that this *cis*-regulatory change drives expression in iridophore (Santos et al. 2014). However, since the *trans*-landscape might be different between zebrafish and cichlid fish, it is still unclear how *fhl2b* affects the emergence of eggspots and whether *cis*- or *trans*- of *fhl2b* is responsible for the formation of eggspots. Another study found that coding sequence change in a xanthophore related gene, *csf1ra,* is highly expressed in two independently evolved fin pigment patterns (Salzburger et al. 2007). The same study also identified positive selection in the ancestral lineage of haplochromine lineage, which might causally link haplochromine eggspots to the radiation of this lineage. Recently, Santos et al., (Santos et al.) suggested that these eggspots and blotches do not share a genetic basis. However, this suggestion was only based on gene expression profile of 46 out of 1229 eggspots candidate genes which might have bias. Therefore, a thorough comparison of gene expression profiles between eggspots and ectodine blotch is needed.

Aim of my PhD project

Therefore, in my PhD project, by applying next generation sequencing technology (RNAseq and Ion Torrent sequencing), I wanted to answer the following questions about the innovative anal fin pigmentation patterns in cichlid fish: 1) What is the genetic basis of the innovative ectodine blotch in *C. macrops*? 2) Do eggspots and ectodine blotch share a common gene network based on transcriptomic data? 3) What are the differences of the gene characters (including upstream and coding region) of the shared and unshared candidate genes between eggspots and blotch? 4) Why both eggspots and ectodine blotch are innovative anal fin pigmentation patterns but possess different evolvability?

Thesis outline

To disentangle the origination and evolution of the innovative anal fin pigmentation patterns in cichlid fish, first of all, I was involved in a project to generate transcriptomic data for eggspots in *A. burtoni,* followed by gene expression profile comparison of 46 highly differential expressed eggspots related candidate genes. This project suggested that eggspots and blotches might not share a common genetic basis. ***Chapter 2*** is the resulting submitted paper from this work. However,

considering that it was only based on 46 out of 1229 eggspots related candidate genes, the conclusion might have bias. Therefore, in ***Chapter 3***, I started my PhD project by making use of a thorough comparative transcriptomic data analysis to find ectodine blotch related candidate genes in *C. macrops* first, and then compare its expression profile with the eggspots candidate genes derived from Chapter 2. Then, in ***Chapter 4***, I wanted to find what are the mechanisms behind the gene network rewiring of eggspots and ectodine blotch by making use of Ion Torrent next generation sequencing, and especially focus on the roles of TEs and TFs. This result also provides clues about the different evolvability between eggspots and ectodine blotch. In ***Chapter 5***, I mainly described the expansion of apolipoprotein D gene family via gene duplication in cluster in teleosts. One of its members was found highly expressed in ectodine blotch in Chapter 3. Orthologous genes in the same physical order along their respective cluster exhibited homologous tissue expression pattern, indicating their conserved roles in the corresponding gene network. Interestingly, novel functions evolved in the lineage specific duplicated genes in cichlid fish. The expansion of this gene family might be related to speciation and radiation of teleosts, especially for cichlid fish. Finally, in ***Chapter 6***, I discussed the results obtained throughout my doctoral work, along with future perspectives.

References

Andrey G, Montavon T, Mascrez B, Gonzalez F, Noordermeer D, Leleu M, Trono D, Spitz F, Duboule D. 2013. A switch between topological domains underlies HoxD genes collinearity in mouse limbs. Science. 340:1234167.

Bhullar B-AS, Morris ZS, Sefton EM, Tok A, Tokita M, Namkoong B, Camacho J, Burnham DA, Abzhanov A. 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. 69:1665–1677.

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.

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. Nature. 376:479–485.

Cheatle Jarvela AM, Hinman VF. 2015. Evolution of transcription factor function as a mechanism for changing metazoan developmental gene regulatory networks. Evodevo. 6:3.

Couldridge VCK. 2002. Experimental manipulation of male eggspots demonstrates female preference for one large spot in Pseudotvopheus lombavdoi. J. Fish Biol. 60:726–730.

Darwin C. 1859. The origin of species. John Murray.

Faria R, Renaut S, Galindo J, Pinho C, Melo-Ferreira J, Melo M, Jones F, Salzburger W, Schluter D, Butlin R. 2014. Advances in Ecological Speciation: an integrative approach. Mol. Ecol. 23:513–521.

Fisher R. 1930. The Genetical Theory of Natural Selection. Clarendon Press.

Fryer G, Iles T. 1972. The cichlid fihses: their biology and evolution. edinburgh, uk.

Glassford WJ, Johnson WC, Dall NR, Smith SJ, Liu Y, Boll W, Noll M, Rebeiz M. 2015. Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty. Dev. Cell. 34:520–531.

Gompel N, Prud'homme B, Wittkopp PJ, Kassner VA, Carroll SB. 2005. Chance caught on the

wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila. Nature. 433:481–487.
Green SA, Simoes-Costa M, Bronner ME. 2015. Evolution of vertebrates as viewed from the crest. Nature. 520:474–482.
Haldane JBS. 1932. The causes of evolution. London: Longmans, Green.
Herpin A, Braasch I, Kraeussling M, Schmidt C, Thoma EC, Nakamura S, Tanaka M, Schartl M. 2010. Transcriptional rewiring of the sex determining dmrt1 gene duplicate by transposable elements. PLoS Genet. 6:e1000844.
Laland KN, Uller T, Feldman MW, Sterelny K, Müller GB, Moczek A, Jablonka E, Odling-Smee J. 2015. The extended evolutionary synthesis: its structure, assumptions and predictions. Proc. Biol. Sci. 282:20151019.
Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 43:1154–1159.
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.
Monteiro A. 2015. Origin, development, and evolution of butterfly eyespots. Annu. Rev. Entomol.60:253–271.
Moriyama Y, Ito F, Takeda H, Yano T, Okabe M, Kuraku S, Keeley FW, Koshiba-Takeuchi K. 2016. Evolution of the fish heart by sub/neofunctionalization of an elastin gene. Nat. Commun. 7:10397.
Müller GB, Wagner GP. 1991. Novelty in Evolution: Restructuring the Concept. Annu. Rev. Ecol. Syst. 22:229–256.
Müller GB. 2007. Evo-devo: extending the evolutionary synthesis. Nat. Rev. Genet. 8:943–949.
Ogino Y, Kuraku S, Ishibashi H, Miyakawa H, Sumiya E, Miyagawa S, Matsubara H, Yamada G, Baker ME, Iguchi T. 2016. Neofunctionalization of Androgen Receptor by Gain-of-Function Mutations in Teleost Fish Lineage. Mol. Biol. Evol.33:228–244.
Pantalacci S, Sémon M. 2015. Transcriptomics of developing embryos and organs: A raising tool for evo-devo. J. Exp. Zool. B. Mol. Dev. Evol. 324:363–371.
Pigliucci M, Müller GB. 2010. Evolution: The Extended Synthesis. MIT Press.
Pougach K, Voet A, Kondrashov FA, Voordeckers K, Christiaens JF, Baying B, Benes V, Sakai R, Aerts J, Zhu B, et al. 2014. Duplication of a promiscuous transcription factor drives the emergence of a new regulatory network. Nat. Commun. 5:4868.
Roux J, Rosikiewicz M, Robinson-Rechavi M. 2015. What to compare and how: Comparative transcriptomics for Evo-Devo. J. Exp. Zool. B. Mol. Dev. Evol. 324:372–382.
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.
Salzburger W, Mack T, Verheyen E, Meyer A. 2005. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 5:17.
Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol.18:169–185.
Santos ME, Baldo L, Gu L, Boileau N, Musilova Z, Salzburger W. The molecular basis of independently evolved anal fin pigment patterns in cichlid fish (submitted).
Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, Belting H-G, Affolter M, Salzburger W. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:5149.
Shubin N, Tabin C, Carroll S. 2009. Deep homology and the origins of evolutionary novelty. Nature. 457:818–823.
Sivanantharajah L, Percival-Smith A. 2015. Differential pleiotropy and HOX functional organization. Dev. Biol. 398:1–10.

Soltis PS, Soltis DE. 2016. Ancient WGD events as drivers of key innovations in angiosperms. Curr. Opin. Plant Biol.30:159–165.
Takahashi T, Sota T. 2016. A robust phylogeny among major lineages of the East African cichlids. Mol. Phylogenet. Evol. 100: 234-42.
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.
Theis A, Salzburger W, Egger B. 2012. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. PLoS One. 7:e29878.
Turetzek N, Pechmann M, Schomburg C, Schneider J, Prpic N-M. 2016. Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. Mol. Biol. Evol.33:109–121.
Wagner GP, Lynch VJ. 2010. Evolutionary novelties. Curr. Biol. 20:R48–R52.
Wagner GP, Pavlicev M, Cheverud JM. 2007. The road to modularity. Nat. Rev. Genet. 8:921–931.
Wagner GP. 2005. The developmental evolution of avian digit homology: an update. Theory Biosci. 124:165–183.
Wagner GP. 2007. The developmental genetics of homology. Nat. Rev. Genet. 8:473–479.
Wagner GP. 2012. Next Gen Devo-Evo. J. Exp. Zool. B. Mol. Dev. Evol. 318:519–520.
Wagner GP. 2014. Homology, genes and evolutionary innovation. Princet. Univ. Press. Princet.
Wake DB. 1994. Comparative terminology. Science. 265:268–269.
Warren IA, Naville M, Chalopin D, Levin P, Berger CS, Galiana D, Volff J-N. 2015. Evolutionary impact of transposable elements on genomic diversity and lineage-specific innovation in vertebrates. Chromosome Res. 23:505–531.
Wickler W. 1962. “Egg-dummies” as Natural Releasers in Mouth-breeding Cichlids. Nature. 194:1092–1093.
Wright S. 1931. EVOLUTION IN MENDELIAN POPULATIONS. Genetics. 16:97–159.
Zhang YE, Long M. 2014. New genes contribute to genetic and phenotypic novelties in human evolution. Curr. Opin. Genet. Dev. 29:90–96.

Chapter 2

Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes

M. Emília Santos[1,2*], Laura Baldo[3], Langyu Gu[1], Nicolas Boileau[1], Zuzana Musilova[1,4] and Walter Salzburger[1*]

1: *Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland,*

2: *Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure, CNRS UMR 5242, 46 Allée d'Italie, 69364 Lyon Cedex 07, France*

3: *Ecology Department, University of Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain*

4: *Department of Zoology, Faculty of Science, Charles University in Prague, Vinicna 7, 128 44 Prague, Czech Republic*

*Correspondence:

M. Emília Santos, e-mail: emilia.p.santos@gmail.com

Walter Salzburger, e-mail: walter.salzburger@unibas.ch

Accepted by BMC Genomics, 2016, 17:712.

Santos et al. BMC Genomics (2016) 17:712
DOI 10.1186/s12864-016-3046-y

BMC Genomics

RESEARCH ARTICLE Open Access

Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes

M. Emília Santos[1,2*], Laura Baldo[3], Langyu Gu[1], Nicolas Boileau[1], Zuzana Musilova[1,4] and Walter Salzburger[1*]

Abstract

Background: Understanding the genetic basis of novel traits is a central topic in evolutionary biology. Two novel pigmentation phenotypes, egg-spots and blotches, emerged during the rapid diversification of East African cichlid fishes. Egg-spots are circular pigmentation markings on the anal fins of hundreds of derived haplochromine cichlids species, whereas blotches are patches of conspicuous anal fin pigmentation with ill-defined boundaries that occur in few species that belong to basal cichlid lineages. Both traits play an important role in the breeding behavior of this group of fishes. Knowledge about the origin, homology and underlying genetics of these pigmentation traits is sparse.

Results: Here, we present a comparative transcriptomic and differential gene expression analysis of egg-spots and blotches. We first conducted an RNA sequencing experiment where we compared egg-spot tissue with the remaining portion of egg-spot-free fin tissue using six individuals of *Astatotilapia burtoni*. We identified 1229 differentially expressed genes between the two tissue types. We then showed that rates of evolution of these genes are higher than average estimated on whole transcriptome data. Using quantitative real-time PCR, we found that 29 out of a subset of 46 differentially expressed genes showed an analogous expression pattern in another haplochromine species' egg-spots, *Cynotilapia pulpican*, strongly suggesting that these genes are involved in the egg-spot phenotype. Among these are the previously identified egg-spot gene *fhl2a*, two known patterning genes (*hoxC12a* and *bmp3*) as well as other pigmentation related genes such as *asip*. Finally, we analyzed the expression patterns of the same gene subset in two species that feature blotches instead of egg-spots, one haplochromine species (*Pseudocrenilabrus philander*) and one ectodine species (*Callochromis macrops*), revealing that the expression patterns in blotches and egg-spots are rather distinct.

Conclusions: We identified several candidate genes that will serve as an important and useful resource for future research on the emergence and diversification of cichlid fishes' egg-spots. Only a limited degree of conservation of gene expression patterns was detected between the egg-spots of the derived haplochromines and blotches from ancestral haplochromines, as well as between the two types of blotches, suggesting an independent origin of these traits.

Keywords: Pigmentation, Diversity, Egg-spot, Blotches, East African cichlids, Gene expression

Abbreviations: Bp, Base pairs; CDS, Coding sequences; CT, Threshold cycle; DE, Differentially expressed; dN/dS, Ratio of non-synonymous substitutions over synonymous substitutions; FDR, False discovery rate; GLM, Generalized linear model; GO, Gene ontology; qPCR, Quantitative real-time polymerase chain reaction; RNAseq, RNA sequencing; RSEM, RNA-Seq by Expectation-Maximization; UTR, Untranslated region

* Correspondence: emilia.p.santos@gmail.com; walter.salzburger@unibas.ch
[1]Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
Full list of author information is available at the end of the article

Background

Animal pigmentation patterns are highly variable phenotypes both at the intra- and inter-specific level, and represent prominent traits to study the genetics of species diversification and adaptation (reviewed in [1–3]). The functionality of color patterns can readily be assessed in most cases, given that these traits often evolve in response to adaptation to the environment via natural selection (e.g. inter- and intra-specific communication, camouflage and mimicry), or co-vary with female choice via sexual selection [4–6]. The outcome of these two types of selection regimes can be different, with the former often producing cryptic phenotypes, where coloration mimics the environment, while the latter generates conspicuous phenotypes, where males typically display bright colors driving female choice or male-male competition [4–6]. Despite the high evolutionary significance of color patterns, the genetic mechanisms underlying their formation and diversification often remain elusive [1–3].

Recent work in fish model systems, especially in zebrafish, has started to uncover the genes and cellular processes involved in pigmentation pattern formation [7–9]. Pigmentation patterns are determined by the specification of different types of neural crest derived pigment cells – the chromatophores [10] – that contain different light absorbing pigments: melanophores contain black eumelanin pigments; erythrophores and xantophores contain yellow-red carotenoid and pteridine pigments; cyanophores contain a blue pigment of unknown composition; and finally, iridophores contain purine crystals that produce metallic iridescence [11]. Differences in the arrangement, position, and density of these cells leads to the diversity of color patterns present in nature. These differences depend on a variety of factors including neural crest cell migration, specification, proliferation, and survival [7–9, 11].

In this study, we address the molecular basis of two novel and conspicuous pigmentation traits found in the anal fin of male cichlid fishes – egg-spots and blotches (Fig. 1). Egg-spots represent an evolutionary novelty that emerged only once in the haplochromine lineage, the most species-rich group of East African cichlids [12, 13]. These circular markings consist of a central circular area containing xanthophores and iridophores, surrounded by an outer transparent ring [14, 15]. They are primarily found in males and show an extreme inter- and intra-specific variability in number, color, and position on the fin [13–16]. Egg-spots have been the subject of intense studies suggesting a signaling function in the peculiar mating behavior of the mouth-brooding haplochromines. They are likely sexually selected via female choice in some species [17, 18] and via male-male competition in others [19–21]. Blotches, on the other hand, are patches of conspicuous anal fin pigmentation with ill-defined boundaries and occur only in a handful of cichlid species, including some basal haplochromines [13–15] and ectodine cichlids from Lake Tanganyika (Fig. 1). As with egg-spots, they are mostly found in males and their function might also be linked to courtship behavior, although this has been less extensively studied [12]. The origin and evolutionary trajectory of these anal fin patterns remains unclear. Due to the phylogenetic position of the species showing blotches as sister-group to the egg-spot bearing haplochromines [13–15], it might be speculated that egg-spots are derived from the blotch-pattern, which would make the two phenotypes homologous.

Convergent evolution is widespread in East African cichlid adaptive radiations, not only between lakes [22, 23], but also within a single lake [24]. For example, haplochromine anal fin blotches are phenotypically similar to the ones found in the genus *Callochromis* (Fig. 1). However, the phylogenetic position of *Callochromis*, which is nested within the Ectodini [25], suggests that these two types of blotches evolved independently. Overall, we envision two possible scenarios for the origin of egg-spots: in one case they represent a derived state of blotches found in haplochromines, whereas blotches found in ectodines evolved independently (two origins); alternatively egg-spots have evolved independently from the blotches of both basal haplochromines and ectodines (three origins).

Understanding the genetic pathways underlying these pigmentation phenotypes can help us to distinguish between these scenarios. While several studies have addressed pigmentation diversity in East African cichlids, little is known about the genetics underlying their

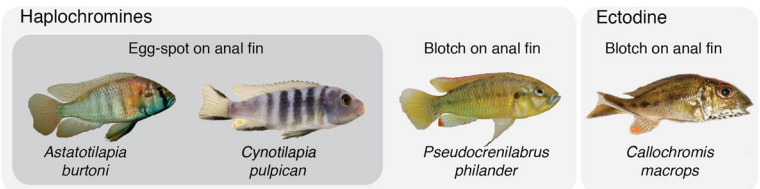

Fig. 1 Representative males from the four species analyzed: two haplochromine species displaying egg-spots in their anal fins (*A. burtoni* and *C. pulpican*), a basal haplochromine species (*P. philander*) and an ectodine species (*C. macrops*), both showing orange blotches in their anal fin

coloration and pigmentation patterning, and only a handful of genes have been studied in detail. Among these genes is *hagoromo*, which shows a greater diversity of alternatively spliced variants and accelerated protein evolution in the haplochromines compared to other cichlids [26, 27]; *paired box 7* (pax7), on the other hand, was shown to be linked to a haplochromine female biased pigmentation phenotype [28]. Three genes have so far been associated with the egg-spot phenotype: the xanthophore marker colony stimulating factor 1 receptor A (*csf1ra*), and the two *four and a half lim domain 2* proteins (*fhl2a* and *fhl2b*). *csf1ra* is expressed in haplochromine egg-spots and in the characteristic "Perl-fleckmuster" (pearly spotted) pattern present in cichlid fins. This gene underwent adaptive sequence evolution in the ancestral lineage of the haplochromines coinciding with the emergence of egg-spots [14]. However, *csf1ra* is downstream in the pathway of egg-spot morphogenesis. More recently, we have shown that *fhl2a* and *fhl2b* are more causally related to egg-spot development and that an alteration in the *cis*-regulatory region of *fhl2b* could have contributed to the emergence of this trait in haplochromines in the first place [15].

In this study, we first addressed the question of the genetic basis of the egg-spots. We then went onto use comparative transcriptomics across species carrying egg-spots and blotches to shed light on the origin of this novel trait. Specifically, we identified a total of 1229 genes that were differentially expressed (DE) between egg-spot and non-egg-spot fin tissues in the haplochromine cichlid *Astatotilapia burtoni*. These genes are evolving at a higher rate than average making this a valuable dataset to study the emergence and rapid diversification of this trait. For a subset of 46 DE genes we measured expression levels in three other species: the egg-spot bearing haplochromine *Cynotilapia pulpican*, carrying egg-spots on a different region of the anal fin than *A. burtoni*, and two blotch-bearing species, the basal haplochromine *Pseudocrenilabrus philander* and the ectodine *Callochromis macrops*. The rationale is that if egg-spots and blotches in haplochromines are controlled by the same genetic components they might show similar expression profiles.

A total of 29 out of 46 genes were found to be DE in *C. pulpican*. By comparing the expression in two haplochromine species with different egg-spot arrangements, we confirmed that the expression of the genes is correlated with the presence of egg-spots (irrespective of their position on the anal fin), whilst excluding potential positional genes and therefore confirming their involvement in egg-spots formation. Both types of blotches showed very distinct expression profiles from the egg-spots, and substantial differences in gene expression were also found between the two types of blotches. A similar gene expression profile between the egg-spots of derived haplochromines and the blotch pattern in the basal haplochromine *P. philander* would be indicative of a common origin for both traits, whereas similar expression profiles between the haplochromine egg-spots and the blotch of *C. macrops* would suggest that convergent evolution of this trait involved the same genetic pathways. Our study reveals the opposite for the genes under investigation, i.e. egg-spots and blotches show different expression profiles and also the two types of blotches differ in gene expression profiles, suggesting that egg-spots and blotches do not share a genetic basis and that convergent phenotypic evolution does not correspond to parallelism at the genetic level.

Results and discussion

Transcript profile in anal fin and egg-spot tissue

In order to identify genes involved in egg-spot morphogenesis we quantified differences in gene expression patterns between egg-spots and the surrounding non-pigmented anal fin of six *Astatotilapia burtoni* males (Fig. 1). Illumina RNAseq (RNA sequencing) provided a total of 193,054,988 high quality reads from the six egg-spot tissue samples and 194,099,061 reads from anal fin tissue samples of the same individuals. The replicates for each tissue were sequenced separately and the average number of reads per sample was 3,226,2837.42 (2,750,960.2–3,226,2837.42). We mapped the reads from each replicate to a reference *A. burtoni* embryonic library, which is a transcript collection from several different embryonic and larval developmental stages, and therefore probably the most comprehensive available representation of the entire gene set from *A. burtoni* [29]. In total we identified 1229 genes that were DE between the two types of tissues, with 620 genes being over-expressed in the egg-spot tissue, whilst 609 were under-expressed (Table 1). The DE transcripts, their identification using tBLASTx and BLASTx searches (against the NCBI non-redundant database [30]), together with the respective expression levels, are provided in Additional file 1. A first inspection of those DE genes between egg-spot and non-egg-spot tissue revealed that our experiment retrieved many genes with a known function in pigment formation and patterning in different model organisms including *paired box 7* (*pax7*),

Table 1 Differential gene expression (DGE) statistics

DGE	Contigs	Contigs with BlastID	Annotated contigs
Over	620	377	178
Under	609	435	241
Total	1229	812	419

Number of genes over-expressed and under-expressed in the egg-spot, number of hits after BLASTx search against NCBI's *Danio rerio* protein database and number of BLAST2GO annotated contigs

endothelin receptor b1 (*ednrb1*), *microphthalmia-associated transcription factor a* (*mitfa*), *Agouti signaling protein 1* (*asip1*), *sex determining region Y box 10* (*sox10*) and *anaplastic lymphoma receptor tyrosine kinase* (*alk*) [31], suggesting that our strategy is a valid approach to identify candidate genes for egg-spot morphogenesis.

Functional annotation of the DE genes

The reference *A. burtoni* transcriptome was annotated by performing a BLASTx search against NCBI's *Danio rerio* protein database [30]. From the 1229 DE genes, 58.6 % (720) had significant BLAST hits against the database (annotated datasets can be found in Additional file 2), while 41.4 % (509) of the DE contigs were non-identified. From the 720 contigs with a BLAST hit we could functionally annotate 495 using BLAST2GO [32]. We further described the Gene Ontology (GO) term composition for egg-spot over-expression and egg-spot under-expression in comparison to the reference transcriptome GO representation (Fig. 2). Overall, the GO terms representation was similar between the two tissues. However, there were several GO terms for "Molecular function" and "Cellular component" that differed significantly between the two data-sets, suggesting, as expected, that the two tissues are functionally different (Fig. 2).

To narrow down the list of relevant GO terms, and to use them as a tool to find candidates, we used a two-sided Fisher's exact test (false discovery rate (FDR) <0.05) to determine which functional GO categories were enriched in the genes over-expressed in the egg-spot in comparison to the total embryonic transcriptome. Five categories were significantly enriched in our over-expression gene dataset: 'Pigmentation' (GO:0043473),

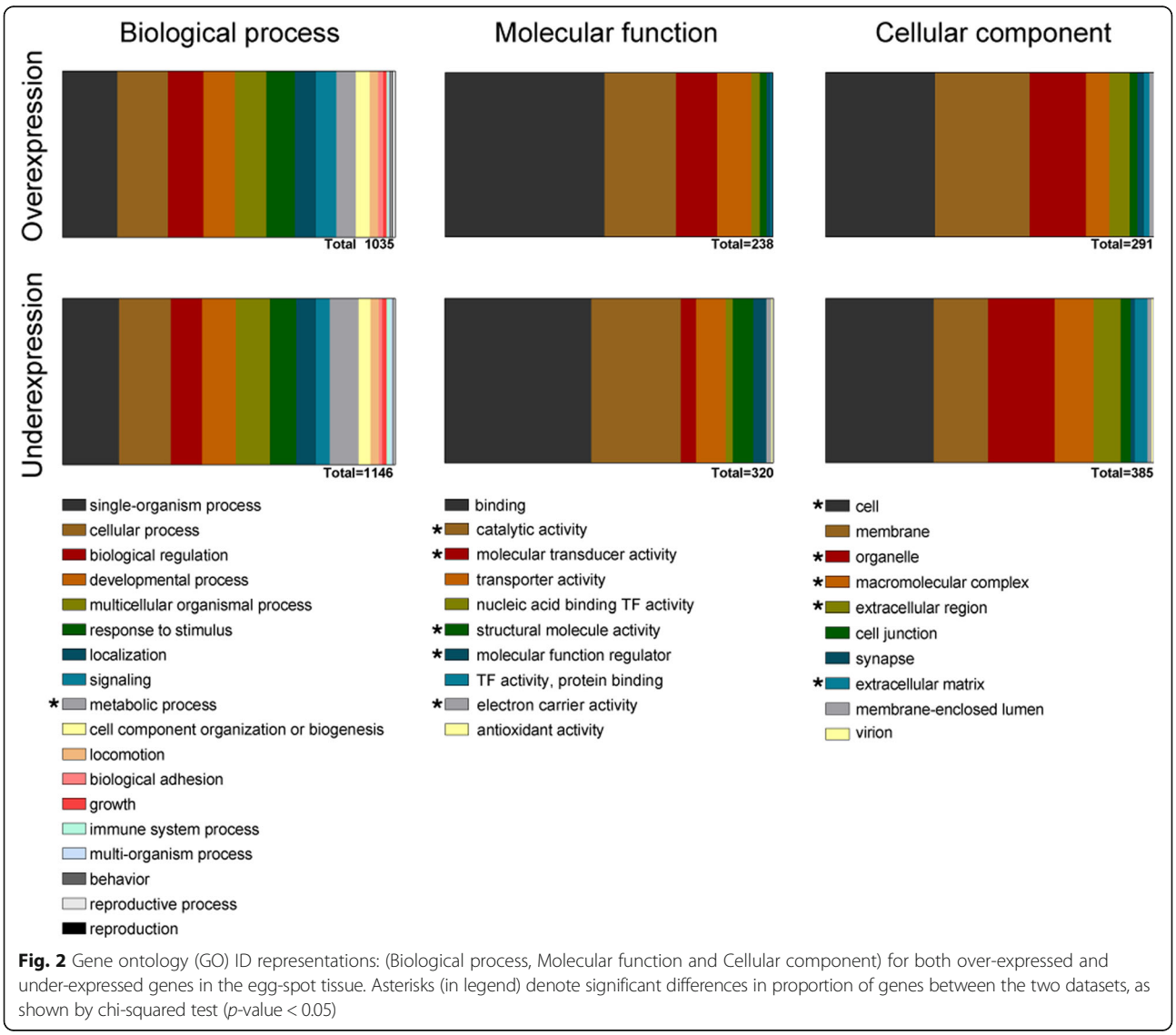

Fig. 2 Gene ontology (GO) ID representations: (Biological process, Molecular function and Cellular component) for both over-expressed and under-expressed genes in the egg-spot tissue. Asterisks (in legend) denote significant differences in proportion of genes between the two datasets, as shown by chi-squared test (p-value < 0.05)

'Developmental pigmentation' (GO:0048066), 'G-protein coupled peptide receptor activity' (GO:0008528), 'Peptide receptor activity' (GO:0001653) and 'Cell adhesion molecule binding' (GO:0050839) (Fig. 3). These are GO functional categories known to play a role in the development of pigmentation patterns. Neural crest cells are precursors of pigment cells and migrate from their original location to the anal fin where they will form the egg-spots [33–35], therefore genes playing a role in cell migration, cell adhesion and pigmentation development are relevant to the formation of this trait. Egg-spot formation relies on pigment production, which in turn is often activated via membrane receptor activity [36–38]. In Table 2 we present the list of genes belonging to these enriched functional categories that are potentially good candidates for egg-spot morphogenesis. The genes belonging to the GO term 'Developmental pigmentation' were overlapping with the ones included in the 'Pigmentation' category and the same is true for the two receptor GO term categories, therefore we only show three of the five enriched functional GO categories. This method of functional description of a gene dataset to extract candidates represents a supervised search, meaning that we might bias our findings towards what is already known. We note, however, that there are many other non-described genes, or known genes with incomplete GO term annotations, which could play a role in egg-spot morphogenesis.

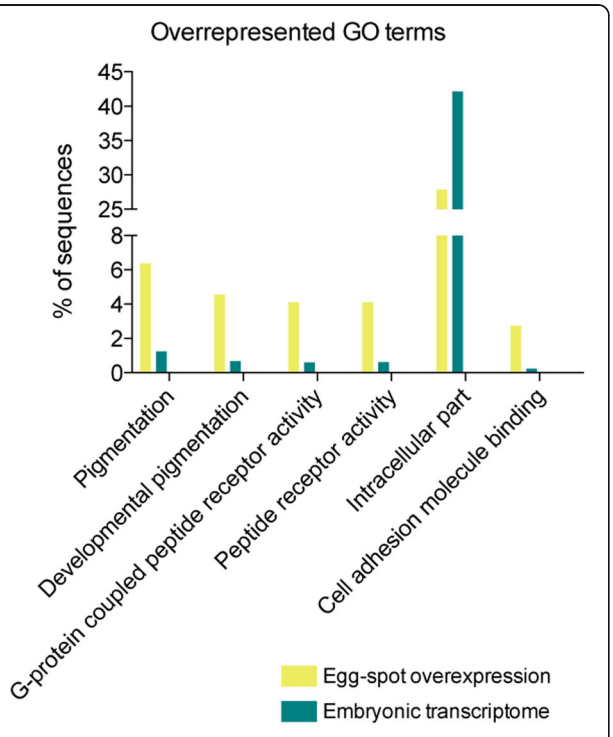

Fig. 3 Enrichment of functional GO terms in the egg-spot over-expressed genes (yellow bar) when compared to the total transcriptome of *A. burtoni* (blue bar). Those were calculated with a two-tailed Fisher exact test (FDR < 0.05)

Potential lineage specific genes are DE in the egg-spot

How novel traits emerge and are modified is one of the many unresolved problems in evolutionary biology [39–41]. It has long been advocated that new traits can emerge via the co-option of conserved regulators [42]. More recently, however, evidence is accumulating that new, i.e. lineage specific, genes can also play an important role in the development of novel traits [43–45]. Around 41 % of our candidate contigs did not have a BLAST hit against the *D. rerio* protein database. This could be due to the incompleteness of this database or to the lack of homologs in this species. To control for these factors we performed BLASTx and tBLASTx searches against the NCBI non-redundant (nr) protein and nucleotide databases [30]. Around 15.5 % (191/1229) of the DE contigs could not be assigned to a specific gene present in either nr database (Additional file 1). The contigs without positive BLAST hits could represent non-coding RNAs, partial sequences of known genes that could not be identified, or lineage specific genes (new or fast evolving genes) [46]. These results add to previous work on comparative transcriptomics of East African cichlids reporting that only 51 % of the total transcriptomes of the species studied (*A. burtoni* and *Ophthalmotilapia ventralis*) have hits on the NCBI nr nucleotide database [46]. In our case, the reduction in percentage of non-identified contigs is, most probably, due to the recent availability of five cichlid genomes [29].

It has previously been shown that lineage specific genes might play a role in the emergence and development of novel traits. In cnidarians 15 % of the transcripts expressed in a phylum specific cell type are lineage-specific, though the functional role of these transcripts was not tested [45]. The relative contribution of novel genes to the evolution of new morphologies, when compared to the co-option of conserved genes, is still under debate and further studies are needed to clarify their role on the evolution of such traits. Therefore, it would be interesting to identify the unknown DE transcripts and assess their role in the development and evolution of egg-spots.

Rates of evolution of the egg-spot DE genes

Changes in gene function can result either from modification in a *cis*-regulatory element that changes gene expression pattern and timing, and/or from a modification in the protein sequence that alters its function [47–50]. To test for protein sequence evolution in the egg-spot DE genes we calculated the rates of evolution in the form of dN/dS (ratio of non-synonymous substitutions over synonymous substitutions) of this gene dataset and compared the values obtained with a previously

Table 2 List of genes belonging to the GO term categories that are enriched in the egg-spot overexpressed dataset

	Gene	Transcript	logFC	BLASTx Identification	Accession	e-value
GO:0043473 Pigmentation						
1	ednrb	c5301_g0	0.926091617	endothelin B receptor [Haplochromis burtoni]	XP_005943243.1	0
2	rab38	c22025_g0	0.696999584	ras-related protein Rab-38 [Pundamilia nyererei]	XP_005720771.1	1.00E-149
3	pax7	c28600_g0	1.145533605	paired box protein Pax-7-like isoform X2 [Haplochromis burtoni]	XP_005948265.1	0
4	alk	c41674_g0	0.825942988	ALK tyrosine kinase receptor-like [Haplochromis burtoni]	XP_014192765.1	0
5	adrb1	c29399_g0	1.214559835	beta-1 adrenergic receptor [Pundamilia nyererei]	XP_005747452.1	0
6	gpnmb	c5056_g0	1.060188549	transmembrane glycoprotein NMB isoform X1 [Haplochromis burtoni]	XP_014191090.1	0
7	sox9a	c11994_g0	0.829590765	transcription factor Sox-9-A-like [Haplochromis burtoni]	XP_005923891.1	1.00E-127
8	mitf	c20716_g0	1.100168154	microphthalmia-associated transcription factor-like isoform X1 [Pundamilia nyererei]	XP_005731764.1	0
9	matp	c18656_g0	0.975922489	membrane-associated transporter protein [Haplochromis burtoni]	XP_005917392.1	0
GO:0001653 Peptide receptor activity						
1	ednrb	c5301_g0	0.926091617	endothelin B receptor [Haplochromis burtoni]	XP_005943243.1	0
2	calcrl	c8691_g0	1.204750964	calcitonin gene-related peptide type 1 receptor-like [Xiphophorus maculatus]	XP_005814950.1	2.00E-50
3	npyr1	c42378_g0	3.144400118	neuropeptide Y receptor type 1 [Haplochromis burtoni]	XP_005927047.1	0
4	rgr	c3216_g0	1.570045325	RPE-retinal G protein-coupled receptor [Haplochromis burtoni]	XP_005919610.1	3.00E-170
5	mc5r	c25961_g0	1.240861041	melanocortin receptor 5-like [Oreochromis niloticus]	XP_003452144.2	0
6	ackr3	c33293_g0	0.951452649	atypical chemokine receptor 3-like [Haplochromis burtoni]	XP_005950282.1	0
7	tacr3	c38449_g0	1.408219331	neuromedin-K receptor [Maylandia zebra]	XP_004549575.1	0
8	gcgr	c15641_g0	1.273842194	glucagon receptor [Haplochromis burtoni]	XP_005940348.1	0
GO:0050839 cell adhesion molecule binding						
1	jup	c20044_g0	0.785483788	junction plakoglobin-like [Haplochromis burtoni]	XP_014185585.1	0
2	postn	c318_g0	0.94955976	periostin-like isoform X2 [Haplochromis burtoni]	XP_005926524.1	0
3	cd200	c1300_g1	0.736112405	OX-2 membrane glycoprotein [Pundamilia nyererei]	XP_005747247.1	0
4	edil3	c4665_g0	1.050287092	EGF-like repeat and discoidin I-like domain-containing protein 3 isoform X1 [Oreochromis niloticus]	XP_005473287.1	0
5	cadm3	c4984_g1	0.941973544	cell adhesion molecule 3 isoform X1 [Haplochromis burtoni]	XP_005918142.1	0

published dataset that estimated transcriptome-wide dN/dS values between cichlid species [46]. We were able to estimate dN/dS values (averages across species pairwise dN/dS) for 196 out of the 1229 contigs (see Additional file 1). As expected, the majority of the genes were under purifying selection ($dN/dS < 1$) and there was no significant difference in the rates of evolution between the over and under-expressed genes (Fig. 4). However, for both the over- and under-expressed genes, the average dN/dS values were significantly higher than those of the entire transcriptome (Fisher's exact test, p-value <0.05), which means that, on average, the genes that are DE between the egg-spot and the anal fin are evolving at a faster rate. The haplochromine egg-spot is a male ornamental trait and, hence, most likely under sexual selection, either directly via female choice or via male-male competition [17–21]. Our results thus provide support to the general finding that genes underlying sexually selected traits evolve more rapidly [51–54].

We found seven genes to be under positive selection ($dN/dS > 1$), four of which were over-expressed in the egg-spot tissue (Table 3). Among them there are genes that play a role in neural crest differentiation (*tenascin*) and in cell migration (*tenascin*, *mucin* and *family with sequence similarity 110c* (*fam110c*)), which are important processes in pigmentation development [55–58]. The other genes have no a priori functional link with egg-spot formation. Nonetheless, due to their difference in expression and their signature of adaptive sequence evolution, they should be considered as good candidates and their functional roles in egg-spot development should be tested in the future.

Comparative gene expression via quantitative real time PCR

To confirm the results obtained via RNAseq, we examined a subset of 46 of the 1229 DE genes and tested their expression in egg-spot versus non-egg-spot tissue via

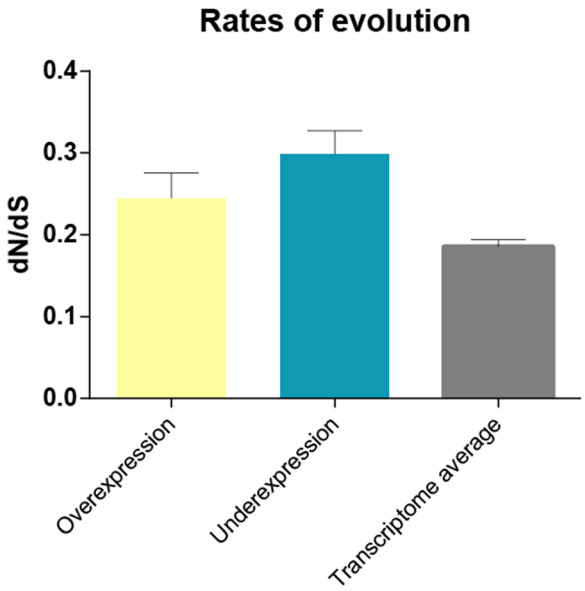

Fig. 4 Rates of evolution (dN/dS) for the over-expressed genes (yellow bar), under-expressed genes (blue bar) and for a previously published dataset that estimated transcriptome-wide dN/dS values between cichlid species. No significant difference was detected between the over- and under-expressed dataset, although both had significantly higher dN/dS than the transcriptome average (as determined by *t*-test, *p*-value < 0.01). Error bars denote standard error of the mean

quantitative real-time PCR (qPCR) in a second haplochromine species with a different egg-spot arrangement on the anal fin, *Cynotilapia pulpican* from Lake Malawi (Fig. 1). Half of these genes were over-expressed and half under-expressed in the egg-spot (Tables 4 and 5, respectively). These candidate genes were chosen randomly across the spectrum of the different levels of expression (from 1.3 to 5 fold differences in gene expression). Under-expressed genes were included as they might be acting as pigmentation inhibitors, thus preventing the appearance of egg-spots in other regions of the anal fin when over-expressed. Overall, there was no obvious trend with respect to functional GO categories associated with the top DE genes (see Additional file 2). Note that six out of the 46 candidates remained unidentified after tBLASTx searches against a non-redundant NCBI database.

While the egg-spots of *A. burtoni* are located in the proximal region of the anal fin, *C. pulpican* has its egg-spots in the distal region of the anal fin. By measuring the expression of these genes in this species, we effectively control for positional effects in gene expression along the proximal-distal axis.

We also aimed to determine whether egg-spots and blotches share a conserved gene expression profile, which would indicate a common origin of these two traits. We thus tested if the candidate genes identified in *A. burtoni* had similar expression levels in the blotches of a basal haplochromine species (*Pseudocrenilabrus philander*) and in the blotches of a member of a distinct cichlid tribe, an ectodine species (*Callochromis macrops*), where this trait has likely evolved independently.

Comparative gene expression in haplochromine egg-spots

The qPCR gene expression analysis in the second haplochromine species revealed that 14 of the 23 genes that were over-expressed in the egg-spots of *A. burtoni* showed a similar expression pattern in *C. pulpican* (Fig. 5a), suggesting they are egg-spot specific and not simply involved in fin patterning. Among them are the previously identified egg-spot gene *fhl2a* [15], two transcription factors well known for their involvement in patterning and cell fate specification (*homeobox C12a* (*hoxC12a*) and *heart and neural crest derivatives expressed 2* (*hand2*)), and an important growth morphogen (*bone morphogenetic protein 3b* (*bmp3b*)) [59–61]. The detection of *fhl2a*, in particular, suggests that our results are robust, since the gene was recently shown to be over-expressed across egg-spot development [15]. Included in the list are five of the unidentified contigs.

The remaining nine genes that were over-expressed in the egg-spots of *A. burtoni* either showed no difference in expression (4) or were under-expressed (5) in the egg-spots of *C. pulpican* (Fig. 5a). These genes are most likely involved in fin rather than egg-spot patterning, as

Table 3 DE genes under positive selection and their identification as determined through BLASTx against the NCBI non-redundant database

	Gene	Transcript	dN/dS	logFC	BLASTx Identification	Accession	e-value
1	FAM110C	c41094_g0	1.0613	−1.077081516	protein FAM110C [Haplochromis burtoni]	XP_005914672.1	6.00E-100
2	mucin-5 AC-like	c21845_g0	1.1477	1.528944965	mucin-5 AC-like [Haplochromis burtoni]	XP_005952554.2	0
3	intestinal mucin-like	c3522_g2	1.1479	0.741029968	intestinal mucin-like protein [Haplochromis burtoni]	XP_005941718.1	0
4	tenascin-like	c2897_g0	1.2524	2.61868262	tenascin-like [Haplochromis burtoni]	XP_005943223.1	0
5	myosin-IIIa	c23722_g0	1.2911	−0.911278787	myosin-IIIa isoform X5 [Haplochromis burtoni]	XP_014192226.1	0
6	polyubiquitin-like	c3172_g0	1.8501	−1.096942463	polyubiquitin-like [Haplochromis burtoni]	XP_014194859.1	1.00E-104
7	testican 1	c4037_g0	1.9352	1.068554755	testican-1 [Maylandia zebra]	XP_004545476.1	0

Table 4 Differentially over-expressed transcripts and their identification as determined through BLASTx against the NCBI non-redundant database

	Gene	Transcript	logFC	BLASTx Identification	Accession	e-value
1	asip1	comp13033_c0	3.143700418	agouti-signaling protein-like [Oreochromis niloticus]	XP_003448419.1	3.00E-25
2	rbp7	comp8091_c0	3.229469794	retinoid-binding protein 7-like [Oreochromis niloticus]	XP_003448369.1	9.00E-91
3	hand2	comp22787_c0	3.511901296	heart- and neural crest derivatives-expressed protein 2-like [Oreochromis niloticus]	XP_003452793.1	2.00E-96
4	*NA*	comp17910_c0	2.484101474	No significant similarity found	*NA*	*NA*
5	*NA*	comp20229_c0	2.626648739	hypothetical protein LOC100708826 [Oreochromis niloticus]	XP_003455230.1	6.00E-19
6	IF ON3	comp1238_c0	2.271395094	intermediate filament protein ON3-like [Oreochromis niloticus]	XP_003441441.1	0
7	*NA*	comp23328_c0	2.751465615	No significant similarity found	*NA*	*NA*
8	akap12	comp28860_c0	2.392200617	A-kinase anchor protein 12 [Danio rerio] > gb|ABQ11279.1| gravin [Danio rerio]	NP_001091654.1	2.00E-49
9	bmp3b	comp14170_c0	1.907176985	bone morphogenetic protein 3B-like [Oreochromis niloticus]	XP_003438593.1	0
10	*NA*	comp23699_c0	2.056104188	No significant similarity found	*NA*	*NA*
11	rbp4a	comp104_c0	1.758056096	retinol-binding protein 4-A-like [Oreochromis niloticus]	XP_003441907.1	2.00E-132
12	hoxC12a	comp21426_c0	2.020913618	Hoxc12a [Haplochromis burtoni]	ABS70754.1	2.00E-172
13	cytl1	comp7733_c0	1.730109411	cytokine-like protein 1-like [Oreochromis niloticus]	XP_003441598.1	4.00E-80
14	*NA*	comp24816_c0	1.803818569	No significant similarity found	*NA*	*NA*
15	sfr5	comp6979_c0	1.70609137	secreted frizzled-related protein 5-like isoform 3 [Oreochromis niloticus]	XP_003451970.1	0
16	*NA*	comp4443_c1	1.661176264	No significant similarity found	*NA*	*NA*
17	fhl2a	comp2939_c0	1.543403442	four and a half LIM domains protein 2-like [Oreochromis niloticus]	XP_003453001.1	0
18	cecr5	comp6479_c0	1.505843782	cat eye syndrome critical region protein 5-like [Oreochromis niloticus]	XP_003457763.1	0
19	zygin1	comp2115_c0	1.527432266	fasciculation and elongation protein zeta-1-like [Oreochromis niloticus]	XP_003449843.1	0
20	vtn	comp7947_c0	1.490014821	vitronectin-like [Oreochromis niloticus]	XP_003458657.1	0.00E + 00
21	igf1	comp17864_c0	1.458424511	insulin-like growth factor 1 [Oreochromis niloticus]	XP_003448107.1	7.00E-94
22	igSF10	comp36206_c0	1.484184706	immunoglobulin superfamily member 10-like [Oreochromis niloticus]	XP_003454869.1	0
23	fmdo	comp19154_c0	1.343960756	fibromodulin-like [Oreochromis niloticus]	XP_003441412.1	0

suggested by the fact that three of these of genes are known to participate in fin development (*retinol binding protein 7* (*rbp7*), *retinol binding protein 4* (*rbp4*) and *insulin-like growth factor 1* (*igf1*)) [62–64]. Overall, we confirmed the over-expression of 14 genes in the adult egg-spots from both *A. burtoni* and *C. pulpican* making them strong candidates genes for egg-spot formation that deserve further investigation.

Among the 23 under-expressed genes in *A. burtoni*, 15 were also consistently under-expressed in the egg-spots of *C. pulpican* (Fig. 5b), including one unidentified contig. Again, this suggests that these genes are egg-spot related. Among them is *aristaless 3* (*Axl3*), a gene belonging to the homeobox gene family, known for its patterning effects [65]. *Axl3* displays the highest expression differences among all genes (under- and over-expression included) and might putatively represent an inhibitor of the pigmentation/egg-spot pattern, although no role in pigmentation has been reported yet. The remaining eight genes showed no differences in gene expression between egg-spot and anal fin tissue on *C. pulpican*, and could therefore be involved in fin patterning. Thus far, none of these eight genes have been related to a function in pigmentation.

We cannot rule out that the genes that did not show the same pattern in both species do not have a function in egg-spots. Although egg-spots in *A. burtoni* and *C. pulpican* are homologous they do not necessarily have to share the exact same genetic network. It is thus possible that the DE genes might be responsible for interspecific differences of the egg-spot phenotype acting in a lineage-specific manner as has been shown in other taxa. For instance, the eyespots (concentric wing pigmentation patterns) of nymphalid butterflies, which are arranged along the distal half of the wing, are considered homologous [43, 66]. Nevertheless, there is a great flexibility in the expression patterns of four genes involved in the development of these structures in the different species studied: *antennapedia* was the only gene where there was a gain of expression associated with the origin of the eyespot phenotype, whereas there were many gain or loss events for *notch*, *distalless* and *spalt* in the different species [67]. Overall, the genetic network underlying the nymphalid eyespot pattern

Table 5 Differentially under-expressed transcripts and their identification as determined through BLASTx against the NCBI non-redundant database

	Gene	Transcript	logFC	BLASTx Identification	Accession	E-value
1	axl3	comp20108_c0	−5.023523546	homeobox protein aristaless-like 3-like [Danio rerio]	XP_695330.1	2.00E-152
2	and1	comp5622_c0	−3.032229958	actinodin1 precursor [Danio rerio]	NP_001184183.1	4.00E-124
3	slc13m5	comp28513_c0	−3.143689749	solute carrier family 13, member 5 [Danio rerio]	NP_001136038.1	0
4	oc	comp5530_c0	−3.008253114	osteocalcin [Oreochromis niloticus]	XP_003443144.1	2.00E-62
5	*NA*	comp36289_c0	−3.547180945	hypothetical protein LOC100695447 [Oreochromis niloticus]	XP_003459280.1	2.00E-50
6	and4	comp2301_c0	−2.691704824	actinodin4 precursor [Danio rerio]	NP_001129716.1	1.00E-85
7	carp	comp10574_c0	−2.74115746	cocaine- and amphetamine-regulated transcript protein-like [Oreochromis niloticus]	XP_003456941.1	3.00E-58
8	*NA*	comp116662_c0	−2.679644343	No significant similarity found	*NA*	*NA*
9	*NA*	comp29518_c0	−2.574270324	No significant similarity found	*NA*	*NA*
10	hdd11	comp1748_c0	−2.191735413	putative defense protein Hdd11-like [Oreochromis niloticus]	XP_003446154.1	8.00E-127
11	iunh	comp29726_c0	−1.991962941	inosine-uridine preferring nucleoside hydrolase-like [Oreochromis niloticus]	XP_003455949.1	6.00E-55
12	hbba	comp70_c0	−1.747612794	hemoglobin subunit beta-A-like isoform 1 [Oreochromis niloticus]	XP_003442119.1	9.00E-99
13	matn4	comp4244_c0	−1.775099485	matrilin-4 [Oreochromis niloticus]	XP_003451941.1	0
14	tsp4	comp2186_c1	−1.66042901	thrombospondin-4-B-like [Oreochromis niloticus]	XP_003451568.1	0
15	mmp13	comp20376_c0	−1.855094613	collagenase 3-like [Oreochromis niloticus]	XP_003441718.1	0
16	col9a1	comp6219_c0	−1.614663219	collagen alpha-1(IX) chain-like, partial [Danio rerio]	XP_003200573.1	2.00E-138
17	caytaxin	comp7321_c0	−1.667845939	caytaxin-like [Oreochromis niloticus]	XP_003448582.1	0
18	ltl	comp656_c0	−1.547716431	lily-type lectin [Epinephelus coioides]	AEA39736.1	3.00E-69
19	phospho1	comp2411_c0	−1.545453078	probable phosphatase phospho1-like [Oreochromis niloticus]	XP_003442063.1	0
20	pai1	comp29400_c0	−1.616263913	plasminogen activator inhibitor 1-like [Oreochromis niloticus]	XP_003460165.1	0
21	hbaa	comp28_c0	−1.541854961	Hemoglobin subunit alpha-A	Q9PVM4.3	1.00E-79
22	loxl4	comp12727_c0	−1.470941331	lysyl oxidase homolog 4-like [Oreochromis niloticus]	XP_003455871.1	0.00E + 00
23	cd81	comp5209_c0	−1.491910445	CD81 antigen-like [Oreochromis niloticus]	XP_003443898.1	0.00E + 00

appears to be highly variable, suggesting that homologous structures are not necessarily controlled by the same set of genes. Perhaps the same is true for cichlid egg-spots, which might initially have been under the control of the same set of genes followed by diversification in the recruitment of different genes. A broader phylogenetic sampling of egg-spot phenotypes would be necessary to clarify this question.

The 29 genes that were consistently over- or under-expressed in the adult egg-spots in both haplochromine species are nevertheless strong candidates genes for egg-spot development and merit further investigation to understand their role in the origin and diversification of this trait. These genes should be studied in detail throughout development and their function should be tested, not only in one species but also across several species of egg-spot bearing haplochromines with variable egg-spot phenotypes. With this approach we will be able to distinguish between a functional role in the evolution of the trait and or merely a function in the development and/or physiology of the trait.

Comparative gene expression between egg-spots and haplochromine blotches

We then measured gene expression of our set of 46 candidate genes in a basal haplochromine species, *Pseudocrenilabrus philander*, which displays a blotch rather than an egg-spot on its anal fin (Fig. 1). It is not known whether the blotches found in basal haplochromines are ancestral to the egg-spots found in 'modern haplochomines' [13, 25]. Homology inferences are typically made according to shared phenotypic criteria between traits and also according to parallelism at the developmental and genetic level [68] Therefore, if egg-spots and blotches are homologous we might expect that the gene expression patterns in both traits are, at least, partially conserved.

According to our results, haplochromine blotch and egg-spots differ substantially in their expression profiles (Fig. 5a, b). None of the 14 genes that were over-expressed in both *A. burtoni* and *C. pulpican* egg-spots were over-expressed in the blotch of *P. philander* (Fig. 5a), and only four of the 15 genes under-expressed

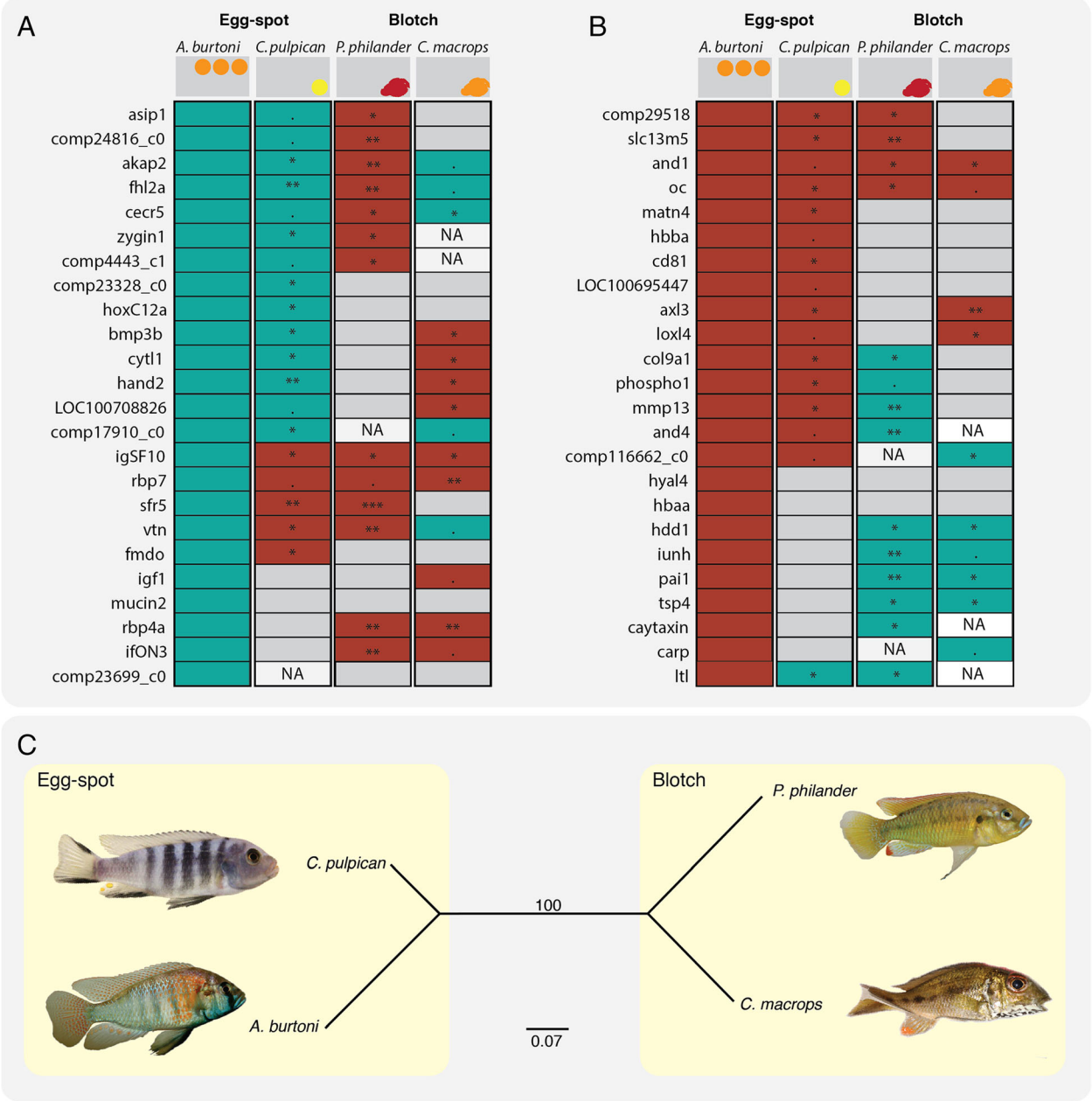

Fig. 5 Gene expression results for 46 DE genes as measured by qPCR. qPCR was performed for *C. pulpican*, *P. philander* and *C. macrops* (Relative position of the egg-spot/blotch on the fin are shown on top of each panel). Expression of these genes was quantified in the egg-spots and blotches relative to the anal fin tissue. Blue box denotes over-expression, red denotes under-expression and grey denotes no significant difference. Instances where it was not possible to measure gene expression are colored white with NA. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, •: $p < 0.1$ (for more details please see Additional files 4, 5 and 6). **a** Results for egg-spot over-expression dataset (Table 4). In the first column are the RNAseq results for *A. burtoni*. In the second, third and fourth column are the results for *C. pulpican*, *P. philander* and *C. macrops* respectively. **b** Results for egg-spot under-expression dataset (Table 5). Details of the statistical analyses used are found in Additional file 4 (*P.pulpican*), Additional file 5 (*P. philander*) and Additional file 6 (*C. macrops*) **c** Distance tree calculated using the gene expression results (over-expression, under-expression and no difference of expression) as characters

in the two modern haplochromines were also under-expressed in *P. philander* (Fig. 5b). Although not conclusive, the poorly conserved expression pattern between the two traits suggests that the haplochromines' egg-spots and the blotches have emerged independently within the Haplochromini lineage.

These results have to be taken with caution, though, as haplochromine egg-spots could have evolved from

blotches by up-regulation of different effector genes within the same genetic network. This has been observed in *Drosophila*, where the phenotypically diverse wing pigmentation patterns are controlled by the key regulator *distalless* (*dll*) [49]. The emergence of this wing spot phenotype was brought by the evolution of regulatory links between *dll* and multiple downstream pigmentation genes, which resulted in their up-regulation in the wing [49].

Comparative gene expression between eggs-spots, haplochromine and ectodine blotches

The blotch phenotype evolved more than once and is also found in some ectodine cichlids from Lake Tanganyika [12]. Ectodine anal fin blotches are similar to the ones found in basal haplochromines (Fig. 1), but apparently have an independent origin [25]. Although non-homologous, ectodine blotches might still share the same genetic network with both haplochromine egg-spots and blotches, as has previously been shown for other convergent traits [69].

In this study, we measured gene expression of our set of 46 candidate genes in the blotch of *Callochromis macrops*. Our gene expression assays revealed that only four of the genes that were over-expressed in *A. burtoni* and *C. pulpican* egg-spots were also over-expressed in the blotch of *C. macrops* (Fig. 5a). They encode transcription factors (*cat eye syndrome critical region 5* (*cecr5*)), co-factors (*fhl2a*) [70], cytoskeleton components and kinases (*a-kinase anchoring protein 2* (*akap2*)) and a non-identified transcript. These genes could be related to the pigmentation patterning or production of pigment in all three species. Furthermore, *C. macrops* also shares with *A. burtoni* and *C. pulpican* four genes that are consistently under-expressed in both species (Fig. 5b). One gene (*vitronectin* [71]) was over-expressed in *C. macrops* blotch and *A. burtoni* egg-spots, but not in *C. pulpican* egg-spots. These two species (*A. burtoni* and *C. macrops*) have in common that their egg-spots and blotches, respectively, contain orange pigments, while the egg-spot of *C. pulpican* is yellow. These genes might therefore correlate with patterning or production of orange pigment, although no such role has been previously described.

The comparison of expression profiles between the blotch bearing *P. philander* and *C. macrops* revealed that the underlying gene expression patterns are different indicating that there is probably no parallel evolution at the genetic level determining the phenotypic resemblance of the blotches. Curiously there are six genes that are under-expressed in the *A. burtoni* egg-spots that show no difference in expression in *C. pulpican*, but are over-expressed in blotches of both *P. philander* and *C. macrops*. The expression pattern of those six genes could be correlated to the blotch phenotype, but the most probable explanation is that they are involved in fin morphogenesis, since the non-pigmented region of *A. burtoni* matches the pigmented one in the two species with blotches.

Gene expression clustering

To determine the relationship between the pigmented anal fin tissues (egg-spots and blotches), we coded the gene expression results of the 46 genes in the four different species into a matrix of discrete data points (0 – under-expression, 1 – no difference, 2 – over-expression) and constructed a distance genealogy (Fig. 5c). The resulting tree diagram shows a clear separation between egg-spot and blotch phenotype. The different species clearly cluster by gene expression phenotype (bootstrap of 100 %) and the observed similarities do not correspond to the species phylogeny (Fig. 5c, Table 6). The character distance matrix also shows that of the two blotches, *C. macrops* blotch is more similar to the haplochromine egg-spots in terms of gene expression (Fig. 5c, Table 6). Our results suggest that egg-spots, haplocromine blotches and ectodine blotches are not regulated by the same genetic components.

Overall our results suggest that haplochromine egg-spots, haplochromine blotches and ectodine blotches are novel pigmentation traits that evolved independently by re-using a limited number of common genes (Fig. 5 and Table 6). The genes in common seem to be related to the cellular composition of the trait, which is re-used every time a new pigmentation pattern emerges, and not with the pigmentation pattern per se. Therefore, a thorough comparison of the different fin phenotypes should be done to assess what are the cellular components of each of the pigmentation phenotypes to better understand and interpret the gene expression underlying it.

These homology inferences have to be taken with caution, as we have only studied a subset of candidate genes (46/1229) derived from the egg-spot versus non-egg-spot tissue transcriptomic comparison in *A. burtoni*. An in-depth comparison of the blotch tissue will certainly require comparative transcriptomics in the blotched species.

Table 6 Mean character distance matrix produced by PAUP

Species comparison		Distance	Genes that differ in expression
P. pulpican	*A. burtoni*	0.38297874	18
P. philander	*A. burtoni*	0.91111112	41
P. philander	*P. pulpican*	0.70454544	31
C. macrops	*A. burtoni*	0.79069769	34
C. macrops	*P. pulpican*	0.69047618	29
C. macrops	*P. philander*	0.47499999	19

Species cluster according to gene expression and not according to phylogeny

Conclusions and future perspectives

Understanding the genetic and molecular basis of both evolutionary innovation and phenotypic variation is a major challenge in evolutionary biology. Using next-generation sequencing we here present a transcriptional survey of egg-spot tissue in the haplochromine cichlid *Astatotilapia burtoni*. This collection of DE transcripts represents the largest set of egg-spot candidate genes available and will greatly contribute to the understanding of the genetics underlying this trait. We provide a list of 1229 genes that are DE between egg-spots and non-egg-spots fin tissues, many of which are fast evolving genes that might be involved in the genetic network determining the egg-spot phenotype.

A closer look at the expression profiles of 46 of the DE genes shows that the expression profiles are not conserved between egg-spots and blotches, which suggests that haplochromine egg-spots, haplochromine blotches and ectodine blotches do not share the same genetic basis. This result indicates that these traits emerged independently in the evolution of this group of fishes. It has been hypothesized that egg-spots are modifications of the "Perlfleckmuster" (pearly spot) pattern that is present in fins of many cichlid species [12, 14]. In the future it will be interesting to determine if the same genes that underlie the egg-spots of haplochromines are also expressed in the "Perlfleckmuster".

With our current approach, we identified 29 genes whose expression patterns are egg-spot specific in two distinct cichlid species, strongly pointing to a role in the formation of this trait. These genes definitely deserve further investigation; in particular, their expression dynamics should be examined during egg-spot development and their function should be assessed with transgenic experiments, now available for cichlids [72]. The functional characterization of these genes during egg-spot development and in a broader phylogenetic context will inform us about the origin and diversification of this innovation in the most species rich vertebrate lineage – the haplochromine cichlid fishes – thus leading to major advances in the understanding of the emergence and diversification of novel traits.

Methods

Samples

Astatotilapia burtoni and *Cynotilapia pulpican* bred laboratory strains were kept at the University of Basel (Switzerland) under standard conditions (12 h light/12 h dark; 26 °C, pH7). All individuals were euthanized with MS222 (Sigma-Aldrich, USA), following approved procedures (permit number 2317 issued by the Basel cantonal veterinary office) before tissue dissections. *Callochromis macrops* individuals were captured at Lake Tanganyika, Mpulungu (Zambia), *P. philander* were captured in a river near Mpulungu (both under a research permit issued by the Department of Fisheries, Republic of Zambia). Dissections were carried out *in situ*, tissues were stored in RNAlater (Ambion, USA) and shipped to the University of Basel.

RNA extractions

Isolation of RNA was performed using TRIzol® (Invitrogen, USA). All dissected tissues were incubated in 750 μl of TRIzol and left at 4 °C overnight (or 8–16 hours). The tissues were homogenized with a BeadBeater (FastPrep-24; MP, Biomedicals, USA). Extractions proceeded according to manufacturer's instructions and DNase treatment was performed with DNA-Free™ (Ambion, USA). RNA quantity and quality was determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA). cDNA was synthetized using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA).

Differential gene expression analysis using RNAseq – Illumina

The anal fins of six *Astatotilapia burtoni* male juveniles were dissected and RNA was extracted from egg-spot and anal fin tissue for each individual. One microgram of RNA per sample was sent for library construction and Illumina sequencing at the Department of Biosystems Science and Engineering (D-BSSE), University of Basel and ETH Zurich. Samples were run in two lanes of an Illumina Genome Analyzer IIx (maximum read length was 50 base pairs (bp)).

The reads from each sample were mapped against a reference *A. burtoni* embryonic transcriptome that contains 171,136 reference transcripts. We mapped the reads from each library against the reference transcriptome using Bowtie2 as aligner [73] and RSEM (RNA-Seq by Expectation-Maximization) [74] as the method to estimate gene abundance. The individual RSEM files were concatenated into one single dataset and analyzed using the Bioconductor R package EdgeR [75]. Transcripts that had less than one count per million in one of the samples were discarded. We tested for differential expression between egg-spot and anal fin samples, using anal fin as reference. Since the samples were paired (each replicate of the egg-spot and anal fin belong to one individual fish), we included the individual information in the statistical model. For that we used a negative binomial generalized linear model (GLM) based on common dispersion using the individual as the blocking factor, i.e. we tested for consistent differences in expression between egg-spot and anal fin within individuals. Transcripts were considered as DE if, after correction for multiple testing, the false discovery rate (FDR) was lower than 0.05 [76].

Functional annotation of differential expressed transcripts

Gene ontology (GO) [77] annotation of the differential expressed transcripts was conducted with Blast2GO version 2.5.0 [32]. BLASTx searches were done against the *Danio rerio* protein database using a threshold of e^{-5} and maximum number of hits of 20. These GO terms were used to estimate transcript function. A table with the list of the differential expressed transcripts, their respective values of expression, and their GO terms is provided in Additional file 1. Between dataset differences in the proportion of genes for individual level 2 GO terms were tested by means of chi-squared tests with *p*-values adjusted for multiple tests using Bonferroni corrections [78]. The enrichment of functional GO terms in the egg-spot over-expressed gene dataset was calculated with a two-sided Fisher's exact test with a FDR of 0.05.

Rates of evolution for the differential expressed transcripts

Transcriptome data from the five available cichlid species (*Pundamila nyererei*, *Neolamprologus brichardi*, *Oreochromis niloticus*, *Maylandia zebra*, and *Astatotilapia burtoni*) were downloaded from Broad Institute [29]. Each species' transcriptome consisted of multiple libraries that were concatenated. The 1229 DE genes from *A. burtoni* were compared using a BLASTn search (threshold: e^{-50}) against each species' transcriptome and DE genes with a hit in all cichlid species were retained (599). The 599 DE genes were then compared using BLASTx (threshold: e^{-20}) against the tilapia (*Oreochromis niloticus*) proteome from the ENSEMBL database and corresponding coding sequences (cds) retrieved (378). Finally, the database of 378 tilapia cds was queried against the individual cichlid transcriptomes using BLASTn (threshold: e^{-35}). BLAST outputs were parsed and filtered to retain hits with identity >90 %, length >200 bp and bit score >200. We obtained 298 tilapia cds that have at least a hit on all cichlid transcriptomes. A concatenated fasta file was built to include the ten top hits from each cichlid transcriptome and the 298 tilapia cds. Sequences were then aligned using MAFFT v7.245 [79] with einsi –adjustdirection options (einsi is suitable for sequences containing large unalignable regions, as expected with the presence of UTRs (untranslated regions) and splicing variants in our transcriptome data). Alignments were trimmed using the tilapia cds as a reference and visually inspected. Alignments with paralogous sequences resulting from recent duplications were discarded. Within each individual alignment a consensus was built across transcripts from each cichlid species with 'cons' from EMBOSS [80] (–plurality 1.5, indicating the cut-off for the number of positive matches below which there is no consensus). Alignments were then translated to proteins and checked for all sequences being in the corresponding tilapia reading frame (no stop codons). The whole pipeline was run with customized R and Unix scripts. We obtained 196 good alignments, 74 % of which comprised of all five cichlid species sequences, while the remaining included at least three species each. Average alignment length was 1716 bp, ranging from 270 to 7794 bp. Alignments are available from the author upon request. dN/dS estimates were calculated using the script kaks.pl in Bioperl [81] which computes the dN/dS for all sequence pairs, using the Nei-Gojobori method [82].

Gene expression analysis using qPCR

The expression of 46 genes (23 over-expressed genes in the egg-spot region and 23 under-expressed genes in the egg-spot) was further studied in three other species - *Cyno-tilapia pulpican*, *Pseudocrenilabrus philander* and *Callochromis macrops*. Primers were designed with GenScript Real-time PCR (TaqMan) Primer Design software available at https://www.genscript.com/ssl-bin/app/primer. Where possible, primers were designed in exon spanning regions to avoid effects of gDNA contamination. Primers were tested in all species and in cases where primers pairs did not work we designed new species-specific primers. Genes studied and primer sequences are available in Additional file 3.

Three qPCR experiments were carried out: *qPCR experiment 1*: Gene expression was compared between the non-egg-spot anal fin tissue and the egg-spot tissue of *C. pulpican*. This species has its egg-spot in a different position in the fin compared to *A. burtoni* (Fig. 1, $n = 4–5$). *qPCR experiment 2*: Gene expression was compared between the non-blotch anal fin tissue and blotch tissue of *P. philander* (Fig. 1, $n = 6$). In this experiment six individuals were used. *qPCR experiment 3*: Gene expression was compared between the non-blotch anal fin tissue and blotch tissue of *C. macrops* (Fig. 1, $n = 4–7$). In this experiment 4 to 7 individuals were used. For all experiments each individual was an independent replicate meaning that there was no pooling of samples.

The reactions were run on the StepOnePlus™ Real-Time PCR system (Applied Biosystems, USA) with FastStart Universal SYBR Green Master mix (Roche, Switzerland), following the manufacturer's protocols. All reactions were performed with an annealing temperature of 58 °C, a final concentration of cDNA of 1 ng/μl and a final primer concentration of 200 ng/μl. The comparative threshold cycle (CT) method [83] was used to calculate the relative concentrations between tissues, where anal fin was taken as the reference tissue and Ribosomal protein L7 (*rpl7*) or the Ribosomal protein SA3 (*rpsa3*) genes as endogenous controls. Primer efficiencies were calculated using standard curves. Efficiency values of test primers were comparable to the efficiency of endogenous control primers (*rpl7*, *rspa3*) and are available in Additional file 3.

Significant differential gene expression between egg-spot/blotch and anal fin was tested with a paired *t*-test. When the data did not conform to the assumptions of a *t*-test (normal distribution and equal variances), an unpaired *t*-test with Welch's correction or a Wilcoxon signed rank test was used. Normality of the data was tested using Shapiro-Wilk test and an *F*-test was used to determine if the variances of the datasets were equal. When the sample size was lower than five a Mann–Whitney test was used. Statistics were carried out using GraphPad Prism version 5.0a for Mac OS X (www.graphpad.com). Individual graphs for each gene studied and the details of the statistical results are given in Additional file 4 (*C. pulpican*), Additional file 5 (*P. philander*) and Additional file 6 (*C. macrops*). We could not test the expression of five of the genes for both datasets because the primers would not amplify at the required efficiency.

Distance calculation and tree based on the genes expression results

The qPCR gene expression results were encoded into a matrix of discrete data points according to their expression level (0 – under-expression, 1 – no difference, 2 – over-expression). Consequently, a neighbor-joining distance tree based on this matrix was constructed using PAUP* 4.0b10 [84] with 100 bootstrap pseudoreplicates (Fig. 5c). To further test the hypothesis that the expression pattern corresponds to the phylogenetic signal, the mean character differences distance for all pairwise comparisons between species were calculated based on the matrix.

Additional files

Additional file 1: Identity of the differential expressed genes between egg-spot and anal fin tissue together with the expression values and dN/dS estimations. (XLSX 222 kb)

Additional file 2: Gene ontology annotation for the DE genes. (XLSX 148 kb)

Additional file 3: Primers used in this study together with their efficiency values. (XLSX 50 kb)

Additional file 4: qPCR results for *P. pulpican*. (*XLSX 257 kb*)

Additional file 5: qPCR results for *P. philander*. (*XLSX 251 kb*)

Additional file 6: qPCR results for *C. macrops*. (*XLSX 253 kb*)

Acknowledgments

We would like to thank Anya Theis for the fish photographs; Adrian Indermaur, Anya Theis, Bernd Egger, Judith Bachman and Fabrizia Ronco for the sampling of *Pseudocrenilabrus philander*. This study was financially supported by a doctoral research fellowship from the Fundação para a Ciência e Tecnologia - FCT (SFRH/BD/43421/2008) to MES and the European Research Council - ERC (Starting Grant 'INTERGENADAPT' and Consolidator Grant 'CICHLID ~ X') to WS.

Availability of supporting data

The sequencing raw data for the twelve egg-spot and anal fin samples can be found in the NCBI Sequence Read Archive (SRA) database under the accession number of SRP075689.

Authors' contributions

MES and WS conceived the study, designed the research and wrote the manuscript. MES performed the RNA sequencing, differential gene expression analysis and the qPCR experiments for *C. pulpican* and *C. macrops* with the help of NB. LB performed the analysis on the rated of evolution of the DE transcripts. LG performed the qPCR experiments for *P. philander*. ZM performed the clustering analyses of the gene expression data. All authors read and approved the manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Not applicable.

Author details

[1]Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland. [2]Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure, CNRS UMR 5242, 46 Allée d'Italie, 69364 LyonCedex 07France. [3]Ecology Department, University of Barcelona, Av. Diagonal, 643, 08028 Barcelona, Spain. [4]Department of Zoology, Faculty of Science, Charles University in Prague, Vinicna 7, 128 44 Prague, Czech Republic.

Received: 10 March 2016 Accepted: 27 August 2016
Published online: 06 September 2016

References

1. Hubbard JK, Uy JAC, Hauber ME, Hoekstra HE, Safran RJ. Vertebrate pigmentation: from underlying genes to adaptive function. Trends Genet. 2010;26:231–9.
2. Mills MJ, Patterson LB. Not just black and white: Pigment patterns development and evolution in vertebrates. Semin Cell Dev Biol. 2009;20:72–81
3. Wittkopp PJ, Beldade P. Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. Semin Cell Dev Biol. 2009;20:65–71.
4. Endler J. Natural selection on color patterns in Poecilia reticulata. Evolution. 1980;34:76–91.
5. Cortesi F, Feeney WE, Ferrari MCO, Waldie PA, Phillips GAC, McClure EC, Sköld HN, Salzburger W, Marshall NJ, Cheney KL. Phenotypic plasticity confers multiple fitness benefits to a mimic. Curr Biol. 2015;25:949–54.
6. Boileau N, Cortesi F, Egger B, Muschick M, Indermaur A, Theis A, Büscher HH, Salzburger W. A complex mode of aggressive mimicry in a scale-eating cichlid fish. Biol Lett. 2015;11:20150521.
7. Kelsh RN. Pigment gene focus genetics and evolution of pigment patterns in fish. Cell Res. 2004;17:326–36.
8. Parichy DM, Spiewak JE. Origins of adult pigmentation: diversity in pigment stem cell lineages and implications for pattern evolution. Pigment Cell Melanoma Res. 2015;28:31–50.
9. Singh AP, Nüsslein-Volhard C. Zebrafish stripes as a model for vertebrate colour pattern formation. Curr Biol. 2015;25:R81–92.
10. Fujii R. The regulation of motile activity in fish chromatophores. Pigment Cell Res. 2000;13:300–19.
11. Bagnara J. Comparative anatomy and physiology of pigment cells in nonmammalian tissues. In: Nordlund JJ, Boissy RE, Hearing VJ, King RA OJ-P, editors. The pigmentary system: physiology and pathophysiology. New York: Oxford University Press; 1998. p. 9–40.
12. Fryer G, Iles T. The cichlid fishes of the great lakes of africa: their biology and evolution. Edinburgh, UK: Oliver & Boyd; 1972.
13. Salzburger W, Mack T, Verheyen E, Meyer A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol Biol. 2005;5:17.

14. Salzburger W, Braasch I, Meyer A. Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biol. 2007;5:51.
15. Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, Belting H-G, Affolter M, Salzburger W. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat Commun. 2014;5:5149.
16. Goldschmidt T, de Visser J. On the possible role of egg mimics in speciation. Acta Biotheor. 1990;38:125–34.
17. Hert E. The function of egg-spots in an African mouth-brooding cichlid fish. Anim Behav. 1989;37:726–32.
18. Hert E. Female choice based on egg-spots in Pseudotropheus aurora Burgess 1976, a rock-dwelling cichlid of Lake Malawi, Africa. J Fish Biol. 1991; 38:951–3.
19. Lehtonen TK, Meyer A. Heritability and adaptive significance of the number of egg-dummies in the cichlid fish Astatotilapia burtoni. Proc R Soc B. 2011; 278:2318–24.
20. Theis A, Salzburger W, Egger B. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. PLoS One. 2012;7:e29878.
21. Theis A, Bosia T, Roth T, Salzburger W, Egger B. Egg-spot pattern and body size asymmetries influence male aggression in haplochromine cichlid fishes. Behav Ecol. 2015;26:1512–9.
22. Kocher T, Conroy J. Similar morphologies of cichlid fish in lakes Tanganyika and Malawi are due to convergence. Mol Phylogenet Evol. 1993;2(2):158–65.
23. Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, Barluenga M, Salzburger W. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol Ecol. 2013;22:670–84.
24. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr Biol. 2012;22:2362–8.
25. Meyer BS, Matschiner M, Salzburger W. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol Phylogenet Evol. 2015;83:56–71.
26. Terai Y, Morikawa N, Kawakami K, Okada N. The complexity of alternative splicing of hagoromo mRNAs is increased in an explosively speciated lineage in East African cichlids. Proc Natl Acad Sci. 2003;100:12798–803.
27. Terai Y, Morikawa N, Kawakami K, Okada N. Accelerated evolution of the surface amino acids in the WD-repeat domain encoded by the hagoromo gene in an explosively speciated lineage of east African cichlid. Mol Biol Evol. 2002;19:574–8.
28. Roberts RB, Ser JR, Kocher TD. Sexual conflict resolved by invasion of a novel sex determiner in Lake Malawi cichlid fishes. Science. 2009;326:998–1001.
29. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, Turner-Maier J, Johnson J, Alcazar R, Noh HJ, Russell P, Aken B, Alföldi J, Amemiya C, Azzouzi N, Baroiller J-F, Barloy-Hubler F, Berlin A, Bloomquist R, Carleton KL, Conte MA, D'Cotta H, Eshel O, Gaffney L, Galibert F, Gante HF, et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 2014;513:375–81.
30. Pruitt KD, Tatusova T, Maglott DR. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 2005;33:D501–4.
31. Hoekstra HE. Genetics, development and evolution of adaptive pigmentation in vertebrates. Heredity. 2006;97:222–34.
32. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005;21:3674–6.
33. Kelsh RN, Harris ML, Colanesi S, Erickson CA. Stripes and belly-spots - a review of pigment cell morphogenesis in vertebrates. Semin Cell Dev Biol. 2009;20:90–104.
34. Dupin E, Sommer L. Neural crest progenitors and stem cells: from early development to adulthood. Dev Biol. 2012;366:83–95.
35. Christiansen JH, Coles EG, Wilkinson DG. Molecular control of neural crest formation, migration and differentiation. Curr Opin Cell Biol. 2000;12:719–24.
36. Braasch I, Liedtke D, Volff J-N, Schartl M. Pigmentary function and evolution of tyrp1 gene duplicates in fish. Pigment Cell Melanoma Res. 2009;22:839–50.
37. Parichy DM, Rawls JF, Pratt SJ, Whitfield TT, Johnson SL. Zebrafish sparse corresponds to an orthologue of c-kit and is required for the morphogenesis of a subpopulation of melanocytes, but is not essential for hematopoiesis or primordial germ cell development. Development. 1999; 126:3425–36.
38. Parichy DM. Temporal and cellular requirements for Fms signaling during zebrafish adult pigment pattern development. Development. 2003;130: 817–33.
39. Wagner GP, Lynch VJ. Evolutionary novelties. Curr Biol. 2010;20:R48–52.
40. Moczek AP. On the origins of novelty in development and evolution. BioEssays. 2008;30:432–47.
41. Pigliucci M. What, if anything, is an evolutionary novelty? Philos Sci. 2008;75: 887–98.
42. Carroll SB. Evolution at two levels: on genes and form. PLoS Biol. 2005;3:e245.
43. Martin A, Reed RD. Wingless and aristaless2 define a developmental ground plan for moth and butterfly wing pattern evolution. Mol Biol Evol. 2010;27: 2864–78.
44. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just orphans: are taxonomically-restricted genes important in evolution? Trends Genet. 2009;25:404–13.
45. Milde S, Hemmrich G, Anton-Erxleben F, Khalturin K, Wittlieb J, Bosch TCG. Characterization of taxonomically restricted genes in a phylum-restricted cell type. Genome Biol. 2009;10:R8.
46. Baldo L, Santos ME, Salzburger W. 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. 2011;3:443–55.
47. Carroll SB. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell. 2008;134:25–36.
48. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007;61:995–1016.
49. Arnoult L, Su K, Manoel D, Minervo C, Magrina J, Gompel N, Prud'homme B. Emergence and diversification of Fly pigmentation through evolution of a gene regulatory module. Science. 2013;6126:1423–6.
50. Brayer K, Lynch VJ, Wagner GP. Evolution of a derived protein–protein interaction between HoxA11 and Foxo1a in mammals caused by changes in intramolecular regulation. Proc Natl Acad Sci. 2011;108:E414–20.
51. Larracuente AM, Sackton TB, Greenberg AJ, Wong A, Singh ND, Sturgill D, Zhang Y, Oliver B, Clark AG. Evolution of protein-coding genes in Drosophila. Trends Genet. 2008;24:114–23.
52. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics. 2007;177: 1321–35.
53. Wright AE, Mank JE. The scope and strength of sex-specific selection in genome evolution. J Evol Biol. 2013;26:1841–53.
54. Ellegren H, Parsch J. The evolution of sex-biased genes and sex-biased gene expression. Nat Rev Genet. 2007;8:689–98.
55. Tucker RP, Mckay SE. The expression of tenascin by neural crest cells and glia. Development. 1991;1039:1031–9.
56. Tucker RP. Abnormal neural crest cell migration after the in vivo knockdown of tenascin-C expression with morpholino antisense oligonucleotides. Dev Dyn. 2001;222:115–9.
57. Lakshmanan I, Rachagani S, Hauke R, Krishn SR, Paknikar S, Seshacharyulu P, Karmakar S, Nimmakayala RK, Kaushik G, Johansson SL, Carey GB, Ponnusamy MP, Kaur S, Batra SK, Ganti AK. MUC5AC interactions with integrin β4 enhances the migration of lung cancer cells through FAK signaling. Oncogene. 2016;35:1–10.
58. Hauge H, Fjelland KE, Sioud M, Aasheim H-C. Evidence for the involvement of FAM110C protein in cell spreading and migration. Cell Signal. 2009;21: 1866–73.
59. Pick L, Heffer A. Hox gene evolution: multiple mechanisms contributing to evolutionary novelties. Ann N Y Acad Sci. 2012;1256:15–32.
60. Yelon D, Ticho B, Halpern ME, Ruvinsky I, Ho RK, Silver LM, Stainier DY. The bHLH transcription factor hand2 plays parallel roles in zebrafish heart and pectoral fin development. Development. 2000;127:2573–82.
61. Gamer LW, Ho V, Cox K, Rosen V. Expression and function of BMP3 during chick limb development. Dev Dyn. 2008;237:1691–8.
62. Blum N, Begemann G. Retinoic acid signaling controls the formation, proliferation and survival of the blastema during adult zebrafish fin regeneration. Development. 2012;139:107–16.
63. Tingaud-Sequeira A, Forgue J, André M, Babin PJ. Epidermal transient down-regulation of retinol-binding protein 4 and mirror expression of apolipoprotein Eb and estrogen receptor 2a during zebrafish fin and scale development. Dev Dyn. 2006;235:3071–9.
64. Chablais F, Jazwinska A. IGF signaling between blastema and wound epidermis is required for fin regeneration. Development. 2010;137:871–9.
65. McGonnell IM, Graham A, Richardson J, Fish JL, Depew MJ, Dee CT, Holland PWH, Takahashi T. Evolution of the Alx homeobox gene family: parallel

retention and independent loss of the vertebrate Alx3 gene. Evol Dev. 2011; 13:343–51.
66. Nijhout HF. Elements of butterfly wing patterns. J Exp Zool. 2001;225:213–5.
67. Shirai LT, Saenko SV, Keller RA, Jerónimo MA, Brakefield PM, Descimon H, Wahlberg N, Beldade P. Evolutionary history of the recruitment of conserved developmental genes in association to the formation and diversification of a novel trait. BMC Evol Biol. 2012;12:21.
68. Rutishauser R, Moline P. Evo-devo and the search for homology ("sameness") in biological systems. Theory Biosci. 2005;124:213–41.
69. Arendt J, Reznick D. Convergence and parallelism reconsidered: what have we learned about the genetics of adaptation? Trends Ecol Evol. 2008;23: 26–32.
70. Johannessen M, Møller S, Hansen T, Moens U, Van Ghelue M. The multifunctional roles of the four-and-a-half-LIM only protein FHL2. Cell Mol Life Sci. 2006;63:268–84.
71. Felding-Habermann B, Cheresh DA. Vitronectin and its receptors. Curr Opin Cell Biol. 1993;5:864–8.
72. Juntti SA, Hu CK, Fernald RD. Tol2-mediated generation of a transgenic haplochromine cichlid, Astatotilapia burtoni. PLoS One. 2013;8:e77647.
73. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
74. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12:323.
75. Robinson MD, McCarthy DJ, Smyth GK. EdgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26:139–40.
76. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc Ser B. 1995;57: 289–300.
77. Ashburner M, Ball C, Blake J. Gene ontology: tool for the unification of biology. Nat Genet. 2000;25(may):25–9.
78. Bonferroni CE. Teoria statistica delle classi e calcolo delle probabilità. Pubbl del R Ist Super di Sci Econ e Commer di Firenze. 1936;8:62.
79. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013; 30:772–80.
80. Rice P, Longden I, Bleasby A. The European Molecular Biology Open Software Suite EMBOSS. Trends Genet. 2000;16:2–3.
81. Stajich JE, Block D, Boulez K, Brenner SE, Chervitz SA, Dagdigian C, Fuellen G, Gilbert JGR, Korf I, Lapp H, Lehva H, Matsalla C, Mungall CJ, Osborne BI, Pocock MR, Schattner P, Senger M, Stein LD, Stupka E, Wilkinson MD, Birney E. The Bioperl Toolkit : Perl Modules for the Life Sciences. Genome Res. 2002;10:1611–1618.
82. Nei M, Gojobori T. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol. 1986;3:418–26.
83. Pfaffl MW. A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res. 2001;29:e45.
84. Swofford DL. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). 2003.

Chapter 3

The genetic basis of convergent evolution of two innovative anal fin pigmentation patterns in East African cichlid fish —— haplochromine eggspots and ectodine blotches

Langyu Gu, M. Emília Santos, Walter Salzburger

Chapter 3

The genetic basis of convergent evolution of two innovative anal fin pigmentation patterns in East African cichlid fish ——haplochromine eggspots and ectodine blotches

Langyu Gu[1], M. Emília Santos[2], Walter Salzburger[1]*

1 Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland.
2 Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure, CNRS UMR 5242, 46 Allée d'Italie, 69364 Lyon Cedex 07, France

*Correspondence:
Walter Salzburger, email: walter.salzburger@unibas.ch

Background:

The origination of evolutionary novelty is one of the most fascinating questions in evolutionary biology. To what extant that overall gene expression levels are associated with the convergent morphologies remains largely unknown. Convergent evolution of innovative anal fin pigmentation patterns in East African cichlid fish is an ideal model to study this question. One pattern is eggspots, i.e. a circular pigmentation pattern with a transparent outer ring in haplochromines with diversified numbers, colours and sizes in different species. Eggspots have been suggested to be involved in female attraction [1], male-male competition [2,3] and species recognition [4]. Another anal fin pigmentation pattern, blotches with ill-defined boundary, has independently evolved in the ectodine lineage. The ectodine blotch pattern was also suggested to be involved in female attraction [5], although less investigated. Unlike eggspots, the ectodine blotch has almost no variations among species.

Results:

On the basis of a comparative RNAseq experiment, we identified 274 ectodine blotch related candidate genes. By re-analyzing the existing raw data from Santos et al., [6] (Chapter 2) using same parameters as we did for ectodine blotch, we identified 812 eggspots related candidate genes. Further expression profile comparison showed that 15.7% (43/274) genes exhibited same expression pattern between ectodine blotch and eggspots. Interestingly, two genes (*col8a1b* and *steap4*) were highly expressed in eggspots, but were down-regulated in blotch, and 15 genes showed highly expression in blotch, but were downregulated in eggspots.

Conclusion:

Our results revealed a number of good candidate genes for further study about the anal fin pigmentation pattern formation in cichlid fish. Common shared gene expression profile suggested that these two pigmentation patterns share at least parts of a common gene network.

Key words

novelty, convergent evolution, ectodine blotch, eggspots, cichlid fish, transcriptome

Introduction

The origination of evolutionary novelty, i.e. "*is a structure that is neither homologous to any structure in the ancestral species nor serially homologous to any part of the same organism*" [7], belongs to the open questions in evolutionary biology. Such novelties provide raw materials for downstream selection and adaptation, and can contribute to biological diversification [8]. Examples of evolutionary novelties are eyespots on the wings of nymphalid butterflies [9], the neural crest in vertebrates [10], or the bird beaks [11,12]. The emergence of novel pigmentation patterns is particularly interesting, because such novel colour traits can have important functions in mimicry [13], sexual selection [14] and camouflage [15]. Yet, the mechanisms involved in the evolution of novel pigmentation patterns often remain unclear [16].

Comparison of convergent evolution cases might help unravel the genetic basis of innovation, considering that it can minimize the noise of individual evolutionary history and stochasticity effects [17]. Many studies have shown that protein-coding sequences changes are associated with convergent phenotypes, for example, antifreeze glycoproteins for cold adaptation was found to be driven by *de novo* mutations in different genes in Arctic cod and Antarctic notothenioid fish independently [18]. Besides, convergent evolution can also be driven by independent changes in gene expression, such as the expression of *prl* related to lactation in primates, mice and elephants [19]. However, only a small portion of studies focused on the whole gene expression level instead of individual gene level, such as the convergent evolution of pregnancy mechanisms between squamate reptile and mamalian [20], the convergent evolution of bioluminescent organs of squid [21], and the convergent thick-lipped phenotype in cichlid fish [22]. To what extent the overall expression level underlies the phenotypic convergence is largely unknown.

Why the question of whether convergent evolution happens in parallel changes in gene expression level is fascinating is because: 1) If gene expression patterns are largely evolutionarily conserved, these convergent traits might represent deep homologies [23,24] and the shared genes

might be related to developmental constraints of the novelties. 2) It can also help us identify the genes that are responsible for the modification of different novel phenotypes. Therefore, studying convergent novelty at the gene expression level will be useful to disentangle the genetic basis of the origination and modification of novelty. Besides, focusing on the gene network level can give a clue about how are these genes recruited independently and whether they are derived from pre-existing gene networks or originated *de novo*, which is related to the potential of trait evolvability [25]. For example, it has been shown that the ancestral *trans*-regulatory landscape was used for producing wing spots in *Drosophila*. Different *trans*- regulatory elements bind to the *cis*-regulatory elements, just like Christmas tree decoration (Christmas tree model) [26]). In this case, the spot gene network is tightly connected to the ancestral wing development gene network. Co-option of pre-existing pattern at new sites was also shown in the case of the recruitment of a *hedgehog* regulatory circuit in the evolution of butterfly eyespots in *Precis coenia* [27]. On the other hand, different transcription factors (TFs) can form their own cross-regulatory network so that the development of corresponding phenotypes can be relatively independent to the ancestral developmental system, such as butterfly eyespots [28]. With the advantages of next generation sequencing technology, more and more studies focus on the gene network level instead of individual genes [29].

Convergent evolution of innovative anal fin pigmentation patterns in East African cichlid fish is an ideal model to study the questions mentioned above (Fig.1) [30,31]. One of the pigmenation patterns, eggspots, the conspicuous pigmentation with circular markings on the anal fin, most likely emerged only once in the ancestor of the haplochromine lineage, the most species richness lineage of cichlids [32,33]. Haplochromine eggspots are highly diverse in different species with different numbers, colours and positions [34]. The presence of eggspots has been suggested to be associated with female attraction [1] and male-male competition [2,3]. Blotches, the pigmentation with ill-identified boundary, are mainly possessed by species in the more ancestral ectodine lineage [6], although some ancestral haplochromine species also have similar patterns, such as *Pseudocrenilabrus philander* [6]. It has also been suggested that blotches might be related to female attraction [5], although this is less be investigated. Noticeably, unlike haplochromine eggspots, ectodine blotches are less varied. Previously, a xanthophore related gene, *csf1ra*, was found to be expressed in anal fin pigmentation patterns in two independent lineages, and positive selection was found in the ancestral branch of haplochromine lineage, which might causally link this gene to the species-richness of haplochromine [33]. Another study suggested that a *cis*-regulatory change in the upstream region of *fhl2b* might be causally related to the formation of eggspots in haplochromines [32]; the same study also linked the expression of *fhl2b* to iridophore pigment cells. However, all these studies focused on individual candidate genes. Recently, based on

gene expression profiles of 46 eggspots candidate genes retrieved from an RNA sequencing experiment, Santos et al. [6] (Chapter 2) found little overlap between genes expressed in haplochromine eggspots and ectodine blotches, and suggested that these traits might not share a genetic basis. However, this study was only based on 46 out of 1229 (3.7%) candidate genes, and these genes are candidate genes related to eggspots instead of ectodine blotches, which could lead to a bias. Therefore, to test the genetic basis of eggspots and ectodine blotches, a more thorough gene expression profile comparison is needed.

In this study, by a thorough RNAseq experiment design controlling position effect and sex-determine effect, we identified 274 ectodine blotches related candidate genes. For better comparison, we re-analyzed the available raw transcriptomic data of eggspots from Santos et al., [6] (Chapter 2) using same parameters as we did for the ectodine blotch, and we identified 812 eggspots related candidate genes. Considering eggspots and blotch share common pigmentation cells (iridophore, xanthophore, melanophore), we predicted that these two convergent anal fin pigmentation patterns might share at least parts of a common gene network.

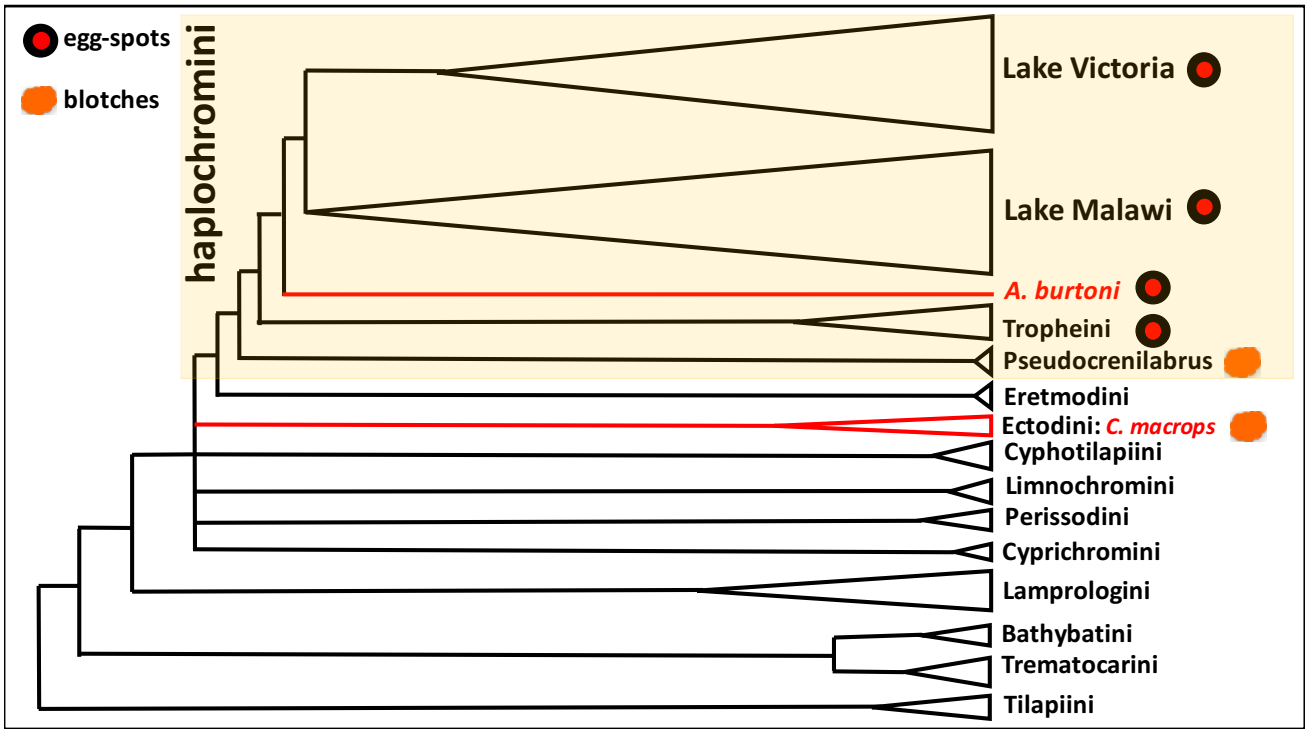

Fig. 1 Convergent evolution of anal fin pigmentation patterns in East African cichlid fish. Schematic molecular phylogeny of the East African cichlid fishes based on combined evidence from [35–37]. Triangle symbol represented species richness based on studies from [35] and https://en.wikipedia.org. Names on the right side indicated rivers or tribes.

Materials and Methods

1. Samples and experiment design

Laboratory strain of *C. macrops* was kept at the University of Basel (Switzerland) under standard conditions (12h light/12h dark; 26°C, pH=7). Prior to tissue dissection, specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following approved procedures (permit nr. 2317 issued by the cantonal veterinary office). For *C. macrops*, we first 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. what would be the corresponding areas to the blotch and non-blotch tissue in males (see Fig. 2). This led to a total of 12 samples for *C. macrops*.

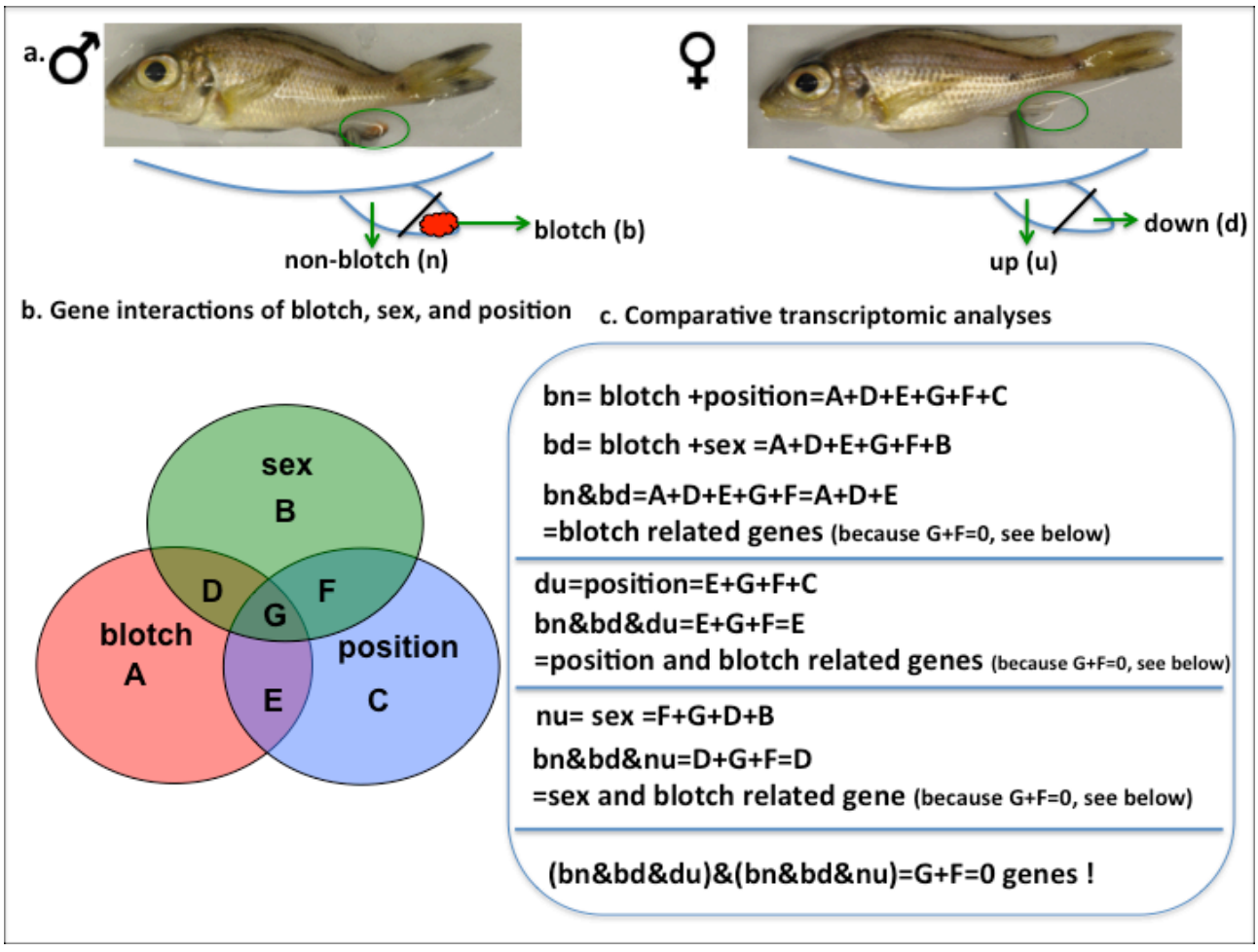

Fig.2 Comparative transcriptomic analyses to identify candidate genes related to blotch on the anal fin of cichlid fish, *C. macrops*. **a**. Four kinds of tissues (blotch tissue (b) and non-blotch tissue (n) in male; corresponding position of male blotch tissue in the up region of anal fin in female, down (d), corresponding position of male non-blotch tissue in female, up (u)) from 3 males and 3 females were sampled (totally 12 samples). **b**. Possible gene interactions of candidate genes derived from comparative transcritptomic analyses of four kinds of tissues. A, represents genes only be responsible for blotch phenotype; B, represents genes only be responsible for sex determination; C, represents genes

only are responsible for position; D, represents genes are responsible for both blotch and sex determination; E, represents genes are responsible for both blotch and position; F, represents genes are responsible for both sex determination and position; G, represents genes are responsible for blotch, position and sex determination. **c**. 1) highly differential expressed genes between b&n are related to blotch and/or position effect, corresponding to A+D+E+G+F+C in Fig.b ; highly differential expressed genes between b&d are related to blotch and sex-determination, corresponding to A+D+E+G+F+B; then the common genes between bn and bd are A+D+E+G+F. Note that here A+D+E are blotch-specific genes, and we will check what genes are responsible for G+F. 2) highly expressed genes between d&u are position related genes, which is E+G+F+C; and the common genes between bn, bd and du are E+G+F. Note that E is blotch+position related genes. 3) highly expressed genes between n&u are sex-determination related genes, which is F+G+D+B; and the common genes between bn, bd, and nu are D+G+F. Note that D is blotch+sex-determination related genes. By comparative transcriptomic data analysis (see below), there are none of genes are common for G+F, (G+F=0).

2. Illumina sequencing for the blotch in C. macrops

For RNA extraction, we used the TRIzol® protocol (Invitrogen, USA). Sample clean-up and DNase treatment were performed with RNA clean&Concentrator™-5 (Zymo Research Corporation, USA). RNA quality and quantity was determined with a Nanodrop 1000 spectrophotometer (Thermo Scientific, USA) and a Bioanalyzer 2100 (Agilent Technologies). Library construction was done in the Life Sciences Training Facility (LSTF), Pharmazentrum/Biozentrum, University of Basel. Libraries were generated using the Illumina TruSeq RNA sample preparation kit (Low-Throughput protocol) according to the manufacturer's instructions. For *C. macrops*, 330ng of RNA was subjected to mRNA selection using poly-T oligo-attached magnetic beads followed by chemical fragmentation. 15 PCR cycles were used to amplify the final libraries. Library quantification and quality assessment was performed on a Bioanalyzer 2100 (Agilent Technologies). The 12 samples were then subjected to Illumina RNA sequencing. 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 Analyzer IIx (maximum read length was 50 bp).

3. Comparative transcriptomic data analysis for the blotch in C. macrops

Quality assessment was conducted with Fastqc 0.10.1 (www.bioinformatics.babraham.ac.uk/projects/) to check per base sequence quality, over representing adapter sequences, and sequence length distribution. Contaminated Illumina adapter (the overrepresented adaptor detected by FastQC) anywhere in the reads were removed by using cutadapt 1.3 [38], with parameter -b ADAPTER -O 15 -e 0.02 -m 40. This means that if the

minimum overlap between read and adapter is more than or equal to 15 bp, the reads will be discarded (since after trimming 15 bp, the read length is for sure less than the minimum length required, 40 bp), allowing a maximum error rate of 0.02, i.e. one mismatch.

Then the reads were aligned to the Tilapia transcriptome assembly available from Broad Institute (ftp://ftp.ensembl.org/pub/release-81/fasta/oreochromis_niloticus/cdna/. version 0.71). This reference was indexed using NOVOINDEX (www.novocraft.com/) using default parameters. We then used NOVOALIGN (www.novocraft.com/) to map the reads against the reference using the following parameters; maximum alignment (t) 30, gap extend penalty (x) 4, and gap opening penalty (g) 30. Although the quality of our reads was very good overall, there were still some low quality reads. In this case, we turned on read trimming so that the reads that failed to align would progressively be shortened by 5bp (s 5) until they either align or the length was reduced to less than the minimum number of good quality bases for a read (l), which was set at 25 (half of the total read length). Considering that the same gene might have different transcripts leading to different alignment locations in the reference, but also to avoid situations where high copy number repeats result in hundreds of alignments for a read, we set 10 as the most randomly selected alignments are reported (r 10).

The output SAM files of read mapping were then transformed into BAM format using SAMtools [36]. Reads were then sorted, indexed and converted to count files. The count files were then concatenated into count tables and analyzed with the Bionconductor R package [39,40], Edge R [41–45] and DESeq [46]. To identify blotch-specific transcripts, we performed multiple comparisons (for details see Fig. 2). In the comparisons between blotch and non-blotch tissues, we used the individual as the block factor, since these tissues were paired. In a second analyses with edgeR, we kept genes that achieved at least one count per million (cpm) in at least three samples. To reduce false positive numbers, differentially expressed transcripts were kept if the false discovery rate (FDR) was smaller than 0.01.

4. Gene Annotation, pathway reconstruction and protein-protein interaction analysis for the blotch

Gene ontology (GO) annotation of the differential expressed transcripts was conducted with Blast2GO version 2.5.0 [47]. BLASTx searches were done using the BLASTx (threshold: e-6) and a number of hits of 10. Enrichment analysis of differentially expressed genes was implemented by the GOseq R package v2.12 [48]. GO terms with corrected p value less than 0.05 were considered

significantly enriched by differential expressed genes. KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway annotation was conducted with KAAS (KEGG Automatic Annotation Server) http://www.genome.jp/tools/kaas/ [49] using zebrafish as the reference with threshold e-value 1e-10. KOBAS v2.0 was used to test the statistical enrichment of differential expression genes in KEGG pathways [50]. Pathway with corrected p values 0.05 were significantly enriched, and top 20 significantly differential expressed genes enriched pathways were displayed in the results. Protein-protein interaction network analysis was conducted with string database (http://string-db.org/) using zebrafish as the reference with default parameters. Visualization of the network was done in Cytoscape v3.0 [51].

5. Gene expression profile comparison between the ectodine blotch and haplochromine eggspots

The existing transcriptomic data from anal fin with eggspots and without eggspots from three males of *A. burtoni* in Santos et al. [6] (Chapter 2) were re-analyzed using same parameters as we did for the ectodine blotch as mentioned above. For example, using Tilapia transcriptome assembly available from Broad Institute (ftp://ftp.ensembl.org/pub/release-81/fasta/oreochromis_niloticus/cdna/. version 0.71) as the reference; same parameters for reads mapping and transform. Since DESeq R package is much more conservative than edgeR package [52], we only used edgeR package to produce final differential expressed transcripts for eggspots. Noticeably, eggspots candidate genes from Santos et al. [6] (Chapter 2) were produced without position effect controlling, so that it could have bias. Therefore, we used the blotch related candidate genes derived from edgeR package from our study as the reference for the comparison.

Results and Discussion

For the ectodine blotch, illumina sequencing generated around 15 to 23 million raw reads per library, the average read quality was above 28 and between 40 to 160 thousand reads were filtered after adaptor trimming (the exact reads numbers see supplementary Table 1). Between eight to ten million reads were finally mapped to the reference. The unmapped reads could be unannotated transcripts, or low quality reads. Illumina reads are available from the Sequence Read Archive (SRA) at NCBI under the accession number SRA..... Multiple comparisons (details see Fig.2) using DESeq identified 51 blotch-related differentially expressed genes (Details see Table 1). Noticeably, they were all included in the 274 candidate genes identified with edgeR package (Table 2). Besides, all 51 genes from DESeq were up-regulated genes, but 11 out of 274 genes from edgeR (*tns4, steap4, rpf, trbc1, hmx1, col2a1b, cd59, keratin, col8a1b, clec19a, clec3a*) were down

regulated in blotch tissues (Table 2). For haplochromine eggspots, we identified 812 candidate genes with edgeR package (Supplementary Table 2). 15.2% (43/274) genes exhibited same expression pattern between the ectodine blotch and eggspots, suggesting that these two pigmentation patterns share at least parts of a common gene network. Besides, genes (*col8a1b and steap4*) were highly expressed in eggspots, but down-regulated in blotch, and 15 genes showed highly expression in blotch, but were down-regulated in eggspots (Table 2). These candidate genes are useful for further study about convergent evolution of anal fin pigmentation patterns in cichlid fish.

Table 1 51 The ectodine blotch related candidate genes derived from DESeq R package

gene	Ensembl ID	bn_condition	Ensembl_Description
MSMP	ENSONIT00000007835	-4.95	Microseminoprotein, prostate associated
APOD	ENSONIT00000023475	-3.59	apolipoprotein D
fhl2b	ENSONIT00000017889	-3.56	Four and a half LIM domains 2b
pnp5a	ENSONIT00000016474	-3.54	Purine nucleoside phosphorylase 5a
pnp4a	ENSONIT00000000731	-3.5	Purine nucleoside phosphorylase 4a
fhl2a	ENSONIT00000015512	-3.45	Four and a half LIM domains 2a
ifi30	ENSONIT00000016006	-3.31	interferon, gamma-inducible protein 30
gpnmb	ENSONIT00000005686	-3.21	Glycoprotein (transmembrane) nmb
rab38	ENSONIT00000010754	-3.18	ras-related protein Rab-38-like
tagln3b	ENSONIT00000017829	-3.17	Transgelin 3b
rgs4	ENSONIT00000013824	-3.15	Regulator of G-protein signaling 4
phlda1	ENSONIT00000026350	-3.07	Pleckstrin homology-like domain, family A, member 1
PRTFDC1	ENSONIT00000008139	-3.01	Phosphoribosyl transferase domain containing 1
tm4sf4	ENSONIT00000005261	-2.99	TM4SF4
keratin	ENSONIT00000023698	-2.96	keratin, type I cytoskeletal 20-like
slc23a1	ENSONIT00000013024	-2.87	solute carrier family 23 member 1-like
vtna	ENSONIT00000010298	-2.69	Vitronectin a
slc2a15a	ENSONIT00000016584	-2.62	Solute carrier family 2 (facilitated glucose transporter), member 15a
crip1	ENSONIT00000001209	-2.57	Cysteine-rich protein 1
TM4SF4	ENSONIT00000005262	-2.54	TM4SF4
pmelb	ENSONIT00000020824	-2.54	Premelanosome protein b
syngr1a	ENSONIT00000017418	-2.51	Synaptogyrin 1a
tfec	ENSONIT00000021193	-2.46	Transcription factor EC
rab27b	ENSONIT00000016030	-2.36	RAB27B, member RAS oncogene family
rab27b	ENSONIT00000021335	-2.36	RAB27B, member RAS oncogene family
retsat	ENSONIT00000007947	-2.3	putative all-trans-retinol 13,14-reductase
CRD2	ENSONIT00000015173	-2.29	cAMP-regulated D2 protein-like isoform X1
tmem235	ENSONIT00000025015	-2.25	transmembrane protein 235
mc1r	ENSONIT00000026719	-2.16	Melanocortin 1 receptor

zdhhc14	ENSONIT00000023686	-2.04	Zinc finger, DHHC-type containing 14
ACSL6	ENSONIT00000021738	-2.04	Acyl-CoA synthetase long-chain family member 6
trpm1a	ENSONIT00000004083	-2.01	Transient receptor potential cation channel, subfamily M, member 1a
novel	ENSONIT00000026203	-1.99	uncharacterize
PLCB2	ENSONIT00000015359	-1.98	Phospholipase C beta 2
cecr1b	ENSONIT00000000025	-1.94	Cat eye syndrome chromosome region, candidate 1b
gatsl2	ENSONIT00000016736	-1.91	GATS protein-like 2
prickle2a	ENSONIT00000021803	-1.9	Prickle homolog 2a
FAM180A	ENSONIT00000013027	-1.877	Family with sequence similarity 180 member A
slc2a5	ENSONIT00000023711	-1.84	solute carrier family 2, facilitated glucose transporter member 5-like
slc7a10a	ENSONIT00000003926	-1.83	Solute carrier family 7 (neutral amino acid transporter light chain, asc system), member 10a
PCBP3	ENSONIT00000013044	-1.83	poly(rC)-binding protein 3-like
ST8SIA3	ENSONIT00000021336	-1.82	sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase-like
ttyh2l	ENSONIT00000006814	-1.74	Tweety homolog 2, like
MYOC	ENSONIT00000023494	-1.63	Myocilin, trabecular meshwork inducible glucocorticoid response
TPD52l1	ENSONIT00000000849	-1.6	Tumor protein D52-like 1
trpm1b	ENSONIT00000019285	-1.57	Transient receptor potential cation channel, subfamily M, member 1b
impdh1b	ENSONIT00000014451	-1.5	IMP (inosine 5,-monophosphate) dehydrogenase 1b
igsf11	ENSONIT00000021051	-1.46	Immunoglobulin superfamily member 11
tubb5	ENSONIT00000016436	-1.42	Tubulin, beta 5
slc25a36b	ENSONIT00000018418	-1.41	Solute carrier family 25 (pyrimidine nucleotide carrier), member 36b
VEGFC	ENSONIT00000002365	-1.41	vascular endothelial growth factor C-like

Table 2 274 The ectodine blotch related candidate genes derived from edgeR R package

	Ensembl	genes_bd_bn	bn_logFC	Ensembl_Description
1	MSMP	ENSONIT00000007835	-4.85	microseminoprotein, prostate associated
2	slc4a11	ENSONIT00000015534	-4.20	solute carrier family 4, sodium borate transporter, member 11
3	TTC39B	ENSONIT00000022072	-3.72	tetratricopeptide repeat domain 39B
4	**APOD**	ENSONIT00000023475	-3.58	Apolipoprotein Da
5	**fhl2b**	ENSONIT00000017889	-3.56	four and a half LIM domains 2b
6	**pnp5a**	ENSONIT00000016474	-3.55	Purine nucleoside phosphorylase 5a
7	**pnp4a**	ENSONIT00000000731	-3.52	Purine nucleoside phosphorylase 4a
8	ly6d	ENSONIT00000024767	-3.50	lymphocyte antigen 6D-like
9	fhl2a	ENSONIT00000015512	-3.46	four and a half LIM domains 2a
10	cecr5	ENSONIT00000000334	-3.32	cat eye syndrome critical region protein 5-like
11	ifi30	ENSONIT00000016006	-3.29	interferon, gamma-inducible protein 30
12	rab38	ENSONIT00000010754	-3.21	ras-related protein Rab-38-like
13	gpnmb	ENSONIT00000005686	-3.21	Glycoprotein (transmembrane) nmb
14	BDH1	ENSONIT00000016804	-3.18	D-beta-hydroxybutyrate dehydrogenase, mitochondrial-like
15	tagln3b	ENSONIT00000017829	-3.164	Transgelin 3b
16	PNPLA1	ENSONIT00000015572	-3.12	Patatin like phospholipase domain containing 1
17	phlda1	ENSONIT00000026350	-3.10	Pleckstrin homology-like domain, family A, member 1
18	rgs4	ENSONIT00000013824	-3.09	Regulator of G-protein signaling 4
19	fzd7b	ENSONIT00000011796	-3.02	Frizzled class receptor 7b
20	fetub	ENSONIT00000013829	-2.96	Fetuin B
21	keratin	ENSONIT00000023698	-2.96	keratin, type I cytoskeletal 20-like
22	soat1	ENSONIT00000002460	-2.95	Sterol O-acyltransferase 1
23	pts	ENSONIT00000024648	-2.94	6-pyruvoyltetrahydropterin synthase
24	TM4SF4	ENSONIT00000005261	-2.88	TM4SF4
25	plin5	ENSONIT00000002684	-2.87	perilipin-5-like
26	slc23a1	ENSONIT00000013024	-2.87	solute carrier family 23 member 1-like
27	PNPL1	ENSONIT00000015571	-2.77	Patatin like phospholipase domain containing 1
28	stmn1a	ENSONIT00000009575	-2.77	Stathmin 1a
29	vtna	ENSONIT00000010298	-2.71	Vitronectin a
30	fam213ab	ENSONIT00000009957	-2.70	Family with sequence similarity 213, member Ab
31	slc2a15a	ENSONIT00000016584	-2.63	Solute carrier family 2 (facilitated glucose transporter), member 15a
32	crip1	ENSONIT00000001209	-2.61	Cysteine-rich protein 1
33	cax2	ENSONIT00000018605	-2.59	Cation/H+ exchanger protein 2
34	LGALS3	ENSONIT00000001154	-2.57	galectin-3-like isoform X1

35	TM4SF4	ENSONIT00000005262	-2.55	TM4SF4
36	syngr1a	ENSONIT00000017418	-2.54	Synaptogyrin 1a
37	pmel1b	ENSONIT00000020824	-2.53	Premelanosome protein b
38	tmeff1b	ENSONIT00000015587	-2.52	Transmembrane protein with EGF-like and two follistatin-like domains 1b
39	tfec	ENSONIT00000021193	-2.44	Transcription factor EC
40	mfi2	ENSONIT00000011253	-2.42	Antigen p97 (melanoma associated) identified by monoclonal antibodies 133.2 and 96.5
41	CRD2	ENSONIT00000015173	-2.35	cAMP-regulated D2 protein-like isoform X1
42	rab27b	ENSONIT00000016030	-2.35	RAB27B, member RAS oncogene family
43	nkx6.2	ENSONIT00000009452	-2.34	NK6 homeobox 2
44	rab27b	ENSONIT00000021335	-2.33	RAB27B, member RAS oncogene family
45	b3gnt2a	ENSONIT00000026370	-2.33	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 2a
46	rxfp2a	ENSONIT00000015985	-2.30	Relaxin/insulin-like family peptide receptor 2a
47	retsat	ENSONIT00000007947	-2.30	putative all-trans-retinol 13,14-reductase
48	unknown	ENSONIT00000020907	-2.25	uncharacterized
49	tmem235	ENSONIT00000025015	-2.25	Transmembrane protein 235
50	hsd3b1	ENSONIT00000019776	-2.25	Hydroxy-delta-5-steroid dehydrogenase, 3 beta- and steroid delta-isomerase 1
51	FIBCD1	ENSONIT00000001032	-2.24	Fibrinogen C domain containing 1
52	RDH11	ENSONIT00000018230	-2.23	retinol dehydrogenase 11-like
53	pth1rb	ENSONIT00000014585	-2.20	Parathyroid hormone 1 receptor b
54	DHRS12	ENSONIT00000012634	-2.20	dehydrogenase/reductase SDR family member 12-like
55	slc2a9	ENSONIT00000025689	-2.19	Solute carrier family 2 (facilitated glucose transporter), member 9
56	mc1r	ENSONIT00000026719	-2.16	Melanocortin 1 receptor
57	unknown	ENSONIT00000015686	-2.09	uncharacterized
58	syngr1b	ENSONIT00000011606	-2.08	Synaptogyrin 1b
59	sec14l2	ENSONIT00000017719	-2.06	SEC14-like protein 2
60	ACSL6	ENSONIT00000021738	-2.05	ENSONIT00000021738
61	rab34b	ENSONIT00000006941	-2.04	RAB34, member RAS oncogene family b
62	zdhhc14	ENSONIT00000023686	-2.04	Zinc finger, DHHC-type containing 14
63	sox9a	ENSONIT00000024737	-2.04	ENSONIT00000024737
64	trpm1a	ENSONIT00000004083	-2.03	Transient receptor potential cation channel, subfamily M, member 1a
65	PLCB2	ENSONIT00000015359	-2.00	ENSONIT00000015359
66	bscl2l	ENSONIT00000006721	-2.00	Bernardinelli-Seip congenital lipodystrophy 2, like
67	tpd52	ENSONIT00000003289	-1.99	Tumor protein D52
68	prkg1b	ENSONIT00000024074	-1.98	Protein kinase, cGMP-dependent, type Ib
69	egfl6	ENSONIT00000020235	-1.96	EGF-like-domain, multiple 6
70	tmem130	ENSONIT00000024971	-1.96	Transmembrane protein 130
71	cecr1b	ENSONIT00000000025	-1.96	Cat eye syndrome chromosome region, candidate 1b

72	unknown	ENSONIT00000026203	-1.96	uncharacterized
73	and2	ENSONIT00000023476	-1.95	Actinodin2
74	gatsl2	ENSONIT00000016736	-1.94	ENSONIT00000016736
75	prickle2a	ENSONIT00000021803	-1.94	Prickle homolog 2a
76	cxxc4	ENSONIT00000022134	-1.93	CXXC finger 4
77	KIF5B	ENSONIT00000004756	-1.93	Kinesin family member 5B
78	URH1	ENSONIT00000017911	-1.91	probable uridine nucleosidase 1 isoform X2
79	myo5aa	ENSONIT00000003312	-1.90	Myosin VAa
80	FAM180A	ENSONIT00000013027	-1.90	Family with sequence similarity 180 member A
81	alk	ENSONIT00000024613	-1.89	Anaplastic lymphoma receptor tyrosine kinase
82	mlphb	ENSONIT00000021146	-1.88	Melanophilin b
83	gpct	ENSONIT00000010244	-1.88	Glutaminyl-peptide cyclotransferase
84	desi1a	ENSONIT00000019699	-1.87	Desumoylating isopeptidase 1a
85	slc2a5	ENSONIT00000023711	-1.86	solute carrier family 2, facilitated glucose transporter member 5-like
86	pcbp3	ENSONIT00000013044	-1.86	poly(rC)-binding protein 3-like
87	slc7a10a	ENSONIT00000003926	-1.85	Solute carrier family 7 (neutral amino acid transporter light chain, asc system), member 10a
88	ST8SIA3	ENSONIT00000021336	-1.84	sia-alpha-2,3-Gal-beta-1,4-GlcNAc-R:alpha 2,8-sialyltransferase-like
89	hmx4	ENSONIT00000004437	-1.83	H6 family homeobox 4
90	paics	ENSONIT00000003650	-1.81	Phosphoribosylaminoimidazole carboxylase, phosphoribosylaminoimidazole succinocarboxamide synthetase
91	tmem182a	ENSONIT00000017894	-1.81	Transmembrane protein 182a
92	RALYL	ENSONIT00000023479	-1.80	RALY RNA binding protein-like
93	calb2a	ENSONIT00000014036	-1.80	Calbindin 2a
94	gch2	ENSONIT00000020050	-1.79	GTP cyclohydrolase 2
95	arhgef33	ENSONIT00000001492	-1.77	Rho guanine nucleotide exchange factor (GEF) 33
96	adra2b	ENSONIT00000025759	-1.76	Adrenoceptor alpha 2B
97	got1l1	ENSONIT00000014700	-1.76	Glutamic-oxaloacetic transaminase 1-like 1
98	ttyh2l	ENSONIT00000006814	-1.76	Tweety homolog 2, like
99	rasd1	ENSONIT00000024743	-1.75	RAS, dexamethasone-induced 1
100	ndrg4	ENSONIT00000013647	-1.73	NDRG family member 4
101	rgmb	ENSONIT00000017244	-1.70	Repulsive guidance molecule family member b
102	slc2a11b	ENSONIT00000017078	-1.69	Solute carrier family 2 (facilitated glucose transporter), member 11b
103	tyrp1b	ENSONIT00000011131	-1.67	Tyrosinase-related protein 1b
104	myoc	ENSONIT00000023494	-1.65	Myocilin, trabecular meshwork inducible glucocorticoid response
105	slc45a2	ENSONIT00000009598	-1.63	Solute carrier family 45, member 2
106	OTOP1	ENSONIT00000025688	-1.62	Otopetrin 1
107	kcnk4	ENSONIT00000006787	-1.61	Potassium channel, two pore domain subfamily K, member 4

108	tpd52l	ENSONIT00000000849	-1.61	Tumor protein D52-like 1
109	GLP2R	ENSONIT00000024649	-1.61	Glucagon-like peptide 2 receptor
110	DENND3	ENSONIT00000007554	-1.60	DENN/MADD domain containing 3
111	FAM19A1	ENSONIT00000021382	-1.60	Family with sequence similarity 19 (chemokine (C-C motif)-like), member A1
112	trpm1b	ENSONIT00000019285	-1.59	Transient receptor potential cation channel, subfamily M, member 1b
113	slc38a11	ENSONIT00000011168	-1.59	Solute carrier family 38, member 11
114	dhrsx	ENSONIT00000009275	-1.58	Dehydrogenase/reductase (SDR family) X-linked
115	impdh1b	ENSONIT00000014451	-1.52	IMP (inosine 5,-monophosphate) dehydrogenase 1b
116	PCDH10	ENSONIT00000004385	-1.52	Protocadherin 10
117	pcolce2b	ENSONIT00000007438	-1.52	Procollagen C-endopeptidase enhancer 2b
118	GPR21	ENSONIT00000004974	-1.51	probable G-protein coupled receptor 21
119	map7a	ENSONIT00000000906	-1.51	Microtubule-associated protein 7a
120	mtnr1aa	ENSONIT00000012798	-1.51	Melatonin receptor 1A a
121	GPR21	ENSONIT00000025721	-1.50	probable G-protein coupled receptor 21
122	igsf11	ENSONIT00000021051	-1.48	Immunoglobulin superfamily member 11
123	mab21l1	ENSONIT00000026730	-1.46	Mab-21-like 1
124	slc25a38a	ENSONIT00000010283	-1.44	Solute carrier family 25, member 38a
125	tubb5	ENSONIT00000016436	-1.44	Tubulin, beta 5
126	rgs6	ENSONIT00000007304	-1.44	Regulator of G-protein signaling 6
127	PRLR	ENSONIT00000004597	-1.43	prolactin receptor-like
128	opn5	ENSONIT00000014212	-1.43	Opsin 5
129	slc25a36b	ENSONIT00000018418	-1.42	Solute carrier family 25 (pyrimidine nucleotide carrier), member 36b
130	VEGFC	ENSONIT00000002365	-1.42	vascular endothelial growth factor C-like
131	npdc1b	ENSONIT00000017456	-1.42	Neural proliferation, differentiation and control, 1b
132	ap1s2	ENSONIT00000003971	-1.41	Adaptor-related protein complex 1, sigma 2 subunit
133	mc4r	ENSONIT00000025784	-1.40	Melanocortin 4 receptor
134	sypb	ENSONIT00000021213	-1.38	Synaptophysin b
135	scarb1	ENSONIT00000014440	-1.38	Scavenger receptor class B, member 1
136	znf385b	ENSONIT00000008702	-1.38	Zinc finger protein 385B
137	rasef	ENSONIT00000018926	-1.37	RAS and EF-hand domain containing
138	prkar1b	ENSONIT00000006488	-1.37	Protein kinase, cAMP-dependent, regulatory, type I, beta
139	drd3	ENSONIT00000017700	-1.37	Dopamine receptor D3
140	tgm1l2	ENSONIT00000003616	-1.35	Transglutaminase 1 like 2
141	akr1b1	ENSONIT00000021207	-1.34	Aldo-keto reductase family 1, member B1 (aldose reductase)
142	bhlhe41	ENSONIT00000018536	-1.32	Basic helix-loop-helix family, member e41
143	slc35f3b	ENSONIT00000012312	-1.31	Solute carrier family 35, member F3b
144	opn4a	ENSONIT00000008980	-1.29	Opsin 4a (melanopsin)

145	urah	ENSONIT00000007140	-1.28	Urate (5-hydroxyiso-) hydrolase
146	b3glctb	ENSONIT00000015990	-1.26	Beta 3-glucosyltransferase b
147	sgtb	ENSONIT00000013978	-1.26	Small glutamine-rich tetratricopeptide repeat (TPR)-containing, beta
148	cacna1hb	ENSONIT00000004185	-1.25	Calcium channel, voltage-dependent, T type, alpha 1H subunit b
149	COL10A1	ENSONIT00000014289	-1.25	Collagen, type X, alpha 1
150	KIF5B	ENSONIT00000004755	-1.25	Kinesin family member 5B
151	C9orf172	ENSONIT00000013714	-1.25	Chromosome 9 open reading frame 172
152	tnfrsfa	ENSONIT00000013989	-1.24	Tumor necrosis factor receptor superfamily, member a
153	cx43	ENSONIT00000026288	-1.24	Connexin 43
154	agtrap	ENSONIT00000000772	-1.24	Angiotensin II receptor-associated protein
155	mfsd9	ENSONIT00000017896	-1.24	Major facilitator superfamily domain containing 9
156	lrrc8aa	ENSONIT00000009316	-1.24	Leucine rich repeat containing 8 family, member Aa
157	rnd1	ENSONIT00000021522	-1.23	Rho family GTPase 1
158	VASH1	ENSONIT00000015141	-1.22	Vasohibin 1
159	amer2	ENSONIT00000018253	-1.22	APC membrane recruitment protein 2
160	LRRC72	ENSONIT00000025031	-1.22	leucine-rich repeat-containing protein 72-like
161	VANGL2	ENSONIT00000023596	-1.21	VANGL planar cell polarity protein 2
162	MFSD1	ENSONIT00000015744	-1.20	major facilitator superfamily domain-containing protein 8-like isoform X1
163	ppfia2	ENSONIT00000000095	-1.19	Protein tyrosine phosphatase, receptor type, f polypeptide (PTPRF), interacting protein (liprin), alpha 2
164	glula	ENSONIT00000024224	-1.19	Glutamate-ammonia ligase (glutamine synthase) a
165	pax7a	ENSONIT00000017970	-1.18	Paired box 7a
166	ckba	ENSONIT00000015656	-1.18	Creatine kinase, brain a
167	DCTN6	ENSONIT00000009520	-1.18	Dynactin subunit 6
168	slc25a48	ENSONIT00000005395	-1.18	Solute carrier family 25, member 48
169	KIF1A	ENSONIT00000012417	-1.15	Kinesin family member 1A
170	stc2a	ENSONIT00000022311	-1.15	Stanniocalcin 2a
171	tgm5l	ENSONIT00000008627	-1.13	Transglutaminase 5, like
172	tfap2e	ENSONIT00000017161	-1.13	Transcription factor AP-2 epsilon
173	mtnr1c	ENSONIT00000022473	-1.12	Melatonin receptor 1C
174	tmem189	ENSONIT00000021135	-1.11	Transmembrane protein 189
175	cdk15	ENSONIT00000011800	-1.10	Cyclin-dependent kinase 15
176	tm6sf1	ENSONIT00000000288	-1.10	Transmembrane 6 superfamily member 1
177	sytl2b	ENSONIT00000014048	-1.10	Synaptotagmin-like 2b
178	MUC2	ENSONIT00000006431	-1.10	mucin-2-like
179	myo16	ENSONIT00000018074	-1.09	Myosin XVI
180	cenpl	ENSONIT00000015263	-1.09	Centromere protein L
181	viml	ENSONIT00000011636	-1.08	ENSONIT00000011636

182	fzd3a	ENSONIT00000001556	-1.08	Frizzled class receptor 3a
183	PDE7B	ENSONIT00000000607	-1.08	Phosphodiesterase 7B
184	FAM135B	ENSONIT00000009646	-1.08	Family with sequence similarity 135 member B
185	adgrb3	ENSONIT00000012277	-1.07	Adhesion G protein-coupled receptor B3
186	gpr143	ENSONIT00000020200	-1.07	G protein-coupled receptor 143
187	tspan9a	ENSONIT00000013039	-1.07	Tetraspanin 9a
188	zdhhc2	ENSONIT00000022221	-1.06	Zinc finger, DHHC-type containing 2
189	kank3	ENSONIT00000015998	-1.06	KN motif and ankyrin repeat domains 3
190	zeb2a	ENSONIT00000007321	-1.04	Zinc finger E-box binding homeobox 2a
191	GNAZ	ENSONIT00000004571	-1.04	Guanine nucleotide binding protein (G protein), alpha z polypeptide
192	GNG5	ENSONIT00000007888	-1.02	guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-5-like
193	neurl1b	ENSONIT00000022470	-1.02	Neuralized E3 ubiquitin protein ligase 1B
194	megf10	ENSONIT00000014149	-1.01	Multiple EGF-like-domains 10
195	DGKD	ENSONIT00000021768	-1.01	Diacylglycerol kinase delta
196	mgat4b	ENSONIT00000005466	-1.01	Mannosyl (alpha-1,3-)-glycoprotein beta-1,4-N-acetylglucosaminyltransferase, isozyme B
197	tgm5l	ENSONIT00000008626	-1.01	Transglutaminase 5, like
198	cx45.6	ENSONIT00000025996	-0.99	Connexin 45.6
199	gnaq	ENSONIT00000018791	-0.99	Guanine nucleotide binding protein (G protein), q polypeptide
200	soul3	ENSONIT00000004455	-0.99	Heme-binding protein soul3
201	acsl3a	ENSONIT00000018452	-0.98	Acyl-CoA synthetase long-chain family member 3a
202	pax7	ENSONIT00000002000	-0.98	Paired box 7
203	stk17a	ENSONIT00000019057	-0.98	Serine/threonine kinase 17a
204	LONRF2	ENSONIT00000021480	-0.97	LON peptidase N-terminal domain and RING finger protein 2
205	EDNRB	ENSONIT00000022254	-0.97	endothelin B receptor-like
206	scgn	ENSONIT00000009096	-0.96	Secretagogin, EF-hand calcium binding protein
207	coro2ba	ENSONIT00000007131	-0.96	Coronin, actin binding protein, 2Ba
208	GGA2	ENSONIT00000004876	-0.95	Golgi-associated, gamma adaptin ear containing, ARF binding protein 2
209	hs3st1l2	ENSONIT00000017333	-0.95	Heparan sulfate (glucosamine) 3-O-sulfotransferase 1-like 2
210	gnao1a	ENSONIT00000012832	-0.95	Guanine nucleotide binding protein (G protein), alpha activating activity polypeptide O, a
211	SHC4	ENSONIT00000016796	-0.95	SHC (Src homology 2 domain containing) family member 4
212	KIAA2026	ENSONIT00000017688	-0.94	KIAA2026
213	oca2	ENSONIT00000006160	-0.94	Oculocutaneous albinism II
214	vcla	ENSONIT00000012125	-0.93	Vinculin a
215	mitfb	ENSONIT00000021398	-0.93	Microphthalmia-associated transcription factor b
216	ednrba	ENSONIT00000023565	-0.92	Endothelin receptor Ba
217	dusp6	ENSONIT00000004893	-0.92	Dual specificity phosphatase 6
218	atp1b3a	ENSONIT00000013367	-0.92	ENSONIT00000013367

219	adcy5	ENSONIT00000021495	-0.90	Adenylate cyclase 5
220	cyth3b	ENSONIT00000021647	-0.90	Cytohesin 3b
221	GIMAP4	ENSONIT00000026771	-0.90	GTPase IMAP family member 4-like
222	sox10	ENSONIT00000022743	-0.89	SRY (sex determining region Y)-box 10
223	chn1	ENSONIT00000010864	-0.89	Chimerin 1
224	ITGA7	ENSONIT00000006575	-0.88	Integrin subunit alpha 7
225	SLC23a2	ENSONIT00000018153	-0.87	Solute carrier family 23 (ascorbic acid transporter), member 2
226	gpr26	ENSONIT00000000673	-0.87	G protein-coupled receptor 26
227	CKAP2	ENSONIT00000014262	-0.86	cytoskeleton-associated protein 2-like
228	crispld2	ENSONIT00000007435	-0.86	Cysteine-rich secretory protein LCCL domain containing 2
229	sacs	ENSONIT00000006843	-0.85	Sacsin molecular chaperone
230	fosb	ENSONIT00000004670	-0.84	FBJ murine osteosarcoma viral oncogene homolog B
231	rabgap1l	ENSONIT00000017800	-0.84	RAB GTPase activating protein 1-like
232	hcst	ENSONIT00000003510	-0.84	Hematopoietic cell signal transducer
233	cspg4	ENSONIT00000016195	-0.83	Chondroitin sulfate proteoglycan 4
234	pip5k1bb	ENSONIT00000020345	-0.83	Phosphatidylinositol-4-phosphate 5-kinase, type I, beta b
235	fscn1b	ENSONIT00000000229	-0.83	Fascin actin-bundling protein 1b
236	mapre3a	ENSONIT00000001359	-0.82	Microtubule-associated protein, RP/EB family, member 3a
237	PPT2	ENSONIT00000002782	-0.81	lysosomal thioesterase PPT2-like
238	csrp2	ENSONIT00000015200	-0.81	Cysteine and glycine-rich protein 2
239	cdc42ep4a	ENSONIT00000026714	-0.80	CDC42 effector protein (Rho GTPase binding) 4a
240	kirrel3l	ENSONIT00000003492	-0.79	Kin of IRRE like 3 like
241	alx4a	ENSONIT00000010641	-0.79	ALX homeobox 4a
242	mef2aa	ENSONIT00000019052	-0.78	Myocyte enhancer factor 2aa
243	trpv1	ENSONIT00000016129	-0.78	Transient receptor potential cation channel, subfamily V, member 1
244	fam150bb	ENSONIT00000023632	-0.76	Family with sequence similarity 150, member Bb
245	PCDH9	ENSONIT00000004068	-0.75	Protocadherin 9
246	hdac7b	ENSONIT00000024411	-0.74	Histone deacetylase 7b
247	foxd3	ENSONIT00000026741	-0.73	Forkhead box D3
248	trim2b	ENSONIT00000017770	-0.73	Tripartite motif containing 2b
249	fsd1	ENSONIT00000003202	-0.73	Fibronectin type III and SPRY domain containing 1
250	sncga	ENSONIT00000008847	-0.73	Synuclein, gamma a
251	syngr3b	ENSONIT00000019720	-0.71	Synaptogyrin 3b
252	sox10	ENSONIT00000010558	-0.70	SRY (sex determining region Y)-box 10
253	prtfdc1	ENSONIT00000008139	-0.70	Phosphoribosyl transferase domain containing 1
254	LFNG	ENSONIT00000000215	-0.70	beta-1,3-N-acetylglucosaminyltransferase lunatic fringe-like
255	marcksl1b	ENSONIT00000009744	-0.69	MARCKS-like 1b

256	rara	ENSONIT00000007968	-0.69	Retinoic acid receptor, alpha
257	afap111b	ENSONIT00000010822	-0.68	Actin filament associated protein 1-like 1b
258	HSPA12A	ENSONIT00000014654	-0.68	heat shock 70 kDa protein 12A-like isoform X1
259	hunk	ENSONIT00000005903	-0.63	Hormonally up-regulated Neu-associated kinase
260	CAPN1	ENSONIT00000010412	-0.60	calpain-1 catalytic subunit-like, partial
261	sypl1	ENSONIT00000021285	-0.60	Synaptophysin-like 1
262	ano5a	ENSONIT00000010429	-0.59	Anoctamin 5a
263	spsb3b	ENSONIT00000025238	-0.59	SplA/ryanodine **receptor** domain and SOCS box containing 3b
264	TNS4	ENSONIT00000011328	0.55	tensin 4
265	steap4	ENSONIT00000007508	0.61	STEAP family member 4
266	Rpf	ENSONIT00000018333	0.82	resuscitation-promoting factor RpfA-like
267	TRBC1	ENSONIT00000000470	0.95	T cell receptor beta constant 1
268	HMX1	ENSONIT00000002272	0.97	H6 family homeobox 1
269	col2a1b	ENSONIT00000006473	1.10	collagen, type II, alpha 1b
270	CD59	ENSONIT00000004672	1.49	CD59 glycoprotein-like
271	keratin	ENSONIT00000023704	1.96	keratin, type I cytoskeletal 13-like
272	col8a1b	ENSONIT00000023612	2.17	Collagen, type XII, alpha 1b
273	CLEC19A	ENSONIT00000012084	2.47	C-type lectin domain family 19 member A
274	CLEC3A	ENSONIT00000007521	2.75	C-type lectin domain family 3 member A

1. *Candidate genes for the anal fin blotch formation in C. macrops*

Based on GO term enrichment analyses and the inspection of published literature, among these 274 candidate genes, 29 genes were found to be related to "pigmentation", including gene *ednrba*, which is expressed by precursors of all three classes of pigment cells [53], genes related to melanophores (*mc1r* [54], *gpr143* [55], *oca2* [56], *sytl2* [57], *dhrsx* [58], *cspg4* [59], *igsf11* [60], *mtnr1aa* [61], *gnpmb* [62], *trpm1a* [63], *trpm1b* [63], *pmelb* [55], *tyrp1b* [63], *RAS family* [64], *slc45a2* [65], *igsf11* [66], *mitfb* [67]), iridohpores (*slc25a38a* [65], *slc45a2* [65], *pnp4a* [68], *tpd52*[65], *tmem* [65], *ifi30* [65], *gpnmb* [65], *tagln3b* [65]) and xanthophores (*pax7* [69], *gch2* [70]). Ten genes are related to "neural crest cell", a multi-potent cell lineage from which various progenitor cells migrate to develop a variety of structures and tissues including pigment cells [71]. Other neural crest cell related genes such as *fhl2a* [32], *fhl2b* [32], *tfap2e* [72], *foxd3* [68], *cax2* [73], *fscn1a* [74], *crip2* [75] were also found in our list. These candidate genes further prove the robustness of our experiment design, and will be useful for further study of anal fin pigmentation pattern formation.

Except cell-cell interactions, the tissue environment might also be important for pattern formation. For example, thyroid hormones can control pattern development as a global factor interacting with local cells [76]. Interestingly, the parathyroid hormone receptor activity gene *pth1rb* and the thyroid hormone related genes *urah* [77] and *mc4r* [78] were included in our list and it will be interesting to see whether they have similar role in pattern formation as thyroid hormone. Other hormone-related genes such as androgen related genes (*gpnmb* [79], *fhl2a* [80], *fhl2b* [80], *apod* [81]), a growth hormone-related gene (*ghra* [82]) and a melatonin hormone related gene (*mtnr1aa* [83]) were also found in our list. Besides, chemical signalling molecules can also play a role in driving pattern formation. For example, damage experiments in butterfly eyespots found that spontaneous low-frequency Ca^{2+} waves in vivo was involved in the wing development [84]. Genes related to other molecules such as H^+, Na^+ and K^+ (*atp1b3a* [85], *cax2* [73]) were among the candidate genes retrieved from our experiment.

The ability of a morphogen to affect pattern formation depends on cellular-specific responding. Therefore, cell-surface receptors may have important functions in intracellular signalling. For example, integrins as signaling receptors have a function in sending signals to the cell in response to the extracellular environment by binding to the ligands including vitronectin and collagen [86]. Genes related to integrin, vitronectin and collagen were also retrieved from our RNAseq experiment (*vtna, COL10A1, ITGA7, col2a1b, col8a1b*) (Table 2). Besides, membrane proteins such as G protein–coupled receptors (GPCRs), which can detect signals, such as hormones,

ions, small organic molecules and entire proteins are also important [87]. Many GPCRs were found in our candidate gene list, such as *opn4a* [88], *mtnraa* [89] and *adra2b* [90]. In addition, except these signaling receptors located on the cell surface, gap junction channels located between neighbor cells for intercellular communication are also important for pattern formation [91]. For example, mutants of candidate genes *connexin41.8* [60] and *connexion 39.4* [92] related to gap junctions can form spots instead of stripes in zebrafish.

Several pathways were enriched significantly for these differential expressed genes (Fig.3). A large portion of genes were involved in the GO category "metabolic pathways" (Fig. 3). As mentioned above, anal fin pigmentation patterns could play a role in sexual selection, and if the development of this trait is costly, it could be a "fitness indicator", because individuals have to balance the costs between fitness and survival. Purine metabolism and melanogenesis pathways were also enriched (Fig. 3). This is not unexpected since both of them are important for pigment pattern formation [93,94]. Wnt signaling pathway was also enriched, which is well known as a morphogen to activate signals of neighbouring cells or tissues to control pattern formation [95,96].

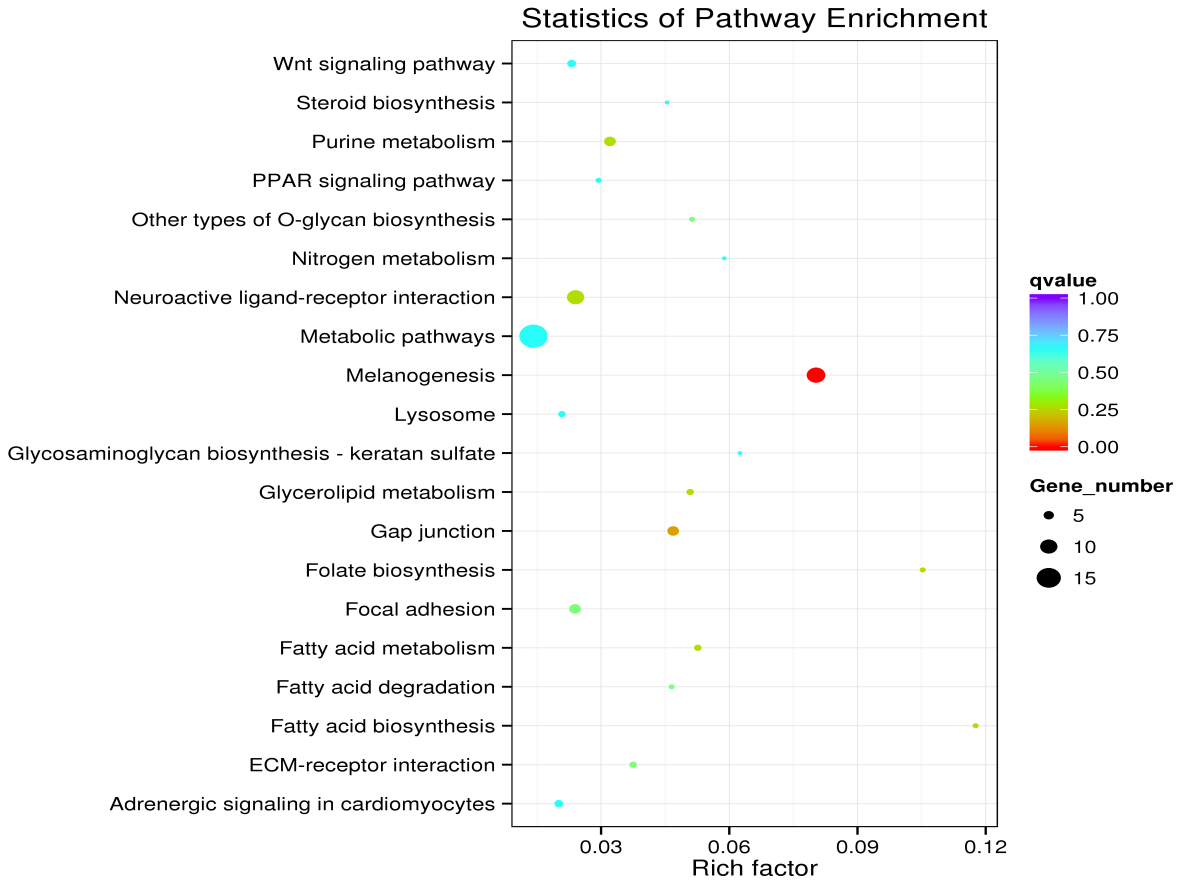

Fig. 3 KEGG pathways enrichment of differential expressed genes of the blotch tissue in cichlid fish *C. macrops*. Rich factor represents the ratio of the number of differential expressed genes and the number of all genes annotated in the pathway. Q-value is the normalization of the *p* value. Here we displayed the top 20 significantly differential expressed genes enriched pathways.

2. The gene network underlying the anal fin blotch in C. macrops

In addition to genes with known functions in pigmentation pattern, we also found candidate genes in our list with other functions, such as immunity, circadian rhythm, and posterior lateral line development (Table 2). For example, candidate gene *igsf11* belongs to immunoglobulin superfamily, but it can also mediate adhesive interactions between melanophores and neighbouring cells for stripe pattern formation in zebrafish [60]. It has previously been shown that the environment can have a large effect on pigmentation pattern formation [16,60]. For example, fish not exposed to light are often colourless and when fish is affected by diseases, it often appears pale [16]. Correlation between nuptial colours and visual sensitivities tuned by opsin was found in cichlid fish [97]. Also, the posterior lateral line system in fish can detect the pattern of water movement, and might contribute to feeding, schooling behavior, and prey detection [97]. Why are they also highly expressed in the blotch tissue, and what are their roles in this phenotype? 1) One possible explanation is that they could be genes with a pleiotropic effect, i.e. also involved in the pigmentation pathway. 2) Another possibility is that these genes were recruited into a novel gene network responsible for anal fin pigmentation pattern. If these genes interact with pigmentation genes and are integrated into blotch gene network, then the selection on these traits might have an indirect response to the blotch pigmentation pattern. Indeed, it has been suggested that pigment patterns may evolve through direct selection on the patterns themselves, and also as correlated responses to selection on other traits [98].

The question remaining is how are the genes responsible for the ectodine blotch recruited. Based on protein-protein interaction analysis (Fig.4), G-protein related genes were found to rank as the top candidate genes with most interactions in the network (Supplementary Table 3). Examples for such genes are *gnai1* and *gnao1a*, which are part of the Wnt signaling pathway [99], and *gnaq* related to pigmentation [63]. Other pigmentation related genes such as *myo5aa* [100], *ednrb1a* [101], *sox10* [102], *mitfb* [68], *gch2* [64], somite development related TF *mef2aa* [103] also interacted with most genes (Supplementary Table 3). In addition, TFs *foxd3* [68], *sox10* [102], *mitfb* [67], *pax7a* [69] were found to interact with each other directly, and were connected to genes which are most related to melanogenesis pathways (Fig. 4). Indeed, *trans*-regulatory landscape has been shown to play a key role in the emergence and modification of novelty, such as the role of *Dll* in wing spot emergence in fly [104]. It will be interesting to see the roles of the TFs derived from our study in the blotch gene network. Noticeably, for 136 genes there were no interactions indicated with other genes (Table 2; Fig.4). This could be due to the incompleteness of the string database (http://string-db.org/). Another explanation is that we used zebrafish as reference, so that it might be

impossible to detect a gene network specific to the anal fin blotch, which do not exist in zebrafish. Other genes such as *ednr1b, myo5aa, gnai1, impdh1b, fzd7a* were also located at the nodes connecting many other genes (Fig. 4; Supplementary Table 3) whose function should also be studied in the future. Besides, there are three uncharacterized candidate genes (*20907, 15686, 26203*) (Table 2), which could be cichlid specific genes. It has previously been shown that lineage specific genes might play a role in the emergence of eggspots [6].

Besides, does the gene network of the ectodine blotch evolve from a pre-existing network or originate *de novo*? The extant of the independence of a core gene network is a key factor for phenotypic evolvability [25], for example, the "Christmas tree model" [26] as mentioned above, which can limit the evolvability of wing spot in fly [26]; and the flexible evolvability of eyespot with different numbers, sizes and colours in butterfly is because of their independent gene network [25]. In our study, at least six candidate genes are likely involved in the pre-existing fin gene regulatory landscape, including genes related to fin regeneration and development (*st8, raraa, dusp6, dhrsx, and2*) [58,105–108] and blood vessel development (*vegfc*) [109]. What are the roles of our candidate genes in the ancestral fin gene network and whether they belong to the core gene network in the innovative blotch needs further investigation.

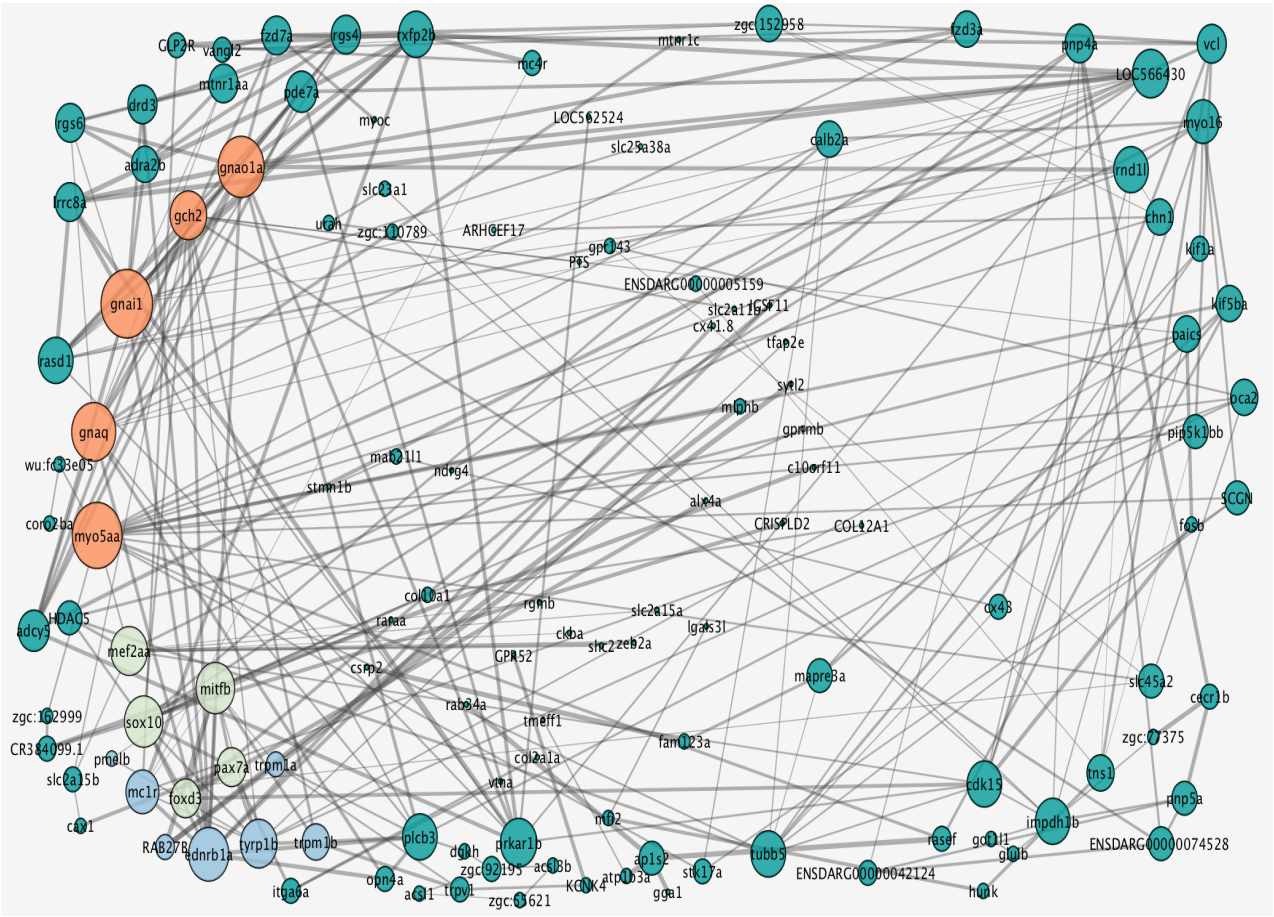

Fig. 4 Protein-protein interaction network The blotch related candidate gene network constructed based on the Protein–Protein Interaction data from the STRING database. Orange round circle represents the top rank genes with most interactions, light green circle represents transcripts interacted each other, light blue circle represents genes related to melanogenesis connecting directly with transcriptions.

3. Genetic basis of the convergent evolution of the ectodine blotch and haplochromine eggspots

By comparing 274 ectodine blotches related candidate genes with 812 eggspots related candidate genes identified by our study, 15.7% (43/274) genes showed same expression pattern between blotch and eggspots (Table 2), suggesting that a common genetic basis at least in part are shared between the ectodine blotch and haplochromine eggspots, which is in contrast with the conclusion from Santos et al, [6] (Chapter 2). The possible explanation of this difference could be that we used the whole gene exrepssion profile derived from comparative transcriptomic data of the ectodine blotch instead of using several individual candidate genes. Besides, compare to eggspots which have different numbers, the ectodine blotch is much easier to design experiment to control position effect to get candidate genes mostly possible related to blotch itself.

These shared 43 genes were related to ligand transporter (*APOD*) [110,111], neural crest cell (*fhl2b, fhl2a*) ([32]), vitronectin (*vtna*) [112], iridohpore (*fhl2b, pnp4a*) [32], melanophore (*gpnmb* [62], *trpm1a* [113], *trpm1b* [113], *mlphb* [114]), chemical signal (*cax2* [73]), connexin (*cx45.6* [115]), and one unannotated gene (*26203*). All these functions might be important for the pigmentation pattern formation as we mentioned above. Besides, the common shared TFs are found related to MiT family (*mitfb*, *tfec*) [67] and xanthopore *pax7* [116]. One of MiT family member *mitf* was related to melanocyte development [117]. Except the same expression pattern in these 43 common shared genes, two shared genes showed different expression pattern, with highly expression in eggspots, but down-regulated in the blotch. One is gene *col8a1b* related to collagen. Collagen has a role as extracellular signals which might affect pattern formation [118]. Another gene is *steap4* which was shown to be related to immunity [119]. Besides, 15 genes showed highly expression in the blotch but were down-regulated in eggspots (Table 2). Noticeably, duplicated genes *pnp4a* and *pnp5a* showed different expression patterns, with *pnp4a* was highly expressed in eggspots, but *pnp5a* was highly expressed in blotch. It has been suggested that iridophore might be the pre-requisite for zebrafish stripe pattern formation [16], and previously, it was shown that an eggspots specific *cis*-regulatory element in the upstream of gene *fhl2b* related to iridophore might be causally linked to eggspots formation [32]. In addition, duplicated genes *pax7a* and *pax7* also showed different expression patterns, with xanthophore related gene *pax7* highly expressed in eggspots and blotch, but *pax7a* was highly specific expressed in blotch (Table 2). Two connexin relatd genes also showed opposite expression pattern in eggspots and the blotch, with *cx 45.6* highly expressed in eggspots, but *cx43* highly expressed in the blotch. It has been shown that mutant of connexin genes can affect stripe pattern formation in zebrafish [115]. Besides, homeobox related TF *hmx4*, and TF *fosb* also showed opposite expression pattern in these two convergent anal fin

pigmentation patterns. What are the roles of these shared genes but with differential expression patterns in these two different novelties will be interesting to do further investigation.

Based on these results, two hypothesis are proposed for the evolution of these common shared and unshared genes: 1) The gene network formed by the shared genes might mean that eggspots and ectodine blotches are homologous at a deep level [23,24]; and the unshared genes are co-opted independently in two lineages. And if these shared genes are responsible for the core gene network of anal fin pigmentation pattern formation, then the common ancestor of haplochromine and ectodine lineages might have innovative anal fin pigmentation, but lost secondarily in most species in ectodine lineage. 2) It also could be that the shared genes were independently co-opted into two different gene networks in two lineages. The roles of the shared and unshared genes in these two anal fin pigmentation patterns could be 1) the common shared genes are responsible for the basic pigmentation pattern formation, such as the pigmentation related genes as we mentioned above. While the shared genes but with opposite gene expression patterns or the unshared genes might be responsible for the differences between eggspots and the blotch. 2) The shared genes could also be responsible for the differences between eggspots and the blotch, since we did not test different developmental points here. For example, iridophore has been suggested to be important for pre-pattern formation in zebrafish stripe formation and the appearance of iridophore during different developmental stages could affect stripe pattern formation [16]. 3) Besides, the unshared genes could also be responsible for the common basic pattern formation, for example, if they are different effector genes but with same phenotypic results. The roles of these shared and unshared genes for the emergence and modification in the convergent evolution of these two anal fin pigmentation patterns needs further investigation.

Conclusion

By using a relatively comprehensive comparative transcriptomic experiment design, we identified 274 candidate genes for the morphogenesis of the anal fin blotch in an ectodine species, *C. macrops*. Further re-analysis using existing transcriptomic data from Santos et al. [6] (Chapter 2), we identified 812 eggspots related candidate genes [6]. Gene expression profile comparison showed that 15.7% (43/274) genes exhibited similar expression pattern between these two convergent novelties, suggesting that these two pigmentation patterns could share at least parts of a common gene network. Besides, two genes (*col8a1b* and *steap4*) were highly expressed in eggspots, but down-regulated in blotch, and 15 genes showed highly expression in the blotch, but were down-regulated in eggspots. Candidate genes identified by our study will be useful for further

analysis of the genetic basis of the innovative anal fin pigmentation patterns in cichlid fish in the future.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LG and WS designed the experiment and wrote the manuscript. LG performed the RNAseq library construction of the ectodine blotch in *C. macrops*. MS provided the raw transcriptomic data of eggspots in *A. burtoni*. LG performed all the data analysis. All authors read and approved the manuscript.

Acknowledgements

We would like to thank Astrid Böhne, Tingting Zhou, Jie Zhang, Shengkai Pan and Madlen Stange for the discussion; Philippe Demougin and Ina Nissen for the assistance with Illumina sequencing. We also thank Nicolas Boileau for assistance in the lab and Brigitte Aeschbach for the laboratory aparatus organization. This project was supported by the University of Basel, Switzerland and the European Research Council (ERC).

References

1. Wickler W. "Egg-dummies" as Natural Releasers in Mouth-breeding Cichlids. Nature. 1962; 194:1092–3.

2. Theis A, Salzburger W, Egger B. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. PLoS One. 2012; 7:e29878.

3. Theis A, Bosia T, Roth T, Salzburger W, Egger B. Egg-spot pattern and body size asymmetries influence male aggression in haplochromine cichlid fishes. Behav. Ecol. 2015; arv104.

4. Couldridge VCK. Experimental manipulation of male eggspots demonstrates female preference for one large spot in Pseudotvopheus lombavdoi. J. Fish Biol. 2002; 60:726–30.

5. Fryer G, Iles T. The cichlid fihses: their biology and evolution. edinburgh, uk. 1972;

6. Santos ME, Baldo L, Gu L, Boileau N, Musilova Z, Salzburger W. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. BMC Genomics. 2016. 17: 712.

7. Müller GB, Wagner GP. Novelty in Evolution: Restructuring the Concept. Annu. Rev. Ecol. Syst. 1991; 22:229–56.

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

9. Monteiro A. Origin, development, and evolution of butterfly eyespots. Annu. Rev. Entomol. 2015; 60:253–71.
10. Shimeld SM, Holland PWH. Vertebrate innovations. Proc. Natl. Acad. Sci. 2000; 97:4449–52.
11. Bright JA, Marugán-Lobón J, Cobb SN, Rayfield EJ. The shapes of bird beaks are highly controlled by nondietary factors. Proc. Natl. Acad. Sci. U. S. A. 2016; 113:5352–7.
12. Bhullar B-AS, Morris ZS, Sefton EM, Tok A, Tokita M, Namkoong B, et al. 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. 2015; 69:1665–77.
13. Mallet J, Joron M. Evolution of Diversity in Warning Color and Mimicry: Polymorphisms, Shifting Balance, and Speciation. Annu. Rev. Ecol. Syst. 1999; 30:201–33.
14. Wellenreuther M, Svensson EI, Hansson B. Sexual selection and genetic colour polymorphisms in animals. Mol. Ecol. 2014; 23:5398–414.
15. Stuart-Fox D, Moussalli A. Camouflage, communication and thermoregulation: lessons from colour changing organisms. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 2009; 364:463–70.
16. Singh AP, Nüsslein-Volhard C. Zebrafish stripes as a model for vertebrate colour pattern formation. Curr. Biol. 2015; 25:R81–92.
17. Faria R, Renaut S, Galindo J, Pinho C, Melo-Ferreira J, Melo M, et al. Advances in Ecological Speciation: an integrative approach. Mol. Ecol. 2014; 23:513–21.
18. Chen L, DeVries AL, Cheng C-HC. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. 1997; 94:3811–6.
19. Emera D, Casola C, Lynch VJ, Wildman DE, Agnew D, Wagner GP. Convergent evolution of endometrial prolactin expression in primates, mice, and elephants through the independent recruitment of transposable elements. Mol. Biol. Evol. 2012;29:239–47.
20. Brandley MC, Young RL, Warren DL, Thompson MB, Wagner GP. Uterine gene expression in the live-bearing lizard, Chalcides ocellatus, reveals convergence of squamate reptile and mammalian pregnancy mechanisms. Genome Biol. Evol. 2012; 4:394–411.
21. Pankey MS, Minin VN, Imholte GC, Suchard MA, Oakley TH. Predictable transcriptome evolution in the convergent and complex bioluminescent organs of squid. Proc. Natl. Acad. Sci. U. S. A. 2014; 111:E4736–42.
22. Colombo M, Diepeveen ET, Muschick M, Santos ME, Indermaur A, Boileau N, et al. The ecological and genetic basis of convergent thick-lipped phenotypes in cichlid fishes. Mol. Ecol. 2013; 22:670–84.
23. Shubin N, Tabin C, Carroll S. Deep homology and the origins of evolutionary novelty. Nature. 2009; 457:818–23.
24. McCune AR, Schimenti JC. Using genetic networks and homology to understand the evolution

of phenotypic traits. Curr. Genomics. 2012;13:74–84.
25. Wagner GP. Homology, genes and evolutionary innovation. Princet. Univ. Press. Princet. 2014;
26. Wagner GP, Lynch VJ. The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. 2008; 23:377–85.
27. Keys DN. Recruitment of a hedgehog Regulatory Circuit in Butterfly Eyespot Evolution. Science. 1999; 283:532–4.
28. Shirai LT, Saenko S V, Keller RA, Jerónimo MA, Brakefield PM, Descimon H, et al. Evolutionary history of the recruitment of conserved developmental genes in association to the formation and diversification of a novel trait. BMC Evol. Biol. 2012;12:21.
29. Lynch VJ, Leclerc RD, May G, Wagner GP. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 2011;43:1154–9.
30. Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 2009;18:169–85.
31. Santos ME, Salzburger W. Evolution. How cichlids diversify. Science. 2012; 338:619–21.
32. Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 2014; 5:5149.
33. Salzburger W, Braasch I, Meyer A. Adaptive sequence evolution in a color gene involved in the formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biol. 2007; 5:51.
34. Salzburger W, Mack T, Verheyen E, Meyer A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 2005; 5:17.
35. Salzburger W, Mack T, Verheyen E, Meyer A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 2005; 5:17.
36. Meyer BS, Matschiner M, Salzburger W. A tribal level phylogeny of Lake Tanganyika cichlid fishes based on a genomic multi-marker approach. Mol. Phylogenet. Evol. 2015; 83:56–71.
37. Takahashi T, Sota T. A robust phylogeny among major lineages of the East African cichlids. Mol. Phylogenet. Evol. 2016;100:234–42.
38. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17:10.
39. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80.
40. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al. Orchestrating

high-throughput genomic analysis with Bioconductor. Nat. Methods. 2015;12:115–21.
41. Robinson MD, Smyth GK. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics. 2007; 23:2881–7.
42. Robinson MD, Smyth GK. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics. 2008; 9:321–32.
43. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010; 26:139–40.
44. McCarthy DJ, Chen Y, Smyth GK. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. Nucleic Acids Res. 2012; 40:4288–97.
45. Zhou X, Lindsay H, Robinson MD. Robustly detecting differential expression in RNA sequencing data using observation weights. Nucleic Acids Res. 2013; 18.
46. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol. 2010; 11:R106.
47. Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics. 2005; 21:3674–6.
48. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. 2010; 11:R14.
49. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007; 35:W182–5.
50. Mao X, Cai T, Olyarchuk JG, Wei L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. Bioinformatics. 2005; 21:3787–93.
51. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003; 13:2498–504.
52. Seyednasrollah F, Laiho A, Elo LL. Comparison of software packages for detecting differential expression in RNA-seq studies. Brief. Bioinform. 2013; bbt086 – .
53. Shin MK, Levorse JM, Ingram RS, Tilghman SM. The temporal requirement for endothelin receptor-B signalling during neural crest development. Nature.1999; 402:496–501.
54. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006; 313:101–4.
55. Burgoyne T, O'Connor MN, Seabra MC, Cutler DF, Futter CE. Regulation of melanosome number, shape and movement in the zebrafish retinal pigment epithelium by OA1 and PMEL. J. Cell Sci. 2015; 128:1400–7.
56. Beirl AJ, Linbo TH, Cobb MJ, Cooper CD. oca2 Regulation of chromatophore differentiation

and number is cell type specific in zebrafish. Pigment Cell Melanoma Res. 2014; 27:178–89.
57. Ménasché G, Ménager MM, Lefebvre JM, Deutsch E, Athman R, Lambert N, et al. A newly identified isoform of Slp2a associates with Rab27a in cytotoxic T cells and participates to cytotoxic granule secretion. Blood. 2008; 112:5052–62.
58. Mellgren EM, Johnson SL. pyewacket, a new zebrafish fin pigment pattern mutant. Pigment Cell Res. 2006; 19:232–8.
59. Price MA, Colvin Wanshura LE, Yang J, Carlson J, Xiang B, Li G, et al. CSPG4, a potential therapeutic target, facilitates malignant progression of melanoma. Pigment Cell Melanoma Res. 2011; 24:1148–57.
60. Eom DS, Inoue S, Patterson LB, Gordon TN, Slingwine R, Kondo S, et al. Melanophore migration and survival during zebrafish adult pigment stripe development require the immunoglobulin superfamily adhesion molecule Igsf11. PLoS Genet. 2012; 8:e1002899.
61. Li DY, Smith DG, Hardeland R, Yang MY, Xu HL, Zhang L, et al. Melatonin receptor genes in vertebrates. Int. J. Mol. Sci. 2013; 14:11208–23.
62. Zhang P, Liu W, Zhu C, Yuan X, Li D, Gu W, et al. Silencing of GPNMB by siRNA inhibits the formation of melanosomes in melanocytes in a MITF-independent fashion. PLoS One. 2012; 7:e42955.
63. Braasch I, Liedtke D, Volff J-N, Schartl M. Pigmentary function and evolution of tyrp1 gene duplicates in fish. Pigment Cell Melanoma Res. 2009; 22:839–50.
64. Lee Y, Nachtrab G, Klinsawat PW, Hami D, Poss KD. Ras controls melanocyte expansion during zebrafish fin stripe regeneration. Dis. Model. Mech. 2010; 3:496–503.
65. Higdon CW, Mitra RD, Johnson SL. Gene expression analysis of zebrafish melanocytes, iridophores, and retinal pigmented epithelium reveals indicators of biological function and developmental origin. PLoS One. 2013; 8:e67801.
66. Eom DS, Inoue S, Patterson LB, Gordon TN, Slingwine R, Kondo S, et al. Melanophore migration and survival during zebrafish adult pigment stripe development require the immunoglobulin superfamily adhesion molecule Igsf11. PLoS Genet. 2012; 8:e1002899.
67. Lister JA, Lane BM, Nguyen A, Lunney K. Embryonic expression of zebrafish MiT family genes tfe3b, tfeb, and tfec. Dev. Dyn. 2011; 240:2529–38.
68. Curran K, Lister JA, Kunkel GR, Prendergast A, Parichy DM, Raible DW. Interplay between Foxd3 and Mitf regulates cell fate plasticity in the zebrafish neural crest. Dev. Biol. 2010;344:107–18.
69. Lacosta AM, Canudas J, Gonzalez C, Muniesa P, Sarasa M, Dominguez L. Pax7 identifies neural crest, chromatophore lineages and pigment stem cells during zebrafish development. Int. J. Dev. Biol. 2007; 51:327–31.

70. Parichy D, Ransom D, Paw B, Zon L, Johnson S. An orthologue of the kit-related gene fms is required for development of neural crest-derived xanthophores and a subpopulation of adult melanocytes in the zebrafish, Danio rerio. Development. 2000;127:3031–44.
71. Parichy DM, Spiewak JE. Origins of adult pigmentation: diversity in pigment stem cell lineages and implications for pattern evolution. Pigment Cell Melanoma Res. 2015; 28:31–50.
72. Van Otterloo E, Li W, Garnett A, Cattell M, Medeiros DM, Cornell RA. Novel Tfap2-mediated control of soxE expression facilitated the evolutionary emergence of the neural crest. Development. 2012; 139:720–30.
73. Manohar M, Mei H, Franklin AJ, Sweet EM, Shigaki T, Riley BB, et al. Zebrafish (Danio rerio) endomembrane antiporter similar to a yeast cation/H(+) transporter is required for neural crest development. Biochemistry. 2010; 49:6557–66.
74. Boer EF, Howell ED, Schilling TF, Jette CA, Stewart RA. Fascin1-dependent Filopodia are required for directional migration of a subset of neural crest cells. PLoS Genet. 2015; 11:e1004946.
75. Yin C, Evason KJ, Maher JJ, Stainier DYR. The basic helix-loop-helix transcription factor, heart and neural crest derivatives expressed transcript 2, marks hepatic stellate cells in zebrafish: analysis of stellate cell entry into the developing liver. Hepatology. 2012; 56:1958–70.
76. McMenamin SK, Bain EJ, McCann AE, Patterson LB, Eom DS, Waller ZP, et al. Thyroid hormone-dependent adult pigment cell lineage and pattern in zebrafish. Science. 2014; 345:1358–61.
77. Zanotti G, Cendron L, Ramazzina I, Folli C, Percudani R, Berni R. Structure of zebra fish HIUase: insights into evolution of an enzyme to a hormone transporter. J. Mol. Biol. 2006; 363:1–9.
78. Vella KR, Ramadoss P, Lam FS, Harris JC, Ye FD, Same PD, et al. NPY and MC4R signaling regulate thyroid hormone levels during fasting through both central and peripheral pathways. Cell Metab. 2011; 14:780–90.
79. Tsui K-H, Chang Y-L, Feng T-H, Chang P-L, Juang H-H. Glycoprotein transmembrane nmb: an androgen-downregulated gene attenuates cell invasion and tumorigenesis in prostate carcinoma cells. Prostate. 2012;72:1431–42.
80. Müller JM, Isele U, Metzger E, Rempel A, Moser M, Pscherer A, et al. FHL2, a novel tissue-specific coactivator of the androgen receptor. EMBO J. 2000;19:359–69.
81. Appari M, Werner R, Wünsch L, Cario G, Demeter J, Hiort O, et al. Apolipoprotein D (APOD) is a putative biomarker of androgen receptor function in androgen insensitivity syndrome. J. Mol. Med. 2009; 87:623–32.
82. Di Prinzio CM, Botta PE, Barriga EH, Ríos EA, Reyes AE, Arranz SE. Growth hormone receptors in zebrafish (Danio rerio): adult and embryonic expression patterns. Gene Expr. Patterns. 2010; 10:214–25.

83. Ren D-L, Sun A-A, Li Y-J, Chen M, Ge S-C, Hu B. Exogenous melatonin inhibits neutrophil migration through suppression of ERK activation. J. Endocrinol. 2015; 227:49–60.
84. Ohno Y, Otaki JM. Spontaneous long-range calcium waves in developing butterfly wings. BMC Dev. Biol. 2015; 15:17.
85. Blasiole B, Degrave A, Canfield V, Boehmler W, Thisse C, Thisse B, et al. Differential expression of Na,K-ATPase alpha and beta subunit genes in the developing zebrafish inner ear. Dev. Dyn. 2003; 228:386–92.
86. Hood JD, Cheresh DA. Role of integrins in cell invasion and migration. Nat. Rev. Cancer. 2002; 2:91–100.
87. Venkatakrishnan AJ, Deupi X, Lebon G, Tate CG, Schertler GF, Babu MM. Molecular signatures of G-protein-coupled receptors. Nature. 2013; 494:185–94.
88. Ramos BCR, Moraes MNCM, Poletini MO, Lima LHRG, Castrucci AML. From blue light to clock genes in zebrafish ZEM-2S cells. PLoS One. 2014; 9:e106252.
89. Shiu SY, Ng N, Pang SF. A molecular perspective of the genetic relationships of G-protein coupled melatonin receptor subtypes. J. Pineal Res. 1996; 20:198–204.
90. Kurko D, Kapui Z, Nagy J, Lendvai B, Kolok S. Analysis of functional selectivity through G protein-dependent and -independent signaling pathways at the adrenergic α(2C) receptor. Brain Res. Bull. 2014; 107:89–101.
91. Naus CC, Laird DW. Implications and challenges of connexin connections to cancer. Nat. Rev. Cancer. 2010; 10:435–41.
92. Irion U, Frohnhöfer HG, Krauss J, Çolak Champollion T, Maischein H-M, Geiger-Rudolph S, et al. Gap junctions composed of connexins 41.8 and 39.4 are essential for colour pattern formation in zebrafish. Elife. 2014; 3:e05125.
93. Videira IF dos S, Moura DFL, Magina S. Mechanisms regulating melanogenesis. An. Bras. Dermatol. 2013; 88:76–83.
94. Ng A, Uribe RA, Yieh L, Nuckels R, Gross JM. Zebrafish mutations in gart and paics identify crucial roles for de novo purine synthesis in vertebrate pigmentation and ocular development. Development . 2009; 136:2601–11.
95. Martin A, Reed RD. Wnt signaling underlies evolution and development of the butterfly wing pattern symmetry systems. Dev. Biol. 2014; 395:367–78.
96. Martin A, Papa R, Nadeau N. Diversification of complex butterfly wing patterns by repeated regulatory evolution of a Wnt ligand. Proc. Natl Acad Sci U S A. 2012; 109: 12632-7.
97. Chitnis AB, Nogare DD, Matsuda M. Building the posterior lateral line system in zebrafish. Dev. Neurobiol. 2012; 72:234–55.
98. Parichy DM. Animal pigment pattern: an integrative model system for studying the

development, evolution, and regeneration of form. Semin. Cell Dev. Biol. 2009; 20:63–4.
99. Schneider PN, Slusarski DC, Houston DW. Differential role of Axin RGS domain function in Wnt signaling during anteroposterior patterning and maternal axis formation. PLoS One. 2012; 7:e44096.
100. Sonal, Sidhaye J, Phatak M, Banerjee S, Mulay A, Deshpande O, et al. Myosin Vb mediated plasma membrane homeostasis regulates peridermal cell size and maintains tissue homeostasis in the zebrafish epidermis. PLoS Genet. 2014; 10:e1004614.
101. Krauss J, Frohnhöfer HG, Walderich B, Maischein H-M, Weiler C, Irion U, et al. Endothelin signalling in iridophore development and stripe pattern formation of zebrafish. Biol. Open. 2014; 3:503–9.
102. Aoki Y, Saint-Germain N, Gyda M, Magner-Fink E, Lee Y-H, Credidio C, et al. Sox10 regulates the development of neural crest-derived melanocytes in Xenopus. Dev. Biol. 2003; 259:19–33.
103. Ticho BS, Stainier DY, Fishman MC, Breitbart RE. Three zebrafish MEF2 genes delineate somitic and cardiac muscle development in wild-type and mutant embryos. Mech. Dev. 1996; 59:205–18.
104. Arnoult L, Su KFY, Manoel D, Minervino C, Magriña J, Gompel N, et al. Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. Science. 2013; 339:1423–6.
105. Zhang J, Wagh P, Guay D, Sanchez-Pulido L, Padhi BK, Korzh V, et al. Loss of fish actinotrichia proteins and the fin-to-limb transition. Nature. 2010; 466:234–7.
106. Bentrop J, Marx M, Schattschneider S, Rivera-Milla E, Bastmeyer M. Molecular evolution and expression of zebrafish St8SiaIII, an alpha-2,8-sialyltransferase involved in myotome development. Dev. Dyn. 2008; 237:808–18.
107. White JA, Boffa MB, Jones B, Petkovich M. A zebrafish retinoic acid receptor expressed in the regenerating caudal fin. Development. 1994; 120:1861–72.
108. Stewart S, Stankunas K. Limited dedifferentiation provides replacement tissue during zebrafish fin regeneration. Dev. Biol. 2012; 365:339–49.
109. Song M, Yang H, Yao S, Ma F, Li Z, Deng Y, et al. A critical role of vascular endothelial growth factor D in zebrafish embryonic vasculogenesis and angiogenesis. Biochem. Biophys. Res. Commun. 2007; 357:924–30.
110. Weech P, Provost P, Tremblay N, Camato R, Milne R, Marcel Y, et al. Apolipoprotein D—An atypical apolipoprotein. Prog. Lipid Res. 1991; 30:259–66.
111. Rassart E, Bedirian A, Do Carmo S, Guinard O, Sirois J, Terrisse L, et al. Apolipoprotein D. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 2000; 1482:185–98.

112. Salazar-Peláez LM, Abraham T, Herrera AM, Correa MA, Ortega JE, Paré PD, et al. Vitronectin expression in the airways of subjects with asthma and chronic obstructive pulmonary disease. PLoS One. 2015; 10:e0119717.
113. Braasch I, Brunet F, Volff J-N, Schartl M. Pigmentation pathway evolution after whole-genome duplication in fish. Genome Biol. Evol. 2009; 1:479–93.
114. Sheets L, Ransom DG, Mellgren EM, Johnson SL, Schnapp BJ. Zebrafish melanophilin facilitates melanosome dispersion by regulating dynein. Curr. Biol. 2007;17:1721–34.
115. Watanabe M, Watanabe D, Kondo S. Polyamine sensitivity of gap junctions is required for skin pattern formation in zebrafish. Sci. Rep. 2012; 2:473.
116. Minchin JEN, Hughes SM. Sequential actions of Pax3 and Pax7 drive xanthophore development in zebrafish neural crest. Dev. Biol. 2008; 317:508–22.
117. Hou L, Arnheiter H, Pavan WJ. Interspecies difference in the regulation of melanocyte development by SOX10 and MITF. Proc. Natl. Acad. Sci. U. S. A. 2006;103:9081–5.
118. Shellswell GB, Bailey AJ, Duance VC, Restall DJ. Has collagen a role in muscle pattern formation in the developing chick wing? 1. An immunofluorescence study. J. Embryol. Exp. Morphol. 1980. 60:245–54.
119. Benard EL, Roobol SJ, Spaink HP, Meijer AH. Phagocytosis of mycobacteria by zebrafish macrophages is dependent on the scavenger receptor Marco, a key control factor of pro-inflammatory signalling. Dev. Comp. Immunol. 2014; 47:223–33.

Supplementary Table 1 Illumina sequencing reads before and after treatment for the ectodine blotch in *C. macrops*.

	reads	
tissues	raw	after
b1	20434263	20275011
b2	18203790	18203790
b3	22905154	22849275
n1	17010414	16972228
n2	15783093	15632290
n3	16487660	16438307
u1	15026915	14868869
u2	15692799	15635502
u3	15143728	15077115
d1	18704476	18657477
d2	18245783	18245783
d3	16161261	16105774

Supplementary Table 2 812 eggspots candidate genes of *A. burtoni*

gene	logFC	FDR
ENSONIT00000010472	-4.819645056	8.92E-21
ENSONIT00000013880	-4.370675902	2.76E-15
ENSONIT00000005086	-3.954814041	2.13E-41
ENSONIT00000010471	-3.512929101	2.97E-19
ENSONIT00000014453	-3.39814272	2.24E-08
ENSONIT00000005568	-2.871687496	2.62E-19
ENSONIT00000010442	-2.831527398	3.44E-23
ENSONIT00000011341	-2.752389288	1.37E-09
ENSONIT00000015659	-2.519736931	6.48E-11
ENSONIT00000011340	-2.507224945	5.15E-11
ENSONIT00000006331	-2.331811635	2.74E-16
ENSONIT00000020748	-2.206734279	2.16E-14
ENSONIT00000024871	-2.002821302	5.53E-27
ENSONIT00000018605	-1.957011268	1.11E-09
ENSONIT00000005936	-1.946554814	1.02E-08
ENSONIT00000018202	-1.925956792	3.34E-06
ENSONIT00000024469	-1.920078636	1.61E-17
ENSONIT00000015402	-1.826811046	4.02E-12
ENSONIT00000000868	-1.815833842	6.09E-10
ENSONIT00000018204	-1.788228929	2.83E-05
ENSONIT00000010127	-1.765826352	1.34E-06
ENSONIT00000014389	-1.750548839	0.000132035
ENSONIT00000022458	-1.749073587	9.15E-17
ENSONIT00000017889	-1.713415471	2.67E-20
ENSONIT00000010841	-1.659214849	1.18E-17
ENSONIT00000000334	-1.656990445	2.79E-17
ENSONIT00000012627	-1.637641578	9.16E-10
ENSONIT00000012628	-1.617723249	4.87E-08

ENSONIT00000019895	-1.592384592	6.24E-09
ENSONIT00000020694	-1.580009004	4.94E-16
ENSONIT00000012616	-1.523945978	0.000646484
ENSONIT00000021130	-1.503183969	0.005057278
ENSONIT00000013246	-1.441471435	3.57E-16
ENSONIT00000023141	-1.440920893	3.75E-06
ENSONIT00000016343	-1.419783934	6.32E-08
ENSONIT00000025532	-1.416067798	3.59E-07
ENSONIT00000013245	-1.39241415	1.77E-14
ENSONIT00000002816	-1.38946264	0.000230859
ENSONIT00000013247	-1.380750227	2.02E-14
ENSONIT00000003034	-1.376243388	0.002895408
ENSONIT00000000731	-1.369639545	3.11E-09
ENSONIT00000003588	-1.369284973	2.29E-08
ENSONIT00000016006	-1.368484544	5.15E-13
ENSONIT00000011169	-1.359946463	0.000181671
ENSONIT00000009796	-1.356547823	0.000105609
ENSONIT00000000542	-1.343731568	0.002887064
ENSONIT00000002815	-1.336512107	1.01E-05
ENSONIT00000018616	-1.336478826	9.79E-18
ENSONIT00000013731	-1.325846706	2.51E-07
ENSONIT00000006220	-1.312850446	7.47E-05
ENSONIT00000022472	-1.30108641	5.15E-13
ENSONIT00000001675	-1.296108402	1.89E-08
ENSONIT00000024288	-1.264212196	1.30E-14
ENSONIT00000019097	-1.253216908	6.14E-15
ENSONIT00000003320	-1.252562039	2.19E-05
ENSONIT00000010294	-1.251196638	0.003068286
ENSONIT00000003587	-1.236725093	2.62E-08
ENSONIT00000005048	-1.233823781	0.001709475
ENSONIT00000008721	-1.232549753	0.009397725
ENSONIT00000015130	-1.232399726	4.03E-15
ENSONIT00000019092	-1.230852404	6.79E-15
ENSONIT00000020905	-1.224354354	6.49E-06
ENSONIT00000015625	-1.222359565	0.000285762
ENSONIT00000000600	-1.220541815	0.0084186
ENSONIT00000002486	-1.22046903	0.008041634
ENSONIT00000005833	-1.209980864	1.37E-09
ENSONIT00000003326	-1.20818657	1.65E-12
ENSONIT00000013026	-1.20511568	0.003391351
ENSONIT00000001371	-1.204956183	3.44E-05
ENSONIT00000010298	-1.203952816	1.57E-06
ENSONIT00000025991	-1.194950098	0.001474614
ENSONIT00000017522	-1.188330906	0.009264907
ENSONIT00000020904	-1.187385817	3.02E-06
ENSONIT00000000626	-1.186619385	3.00E-05

ENSONIT00000016266	-1.176699756	4.06E-11
ENSONIT00000001209	-1.174954822	5.07E-10
ENSONIT00000005832	-1.170233063	3.22E-08
ENSONIT00000017728	-1.16986074	8.79E-06
ENSONIT00000011837	-1.167532951	5.94E-07
ENSONIT00000025491	-1.161145946	4.01E-12
ENSONIT00000013024	-1.160000271	3.50E-07
ENSONIT00000013919	-1.155754328	0.000682728
ENSONIT00000024553	-1.142169935	0.001863552
ENSONIT00000017521	-1.141500528	0.000361799
ENSONIT00000025490	-1.137169964	2.05E-11
ENSONIT00000021193	-1.136518717	1.17E-07
ENSONIT00000026203	-1.134037177	0.000410493
ENSONIT00000012812	-1.131830086	0.005057596
ENSONIT00000018012	-1.127712846	7.80E-10
ENSONIT00000015626	-1.126868361	0.001957822
ENSONIT00000022239	-1.124767375	3.82E-10
ENSONIT00000015135	-1.116952824	4.30E-06
ENSONIT00000008977	-1.115834228	4.15E-05
ENSONIT00000015512	-1.114354975	2.70E-07
ENSONIT00000024042	-1.114261722	0.008897298
ENSONIT00000004504	-1.113577395	1.90E-08
ENSONIT00000010497	-1.108925533	2.24E-05
ENSONIT00000009064	-1.107677656	1.77E-05
ENSONIT00000002684	-1.106738927	4.15E-09
ENSONIT00000016921	-1.105677699	4.41E-05
ENSONIT00000005686	-1.104777653	0.000145682
ENSONIT00000005687	-1.104520363	0.00016473
ENSONIT00000001323	-1.102914745	2.76E-06
ENSONIT00000002381	-1.096234446	0.002607936
ENSONIT00000023911	-1.08977153	6.47E-05
ENSONIT00000016091	-1.089462978	0.001165793
ENSONIT00000020895	-1.081734521	0.0007748
ENSONIT00000002243	-1.081513172	2.84E-09
ENSONIT00000009598	-1.076073953	3.26E-10
ENSONIT00000014998	-1.074266357	0.000163513
ENSONIT00000013027	-1.073037834	3.46E-06
ENSONIT00000020463	-1.072115336	0.002716683
ENSONIT00000015624	-1.071680267	0.001108069
ENSONIT00000011061	-1.071155421	5.82E-08
ENSONIT00000002242	-1.067384952	3.57E-12
ENSONIT00000017673	-1.054754579	8.41E-05
ENSONIT00000011083	-1.051190858	6.41E-05
ENSONIT00000023632	-1.048743343	3.32E-09
ENSONIT00000014048	-1.04578847	2.62E-10
ENSONIT00000020190	-1.043831026	1.38E-05

ENSONIT00000004792	-1.04297672	0.004715277
ENSONIT00000011188	-1.038389367	0.000849595
ENSONIT00000012634	-1.036643867	4.50E-09
ENSONIT00000011836	-1.036572605	5.12E-08
ENSONIT00000011187	-1.032721499	0.0007919
ENSONIT00000023430	-1.032302244	0.003324935
ENSONIT00000007947	-1.032191588	6.24E-09
ENSONIT00000007140	-1.032153051	0.002680352
ENSONIT00000021871	-1.029704262	3.56E-06
ENSONIT00000021398	-1.02455215	0.001551829
ENSONIT00000024415	-1.017042846	0.000350252
ENSONIT00000022134	-1.015714272	4.04E-05
ENSONIT00000013044	-1.010301351	5.32E-06
ENSONIT00000021977	-1.009790457	3.57E-06
ENSONIT00000013597	-1.006144153	1.87E-05
ENSONIT00000022921	-1.004825245	1.84E-06
ENSONIT00000017059	-0.991994777	7.79E-08
ENSONIT00000024552	-0.991198285	0.002467025
ENSONIT00000007390	-0.990932866	0.005446957
ENSONIT00000015852	-0.989655655	9.75E-05
ENSONIT00000004503	-0.986499065	2.24E-07
ENSONIT00000021979	-0.979985857	0.000177919
ENSONIT00000015424	-0.978182147	1.74E-05
ENSONIT00000020208	-0.973271168	1.59E-07
ENSONIT00000024032	-0.966554412	0.009694615
ENSONIT00000000536	-0.96528797	0.000784456
ENSONIT00000002274	-0.95996214	1.75E-08
ENSONIT00000006998	-0.956400741	5.52E-05
ENSONIT00000024421	-0.955674275	2.45E-08
ENSONIT00000016322	-0.945058988	0.001212437
ENSONIT00000025338	-0.942981765	7.08E-05
ENSONIT00000025337	-0.936513781	7.32E-05
ENSONIT00000023612	-0.935106495	9.10E-07
ENSONIT00000006997	-0.93399412	0.000158966
ENSONIT00000014423	-0.933706142	0.001030432
ENSONIT00000024618	-0.932256497	1.49E-07
ENSONIT00000021978	-0.931734846	0.000322495
ENSONIT00000018013	-0.93137833	0.00037604
ENSONIT00000005380	-0.926661349	0.003712117
ENSONIT00000002025	-0.921207507	8.15E-07
ENSONIT00000003015	-0.920734133	0.007370579
ENSONIT00000015422	-0.918507553	3.40E-06
ENSONIT00000021953	-0.914800742	3.10E-05
ENSONIT00000002000	-0.910022348	0.000105302
ENSONIT00000023475	-0.908499789	2.92E-05
ENSONIT00000014999	-0.908407951	0.002467025

ENSONIT00000016183	-0.905151835	3.33E-06
ENSONIT00000002024	-0.904793765	1.05E-07
ENSONIT00000015173	-0.902060797	0.000724969
ENSONIT00000011685	-0.899693691	1.67E-07
ENSONIT00000000955	-0.898558394	0.004232075
ENSONIT00000024767	-0.89147318	1.85E-05
ENSONIT00000012423	-0.887757848	0.003511995
ENSONIT00000021872	-0.885874602	0.000105302
ENSONIT00000000262	-0.885168505	1.18E-05
ENSONIT00000014927	-0.884168204	2.23E-05
ENSONIT00000020074	-0.883451691	1.04E-05
ENSONIT00000009314	-0.882496128	0.000491669
ENSONIT00000018670	-0.879766991	0.007499997
ENSONIT00000024554	-0.879074455	4.04E-05
ENSONIT00000023606	-0.872565473	0.007831797
ENSONIT00000024555	-0.870849477	5.21E-05
ENSONIT00000020246	-0.862042766	5.70E-07
ENSONIT00000012422	-0.853313835	3.33E-06
ENSONIT00000023014	-0.842339018	0.008185131
ENSONIT00000024731	-0.842110832	0.007702723
ENSONIT00000009765	-0.842053196	4.80E-05
ENSONIT00000024712	-0.840322692	7.07E-06
ENSONIT00000016198	-0.83467622	0.001971864
ENSONIT00000018927	-0.830993206	4.26E-05
ENSONIT00000000535	-0.824288181	0.000181839
ENSONIT00000001492	-0.822978043	9.68E-06
ENSONIT00000021647	-0.821364534	0.001165793
ENSONIT00000019760	-0.817745092	0.000201867
ENSONIT00000002780	-0.817602266	0.000365979
ENSONIT00000019548	-0.816613273	0.008712652
ENSONIT00000022255	-0.81593856	0.001325301
ENSONIT00000002275	-0.813063005	0.009787898
ENSONIT00000013910	-0.806269926	0.004544275
ENSONIT00000023565	-0.80615478	9.63E-06
ENSONIT00000019285	-0.804498841	3.01E-05
ENSONIT00000017058	-0.802933449	1.01E-05
ENSONIT00000005882	-0.794721877	6.41E-05
ENSONIT00000015331	-0.7914385	0.004742982
ENSONIT00000012663	-0.790968504	0.000400134
ENSONIT00000000695	-0.777608162	0.000110195
ENSONIT00000018664	-0.776100565	0.001818513
ENSONIT00000021308	-0.775778869	0.00036263
ENSONIT00000002415	-0.772325355	5.41E-05
ENSONIT00000020529	-0.771841432	0.004517295
ENSONIT00000007794	-0.771125517	0.000569407
ENSONIT00000021147	-0.770502392	0.008483549

ENSONIT00000003313	-0.76981344	0.000165058
ENSONIT00000005162	-0.768297715	4.26E-05
ENSONIT00000000545	-0.768161761	0.001080714
ENSONIT00000003173	-0.767431084	0.003391351
ENSONIT00000014043	-0.76728789	6.48E-06
ENSONIT00000021952	-0.763131589	0.000194017
ENSONIT00000026559	-0.75982387	0.00178085
ENSONIT00000009152	-0.759448959	0.004639521
ENSONIT00000014493	-0.757100217	0.000941316
ENSONIT00000013289	-0.755764121	0.009290079
ENSONIT00000021311	-0.752387989	0.000222645
ENSONIT00000010555	-0.747073776	0.003511995
ENSONIT00000005156	-0.746571131	3.00E-05
ENSONIT00000025157	-0.745143225	0.005442029
ENSONIT00000000102	-0.745127464	1.77E-05
ENSONIT00000010437	-0.745050665	4.04E-05
ENSONIT00000003172	-0.742466292	0.002502392
ENSONIT00000021307	-0.741235431	0.000380778
ENSONIT00000019037	-0.740756885	2.46E-05
ENSONIT00000003492	-0.740086305	0.007670686
ENSONIT00000001673	-0.739281101	0.003288375
ENSONIT00000021146	-0.738956322	0.000556907
ENSONIT00000019036	-0.737911302	1.45E-05
ENSONIT00000012427	-0.737214743	0.004134143
ENSONIT00000002706	-0.735342397	0.006295975
ENSONIT00000021951	-0.734241074	4.79E-05
ENSONIT00000026602	-0.733678066	0.004989379
ENSONIT00000018098	-0.731534099	0.000360105
ENSONIT00000021312	-0.730935996	0.000513862
ENSONIT00000005640	-0.730460127	2.27E-05
ENSONIT00000025996	-0.728519296	0.008183484
ENSONIT00000008936	-0.728298922	0.002616354
ENSONIT00000001255	-0.72083763	0.000461378
ENSONIT00000018771	-0.720751958	3.01E-05
ENSONIT00000018821	-0.719198965	0.005875392
ENSONIT00000024983	-0.715459359	6.06E-05
ENSONIT00000000906	-0.709416319	7.01E-05
ENSONIT00000009671	-0.709018716	0.000351512
ENSONIT00000004659	-0.701390369	0.000266698
ENSONIT00000008565	-0.700963216	0.006931814
ENSONIT00000004083	-0.699876067	0.001257622
ENSONIT00000007131	-0.698558617	0.001138794
ENSONIT00000021990	-0.697625426	0.007342447
ENSONIT00000023613	-0.69291479	9.74E-05
ENSONIT00000017799	-0.691273451	0.002076385
ENSONIT00000002070	-0.689184423	0.004845895

ENSONIT00000003289	-0.68822119	0.009397725
ENSONIT00000014030	-0.687599917	0.002232899
ENSONIT00000024422	-0.686508537	8.41E-05
ENSONIT00000012375	-0.684687486	0.005087231
ENSONIT00000001067	-0.684113968	0.000161072
ENSONIT00000013465	-0.683908369	0.007499997
ENSONIT00000007132	-0.683458284	0.003804963
ENSONIT00000007546	-0.682812841	0.005839422
ENSONIT00000001068	-0.682726746	0.000181839
ENSONIT00000018645	-0.682470749	0.000199612
ENSONIT00000025818	-0.68209346	0.001036872
ENSONIT00000019654	-0.67947344	0.003430467
ENSONIT00000022300	-0.676652918	0.006147845
ENSONIT00000009551	-0.675361622	0.001660495
ENSONIT00000013939	-0.674617473	0.000154508
ENSONIT00000025242	-0.66778583	0.002971282
ENSONIT00000006442	-0.667606236	0.009312357
ENSONIT00000022442	-0.666182719	0.000347171
ENSONIT00000002704	-0.664518051	0.000203978
ENSONIT00000005261	-0.664508232	0.005459179
ENSONIT00000004791	-0.662246859	0.003511995
ENSONIT00000006051	-0.662232472	0.007310583
ENSONIT00000006843	-0.657854072	0.007310583
ENSONIT00000025780	-0.65745318	0.00436975
ENSONIT00000004074	-0.657375106	0.000675188
ENSONIT00000007707	-0.656853815	0.004373446
ENSONIT00000016417	-0.656410979	0.009397725
ENSONIT00000010156	-0.656348285	0.00018103
ENSONIT00000018926	-0.65428274	0.00063659
ENSONIT00000011404	-0.65410732	0.001691279
ENSONIT00000016823	-0.652836872	0.004276197
ENSONIT00000022254	-0.65084927	0.003391351
ENSONIT00000006543	-0.649819535	0.0036089
ENSONIT00000003245	-0.647222824	0.000380778
ENSONIT00000011588	-0.641235822	0.00032005
ENSONIT00000001121	-0.641228806	0.002027414
ENSONIT00000013993	-0.641127134	0.002871195
ENSONIT00000000607	-0.639476633	0.002308431
ENSONIT00000019027	-0.635435214	0.000375158
ENSONIT00000012055	-0.633629696	0.003511995
ENSONIT00000022018	-0.631164262	0.000297246
ENSONIT00000007136	-0.627490767	0.008360172
ENSONIT00000019915	-0.627139569	0.004742982
ENSONIT00000011587	-0.625988693	0.000355786
ENSONIT00000003019	-0.625856701	0.007699881
ENSONIT00000021863	-0.624649374	0.002275524

ENSONIT00000024419	-0.62265668	0.001156681
ENSONIT00000007850	-0.622069574	0.007776252
ENSONIT00000026646	-0.619269551	0.006984278
ENSONIT00000003574	-0.61758898	0.009404029
ENSONIT00000014492	-0.617155453	0.001724094
ENSONIT00000017444	-0.616838095	0.002938871
ENSONIT00000014421	-0.615137402	0.004639891
ENSONIT00000007107	-0.612528362	0.001798636
ENSONIT00000016843	-0.611174966	0.003722468
ENSONIT00000003278	-0.61023223	0.001080714
ENSONIT00000005853	-0.601434086	0.008465388
ENSONIT00000012662	-0.599521539	0.002362591
ENSONIT00000018097	-0.595642356	0.001165119
ENSONIT00000004096	-0.595289851	0.002903901
ENSONIT00000007659	-0.594282559	0.009900364
ENSONIT00000013987	-0.590522285	0.005257192
ENSONIT00000024082	-0.588660237	0.002948495
ENSONIT00000018590	-0.584147302	0.002010424
ENSONIT00000003735	-0.583146216	0.001857751
ENSONIT00000019317	-0.58249141	0.008283889
ENSONIT00000023061	-0.581224746	0.002758502
ENSONIT00000024146	-0.580784391	0.008386027
ENSONIT00000003312	-0.576472695	0.00116471
ENSONIT00000004369	-0.574828493	0.002952688
ENSONIT00000024036	-0.57447104	0.002854752
ENSONIT00000024035	-0.574176014	0.002537288
ENSONIT00000016408	-0.571780283	0.003559054
ENSONIT00000004658	-0.571291906	0.00738547
ENSONIT00000024030	-0.566918113	0.003006611
ENSONIT00000024081	-0.565096547	0.006295975
ENSONIT00000003277	-0.564330165	0.002423542
ENSONIT00000008322	-0.561698299	0.005477629
ENSONIT00000011161	-0.559050732	0.002648187
ENSONIT00000020912	-0.555292022	0.009386797
ENSONIT00000020765	-0.554895805	0.009410492
ENSONIT00000011223	-0.549655132	0.006443757
ENSONIT00000023267	-0.540817087	0.00956493
ENSONIT00000015358	-0.540501689	0.008551774
ENSONIT00000023393	-0.539019719	0.008973013
ENSONIT00000004293	-0.53891918	0.006783549
ENSONIT00000022260	-0.53666744	0.003391351
ENSONIT00000007508	-0.533146681	0.00379638
ENSONIT00000022528	-0.532198597	0.007499997
ENSONIT00000015468	-0.532096354	0.00973839
ENSONIT00000002295	-0.530513398	0.003602463
ENSONIT00000024809	-0.518994379	0.008465388

ENSONIT00000000872	-0.508574024	0.009787898
ENSONIT00000015337	0.486001499	0.00998685
ENSONIT00000014110	0.500610037	0.008404555
ENSONIT00000020393	0.503834171	0.009957145
ENSONIT00000005539	0.503964461	0.009765874
ENSONIT00000018951	0.509381453	0.009333258
ENSONIT00000025282	0.512827733	0.007920934
ENSONIT00000021578	0.512924837	0.009887581
ENSONIT00000018117	0.513347586	0.009705242
ENSONIT00000018139	0.516630543	0.006752999
ENSONIT00000021313	0.516815034	0.008350853
ENSONIT00000015142	0.516932144	0.008449408
ENSONIT00000016881	0.517777518	0.008465388
ENSONIT00000023947	0.521464717	0.004989379
ENSONIT00000010686	0.522971099	0.005282079
ENSONIT00000013799	0.523110786	0.00756491
ENSONIT00000015508	0.52376048	0.005388465
ENSONIT00000013666	0.523954535	0.007141173
ENSONIT00000017758	0.526333243	0.005346321
ENSONIT00000026375	0.526616126	0.006984278
ENSONIT00000019783	0.527754825	0.009313055
ENSONIT00000021523	0.528583283	0.005265963
ENSONIT00000011037	0.52937336	0.004715277
ENSONIT00000000954	0.531681694	0.003883232
ENSONIT00000008097	0.532378328	0.004276197
ENSONIT00000004824	0.5326259	0.005270513
ENSONIT00000023910	0.534242137	0.009386797
ENSONIT00000003215	0.534751971	0.005477629
ENSONIT00000021797	0.534873477	0.007513599
ENSONIT00000023614	0.535903754	0.003590587
ENSONIT00000004989	0.537663426	0.008849903
ENSONIT00000022142	0.538033954	0.005066177
ENSONIT00000015745	0.538683238	0.007670686
ENSONIT00000012234	0.539250783	0.007702723
ENSONIT00000006158	0.53947167	0.005440291
ENSONIT00000026122	0.539628801	0.003602463
ENSONIT00000025319	0.540448428	0.00716667
ENSONIT00000019463	0.541444952	0.005542856
ENSONIT00000001344	0.541554297	0.006305579
ENSONIT00000020495	0.541622897	0.009876189
ENSONIT00000016318	0.541875172	0.003804963
ENSONIT00000003201	0.542215957	0.00435906
ENSONIT00000003862	0.542625671	0.00847429
ENSONIT00000022602	0.54414977	0.003174095
ENSONIT00000009541	0.54421219	0.008239559
ENSONIT00000010190	0.54514207	0.003603716

ENSONIT00000003214	0.545688142	0.004039281
ENSONIT00000004867	0.546203729	0.004344054
ENSONIT00000010852	0.546531308	0.008904272
ENSONIT00000022501	0.546760145	0.003511995
ENSONIT00000016246	0.547218704	0.007702723
ENSONIT00000004195	0.548210977	0.005192951
ENSONIT00000006311	0.548548633	0.0036089
ENSONIT00000020233	0.549300795	0.006966976
ENSONIT00000024218	0.549532023	0.002798357
ENSONIT00000020078	0.549628561	0.007993466
ENSONIT00000016300	0.549920951	0.007320567
ENSONIT00000025632	0.550476422	0.004210717
ENSONIT00000026220	0.550914523	0.004232075
ENSONIT00000026238	0.551660956	0.008226774
ENSONIT00000014550	0.554664326	0.003574783
ENSONIT00000024719	0.554882784	0.005066177
ENSONIT00000003510	0.555443079	0.002479915
ENSONIT00000005782	0.556260823	0.0036089
ENSONIT00000000243	0.557404033	0.002798357
ENSONIT00000012996	0.55964888	0.002382293
ENSONIT00000007078	0.561006824	0.009397725
ENSONIT00000016944	0.561526797	0.002795397
ENSONIT00000026100	0.56153611	0.00167794
ENSONIT00000013665	0.562235414	0.003505828
ENSONIT00000020492	0.565246606	0.00348236
ENSONIT00000001540	0.567050686	0.002966395
ENSONIT00000003470	0.570346866	0.003391351
ENSONIT00000019128	0.571148949	0.009887581
ENSONIT00000021734	0.572904725	0.001705053
ENSONIT00000004758	0.573307033	0.00756491
ENSONIT00000002768	0.574200335	0.002076234
ENSONIT00000000320	0.578560398	0.001729308
ENSONIT00000002472	0.578882266	0.006018644
ENSONIT00000018308	0.579138174	0.003324935
ENSONIT00000023131	0.579372735	0.002232388
ENSONIT00000023478	0.579650401	0.002948578
ENSONIT00000019130	0.580637335	0.005592454
ENSONIT00000016157	0.580906307	0.001680803
ENSONIT00000001671	0.581183387	0.006249588
ENSONIT00000022604	0.581392284	0.002035004
ENSONIT00000004704	0.582299884	0.001631702
ENSONIT00000002767	0.582847294	0.0012182
ENSONIT00000020266	0.582899518	0.001369804
ENSONIT00000010650	0.584152767	0.001957822
ENSONIT00000019454	0.584717011	0.004999144
ENSONIT00000003183	0.584864453	0.002758794

ENSONIT00000019453	0.586922856	0.004957429
ENSONIT00000025984	0.590823809	0.00530232
ENSONIT00000012995	0.592846307	0.000996317
ENSONIT00000016602	0.593160629	0.005133536
ENSONIT00000022802	0.593311089	0.005734294
ENSONIT00000019367	0.593944954	0.006295975
ENSONIT00000014551	0.595732516	0.001138794
ENSONIT00000026535	0.598350127	0.000736022
ENSONIT00000003616	0.598812324	0.006538847
ENSONIT00000022738	0.599821248	0.000758074
ENSONIT00000004710	0.600528585	0.0029608
ENSONIT00000006087	0.60123066	0.001810085
ENSONIT00000015913	0.602387612	0.001108069
ENSONIT00000012592	0.60368538	0.001412718
ENSONIT00000003753	0.604150547	0.005222305
ENSONIT00000009776	0.604361419	0.001179731
ENSONIT00000019751	0.605370568	0.002232601
ENSONIT00000016563	0.607409045	0.001635214
ENSONIT00000004015	0.6093277	0.004579531
ENSONIT00000022739	0.60933036	0.000576223
ENSONIT00000018423	0.610016474	0.000781656
ENSONIT00000010358	0.612730146	0.003630085
ENSONIT00000018135	0.614068431	0.0024567
ENSONIT00000009761	0.614081537	0.000722919
ENSONIT00000004499	0.620017796	0.000941316
ENSONIT00000024749	0.620478299	0.000576223
ENSONIT00000026187	0.620860624	0.000477519
ENSONIT00000016236	0.621189231	0.001412718
ENSONIT00000003075	0.626629974	0.00164586
ENSONIT00000003484	0.626810185	0.001670464
ENSONIT00000001708	0.628399252	0.003110224
ENSONIT00000012978	0.630542934	0.000718909
ENSONIT00000016319	0.63118753	0.000534915
ENSONIT00000007501	0.632043288	0.00072551
ENSONIT00000019228	0.632063669	0.001108069
ENSONIT00000018886	0.632481902	0.000695102
ENSONIT00000001555	0.632701886	0.005827229
ENSONIT00000022737	0.633657135	0.000357738
ENSONIT00000010649	0.635045345	0.000536517
ENSONIT00000026758	0.636086012	0.000340808
ENSONIT00000026712	0.636177733	0.000880452
ENSONIT00000023683	0.636369966	0.002467025
ENSONIT00000007506	0.636754521	0.000491669
ENSONIT00000015559	0.637193624	0.001390047
ENSONIT00000013305	0.637368598	0.005243549
ENSONIT00000004358	0.638203622	0.000233645

ENSONIT00000011120	0.638673142	0.002974941
ENSONIT00000024828	0.640014743	0.00340041
ENSONIT00000024944	0.640328688	0.003214851
ENSONIT00000013640	0.643262796	0.00133972
ENSONIT00000004639	0.643328634	0.001443278
ENSONIT00000000362	0.644751318	0.004809562
ENSONIT00000023894	0.645428707	0.008616436
ENSONIT00000021018	0.646747067	0.003907884
ENSONIT00000006105	0.647404332	0.003013738
ENSONIT00000002629	0.648454573	0.009696448
ENSONIT00000017257	0.649237131	0.000231465
ENSONIT00000018100	0.650671123	0.000569439
ENSONIT00000018309	0.653865788	0.001971864
ENSONIT00000010599	0.654265961	0.00024289
ENSONIT00000006452	0.655342921	0.000403405
ENSONIT00000005630	0.657830831	0.000550707
ENSONIT00000006451	0.658738861	0.00037604
ENSONIT00000011632	0.659986618	0.001798636
ENSONIT00000004357	0.661862797	0.000131607
ENSONIT00000017475	0.662370938	0.000181839
ENSONIT00000017481	0.662370947	0.000181839
ENSONIT00000001820	0.664366202	0.000259895
ENSONIT00000014557	0.669539613	0.004741065
ENSONIT00000003076	0.67051247	0.000629005
ENSONIT00000003818	0.670653103	0.001810085
ENSONIT00000018200	0.671967828	0.000421667
ENSONIT00000025384	0.672272889	0.004296068
ENSONIT00000019264	0.672445761	0.0074155
ENSONIT00000012330	0.677012972	0.004060825
ENSONIT00000022166	0.678907318	0.001379168
ENSONIT00000022822	0.680103129	0.00039213
ENSONIT00000001775	0.681158777	0.000355786
ENSONIT00000012329	0.68138329	0.001259152
ENSONIT00000026048	0.69067733	0.00638284
ENSONIT00000025312	0.691947922	0.004569167
ENSONIT00000015180	0.692463552	3.27E-05
ENSONIT00000018947	0.692677257	0.000209917
ENSONIT00000007516	0.697361075	0.003324935
ENSONIT00000017466	0.697732464	9.00E-05
ENSONIT00000026495	0.698443008	9.74E-05
ENSONIT00000004999	0.699192211	0.009666239
ENSONIT00000025383	0.705808936	0.001108281
ENSONIT00000019762	0.708082796	0.001559512
ENSONIT00000001125	0.717617949	0.000148745
ENSONIT00000005498	0.718062169	0.000322538
ENSONIT00000015801	0.719812489	0.005388465

ENSONIT00000000763	0.720392096	0.005803678
ENSONIT00000012218	0.72204664	0.000121247
ENSONIT00000021842	0.722341686	0.000392507
ENSONIT00000020155	0.72499525	2.23E-05
ENSONIT00000006598	0.725817978	0.008404555
ENSONIT00000015702	0.727690504	0.000355802
ENSONIT00000013974	0.728408665	0.000984557
ENSONIT00000023203	0.729889193	0.000201867
ENSONIT00000024853	0.732414826	2.61E-05
ENSONIT00000009569	0.73306492	0.009887581
ENSONIT00000019562	0.733719958	0.004639521
ENSONIT00000021827	0.733796565	0.00499538
ENSONIT00000020322	0.73382567	0.008897298
ENSONIT00000011315	0.734544827	0.000180213
ENSONIT00000000762	0.737084732	0.002883398
ENSONIT00000000242	0.739922797	4.03E-05
ENSONIT00000011033	0.740123172	0.001082797
ENSONIT00000002960	0.740227481	0.00233838
ENSONIT00000023202	0.741751626	9.75E-05
ENSONIT00000014036	0.742287051	8.76E-06
ENSONIT00000022096	0.743368331	6.41E-05
ENSONIT00000015636	0.746181834	6.70E-06
ENSONIT00000011314	0.747003273	0.000354492
ENSONIT00000010690	0.748927469	5.47E-05
ENSONIT00000010691	0.74893449	4.89E-05
ENSONIT00000021439	0.748997839	0.009540895
ENSONIT00000026211	0.750017509	4.72E-05
ENSONIT00000025871	0.758308289	0.005478535
ENSONIT00000023918	0.759216655	0.002200348
ENSONIT00000002223	0.761785992	3.50E-06
ENSONIT00000005094	0.761870743	0.00156384
ENSONIT00000017598	0.7634705	0.002161794
ENSONIT00000000523	0.766218282	0.000278886
ENSONIT00000004497	0.767147295	0.000306575
ENSONIT00000006596	0.767308256	1.46E-05
ENSONIT00000016474	0.768568015	0.000180213
ENSONIT00000006684	0.769138734	0.000270987
ENSONIT00000026288	0.769235559	6.47E-05
ENSONIT00000004371	0.771628516	0.007670686
ENSONIT00000006304	0.771848029	0.000421563
ENSONIT00000008117	0.774458226	3.75E-06
ENSONIT00000008031	0.776035094	0.003523351
ENSONIT00000008759	0.77684028	0.0036089
ENSONIT00000026028	0.777754361	2.94E-05
ENSONIT00000007411	0.778685882	0.001116899
ENSONIT00000018051	0.781722817	0.001653434

ENSONIT00000005574	0.781855982	0.007265505
ENSONIT00000025870	0.783636067	0.003965406
ENSONIT00000026666	0.783723261	8.93E-05
ENSONIT00000017688	0.784280269	9.04E-06
ENSONIT00000010116	0.784495449	0.009099556
ENSONIT00000011034	0.787111576	0.000255644
ENSONIT00000025790	0.788020318	0.00156384
ENSONIT00000014884	0.789156084	4.00E-06
ENSONIT00000010837	0.789715058	0.000584435
ENSONIT00000004755	0.790152969	6.20E-06
ENSONIT00000026715	0.79342736	0.000181671
ENSONIT00000001411	0.79413248	7.33E-05
ENSONIT00000020654	0.794811794	0.003675635
ENSONIT00000011634	0.794843522	7.67E-05
ENSONIT00000006305	0.796142906	0.00197647
ENSONIT00000003041	0.796181913	0.000738713
ENSONIT00000005609	0.798544589	0.000380778
ENSONIT00000004518	0.799748391	5.63E-05
ENSONIT00000005610	0.804116008	0.000347342
ENSONIT00000000719	0.804772936	0.001117772
ENSONIT00000010272	0.806324372	9.86E-06
ENSONIT00000003902	0.809159513	0.002391091
ENSONIT00000001253	0.809690531	1.65E-05
ENSONIT00000024683	0.809857306	3.98E-05
ENSONIT00000004085	0.815403077	2.15E-06
ENSONIT00000010838	0.816810405	0.002152063
ENSONIT00000004199	0.817583979	3.12E-06
ENSONIT00000004200	0.817988972	3.20E-06
ENSONIT00000004756	0.818162755	2.44E-05
ENSONIT00000009568	0.820192632	0.001030432
ENSONIT00000024352	0.822649563	1.42E-05
ENSONIT00000007146	0.822853253	0.000762381
ENSONIT00000026359	0.823622326	0.000195549
ENSONIT00000010490	0.828720009	8.42E-07
ENSONIT00000007147	0.831500663	0.000466779
ENSONIT00000003627	0.837564205	1.15E-06
ENSONIT00000006771	0.839596729	7.47E-05
ENSONIT00000019874	0.840207162	7.58E-06
ENSONIT00000019564	0.840208888	0.000819739
ENSONIT00000001254	0.841208171	0.000312591
ENSONIT00000016302	0.842564425	1.53E-05
ENSONIT00000015953	0.842871227	0.001969966
ENSONIT00000001690	0.850618207	0.009188801
ENSONIT00000021427	0.851551844	0.000637243
ENSONIT00000025220	0.852026406	3.90E-06
ENSONIT00000024261	0.854795791	0.00021329

ENSONIT00000005252	0.855309196	0.000284636
ENSONIT00000011721	0.855873374	6.64E-05
ENSONIT00000024197	0.856909172	0.000229304
ENSONIT00000003667	0.859760073	0.002616354
ENSONIT00000006770	0.862612645	4.70E-05
ENSONIT00000020655	0.863303894	0.001453883
ENSONIT00000014882	0.865104583	1.02E-07
ENSONIT00000009275	0.866285798	1.62E-05
ENSONIT00000021645	0.874477701	2.39E-06
ENSONIT00000009449	0.875186214	0.004859295
ENSONIT00000018848	0.882603546	2.15E-06
ENSONIT00000025425	0.885104384	5.87E-07
ENSONIT00000020235	0.886234433	7.32E-08
ENSONIT00000025549	0.887781701	0.000417442
ENSONIT00000026113	0.889604456	1.87E-05
ENSONIT00000009276	0.894970246	7.07E-06
ENSONIT00000025810	0.897630702	0.005670571
ENSONIT00000025765	0.898297751	2.70E-07
ENSONIT00000010153	0.902110019	0.008404555
ENSONIT00000006778	0.905806927	2.18E-06
ENSONIT00000009448	0.913279486	0.000614794
ENSONIT00000006773	0.913638304	1.29E-05
ENSONIT00000001691	0.916966556	0.002118105
ENSONIT00000006737	0.920800014	7.47E-05
ENSONIT00000020351	0.923151186	9.75E-05
ENSONIT00000006776	0.926118233	5.08E-06
ENSONIT00000024188	0.928405893	0.002295482
ENSONIT00000005090	0.929199431	1.24E-06
ENSONIT00000024196	0.932238525	6.72E-05
ENSONIT00000025323	0.937092925	2.73E-08
ENSONIT00000025550	0.938183072	4.72E-05
ENSONIT00000022574	0.940554192	1.49E-06
ENSONIT00000018026	0.944999279	2.22E-08
ENSONIT00000007853	0.945288153	2.15E-06
ENSONIT00000022400	0.945626414	5.98E-05
ENSONIT00000025911	0.945630803	0.006633249
ENSONIT00000022478	0.950177107	6.52E-05
ENSONIT00000006779	0.952813053	1.23E-05
ENSONIT00000009011	0.953978097	1.49E-07
ENSONIT00000003860	0.957902479	0.000107509
ENSONIT00000020789	0.960623852	0.00420737
ENSONIT00000010892	0.963594034	1.33E-05
ENSONIT00000010082	0.96981661	1.86E-05
ENSONIT00000018860	0.970453735	8.93E-05
ENSONIT00000026223	0.974844708	5.74E-05
ENSONIT00000005224	0.976682464	0.000171129

ENSONIT00000000034	0.977477395	0.000514474
ENSONIT00000004606	0.977892834	4.07E-06
ENSONIT00000017612	0.97834112	0.003080593
ENSONIT00000010807	0.978397537	1.68E-07
ENSONIT00000021628	0.981647225	5.25E-05
ENSONIT00000020657	0.983559071	0.003464108
ENSONIT00000018137	0.983986982	8.08E-06
ENSONIT00000018976	0.984473255	0.000229304
ENSONIT00000011958	0.994477731	1.99E-05
ENSONIT00000026637	0.998733482	0.001705053
ENSONIT00000007838	1.003511741	0.00043059
ENSONIT00000009012	1.007515097	2.45E-08
ENSONIT00000024260	1.009460069	1.06E-05
ENSONIT00000010081	1.011343541	7.10E-06
ENSONIT00000009645	1.015038557	0.003729888
ENSONIT00000008451	1.021314245	1.43E-05
ENSONIT00000005805	1.021690754	0.004149224
ENSONIT00000023817	1.034736342	7.55E-06
ENSONIT00000024909	1.042382259	4.85E-07
ENSONIT00000014558	1.054094296	0.000141358
ENSONIT00000016804	1.057766633	1.07E-05
ENSONIT00000014881	1.061863757	1.64E-11
ENSONIT00000006646	1.063033405	2.31E-11
ENSONIT00000003125	1.068735778	1.97E-07
ENSONIT00000013318	1.069106214	0.001043279
ENSONIT00000010975	1.069309844	2.33E-05
ENSONIT00000009623	1.072878629	1.75E-06
ENSONIT00000006647	1.078520066	4.01E-12
ENSONIT00000010498	1.081789178	0.000209917
ENSONIT00000024667	1.085306722	0.003523351
ENSONIT00000015855	1.08566407	2.53E-06
ENSONIT00000003464	1.092297579	2.74E-06
ENSONIT00000006160	1.095674451	0.002438405
ENSONIT00000017323	1.095982489	0.001810085
ENSONIT00000019979	1.110947427	0.000421912
ENSONIT00000001384	1.112027328	0.000158966
ENSONIT00000010893	1.113061606	1.04E-07
ENSONIT00000023827	1.119045518	1.54E-12
ENSONIT00000001264	1.127107153	4.06E-11
ENSONIT00000008365	1.136820469	7.45E-07
ENSONIT00000016998	1.140362789	6.24E-09
ENSONIT00000012779	1.146232922	0.009705242
ENSONIT00000017248	1.14996419	1.38E-06
ENSONIT00000006573	1.153534846	7.85E-07
ENSONIT00000015650	1.156688649	0.000448356
ENSONIT00000014444	1.15792879	0.000388416

ENSONIT00000015859	1.164560302	3.25E-09
ENSONIT00000018320	1.166311957	7.35E-06
ENSONIT00000003463	1.172709195	6.37E-09
ENSONIT00000004115	1.176376666	1.43E-05
ENSONIT00000006572	1.180220559	8.97E-08
ENSONIT00000010665	1.186494933	0.00092288
ENSONIT00000023113	1.191894181	1.26E-05
ENSONIT00000023698	1.192602354	3.06E-09
ENSONIT00000016184	1.209154767	5.01E-11
ENSONIT00000006082	1.210520951	2.08E-10
ENSONIT00000019225	1.216179254	0.000461378
ENSONIT00000023422	1.222545165	0.001390047
ENSONIT00000006767	1.222639886	7.15E-08
ENSONIT00000006436	1.224053015	0.000470672
ENSONIT00000026449	1.224942991	0.002027414
ENSONIT00000022716	1.230162007	3.51E-05
ENSONIT00000010197	1.232494841	3.88E-14
ENSONIT00000013739	1.235543022	4.01E-12
ENSONIT00000006762	1.240521175	1.28E-07
ENSONIT00000016190	1.240633433	1.84E-13
ENSONIT00000021326	1.246370214	0.001095178
ENSONIT00000018319	1.252960629	5.77E-06
ENSONIT00000023340	1.254929492	5.40E-14
ENSONIT00000022715	1.26234961	3.27E-05
ENSONIT00000022712	1.26381908	9.75E-05
ENSONIT00000022159	1.263956808	2.05E-11
ENSONIT00000023423	1.272946606	0.004781657
ENSONIT00000018318	1.273042547	1.85E-07
ENSONIT00000009081	1.274022604	0.00502486
ENSONIT00000016189	1.274610741	1.06E-11
ENSONIT00000003206	1.276504323	0.001988971
ENSONIT00000023427	1.284041233	3.87E-12
ENSONIT00000003858	1.284129801	1.34E-12
ENSONIT00000004670	1.29436048	1.26E-05
ENSONIT00000016188	1.312756086	2.80E-13
ENSONIT00000003157	1.313973769	5.34E-14
ENSONIT00000020547	1.31848579	5.50E-06
ENSONIT00000013796	1.334713623	6.28E-10
ENSONIT00000015035	1.33903583	2.62E-10
ENSONIT00000022711	1.339758947	1.27E-05
ENSONIT00000018609	1.346238802	0.008251537
ENSONIT00000018003	1.350255046	6.93E-07
ENSONIT00000009921	1.365620996	1.08E-05
ENSONIT00000008107	1.370024782	0.000180213
ENSONIT00000020019	1.372332133	1.00E-06
ENSONIT00000008772	1.373133192	5.75E-05

ENSONIT00000010408	1.374217783	1.41E-13
ENSONIT00000017692	1.381733923	0.000681505
ENSONIT00000004956	1.382602739	5.07E-11
ENSONIT00000018002	1.396238442	1.82E-07
ENSONIT00000007865	1.397275186	3.61E-17
ENSONIT00000006471	1.407810774	1.32E-07
ENSONIT00000025426	1.432822965	3.08E-14
ENSONIT00000020017	1.442929556	2.73E-09
ENSONIT00000010657	1.447948802	0.000285762
ENSONIT00000008134	1.453334244	0.005380757
ENSONIT00000020018	1.453772843	2.93E-09
ENSONIT00000022298	1.454109282	1.07E-19
ENSONIT00000010196	1.485596668	2.62E-20
ENSONIT00000020447	1.500197269	2.65E-09
ENSONIT00000009633	1.502479872	4.20E-16
ENSONIT00000009908	1.504505088	0.000181839
ENSONIT00000006870	1.512800785	2.39E-06
ENSONIT00000026256	1.55554354	8.37E-21
ENSONIT00000008106	1.563184819	1.50E-06
ENSONIT00000022299	1.570687349	8.69E-21
ENSONIT00000006869	1.579378512	2.11E-06
ENSONIT00000017362	1.586037285	7.26E-17
ENSONIT00000016614	1.608021806	1.05E-05
ENSONIT00000004437	1.669199133	1.42E-10
ENSONIT00000011917	1.675310447	0.000745842
ENSONIT00000015928	1.704292755	3.50E-06
ENSONIT00000009506	1.749744032	8.29E-21
ENSONIT00000026201	1.775187975	0.000319709
ENSONIT00000018977	1.787239907	4.12E-06
ENSONIT00000009634	1.790134571	1.05E-12
ENSONIT00000000055	1.791135305	1.64E-07
ENSONIT00000001830	1.833435409	1.84E-11
ENSONIT00000001866	1.859993502	4.63E-09
ENSONIT00000020635	1.923688586	9.46E-05
ENSONIT00000008710	1.936756142	4.96E-13
ENSONIT00000014203	2.000784235	2.68E-05
ENSONIT00000020249	2.011737445	7.76E-09
ENSONIT00000004432	2.02310151	3.56E-06
ENSONIT00000006597	2.023297111	7.80E-13
ENSONIT00000016806	2.130994709	2.13E-29
ENSONIT00000000720	2.36439842	2.10E-07
ENSONIT00000024154	2.742190962	3.04E-14
ENSONIT00000013306	2.769961809	3.60E-09
ENSONIT00000008841	2.851810385	6.17E-44
ENSONIT00000010488	3.08020623	1.62E-24
ENSONIT00000008163	3.323935902	2.26E-51

ENSONIT00000008162	3.392575555	2.26E-48
ENSONIT00000003170	4.682409369	3.20E-51

Supplementary Table 3 Protein-protein interactions based on String database

gene name	Connect degree
gnai1	18
myo5aa	17
gnao1a	15
gnaq	14
ednrb1a	12
sox10	11
mitfb	11
gch2	10
mef2aa	10
prkar1b	10
LOC566430	10
tyrp1b	10
plcb3	9
rxfp2b	9
tubb5	9
impdh1b	9
cdk15	9
rnd1l	9
rasd1	9
mc1r	8
myo16	8
adcy5	7
pde7a	7
foxd3	6
rgs4	6
pax7a	6
rgs6	6
vcl	6
mtnr1aa	6
drd3	6
fzd7a	6
pnp4a	6
lrrc8a	6
fzd3a	5
adra2b	5
chn1	5
tns1	5
oca2	5
kif5ba	5
zgc:152958	5
paics	5

calb2a	5
trpm1b	5
pip5k1bb	4
HDAC5	4
slc45a2	4
ENSDARG00000074528	4
pnp5a	4
SCGN	4
mapre3a	4
ap1s2	4
vangl2	3
opn4a	3
slc2a15b	3
GLP2R	3
mc4r	3
zgc:92195	3
RAB27B	3
cx43	3
itga6a	3
cecr1b	3
trpm1a	3
ENSDARG00000042124	3
rasef	3
kif1a	3
trpv1	3
CR384099.1	3
stk17a	3
pmelb	2
col10a1	2
dgkh	2
mlphb	2
cax1	2
acsl1	2
acsl3b	2
zgc:55621	2
coro2ba	2
KCNK4	2
ENSDARG00000005159	2
zgc:110789	2
slc23a1	2
wu:fc33e05	2
ghra	2
gpr143	2
NEURL1B	2
glulb	2
fosb	2

mab21l1	2
mfi2	2
hunk	2
zgc:162999	2
atp1b3a	2
fam123a	2
got1l1	2
zgc:77375	2
urah	2
PTS	1
ARHGEF17	1
COL12A1	1
slc2a11b	1
vtna	1
col2a1a	1
gpnmb	1
zeb2a	1
GPR52	1
IGSF11	1
sgta	1
LOC562524	1
ENSDARG00000074640	1
ckba	1
stmn1b	1
cx41.8	1
shc2	1
myoc	1
slc2a15a	1
tfap2e	1
lgals3l	1
rgmb	1
lfng	1
sytl2	1
rab34a	1
ndrg4	1
gga1	1
csrp2	1
raraa	1
CRISPLD2	1
stc2a	1
slc25a38a	1
alx4a	1
tmeff1	1
c10orf11	1
sypa	1
syngr1a	1

akr1b1	1
ENSDARG00000069926	1
mtnr1c	1

Chapter 4

Gene network rewiring of the repeated evolution of innovative anal fin pigmentation patterns in cichlid fish

Langyu Gu, Natacha Espirito Santo, Zuzana Musilová, Nicolas Boileau, Walter Salzburger

Chapter 4

Gene network rewiring of the repeated evolution of innovative anal fin pigmentation patterns in cichlid fish

Langyu Gu[1], Natacha Espirito Santo[1], Zuzana Musilová[2], Nicolas Boileau[1], Walter Salzburger[1]*

1 Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland
2 Department of Zoology, Faculty of Science, Charles University in Prague, Vinicna 7,128 44 Prague, Czech Republic

*Correspondence:
Walter Salzburger, email: walter.salzburger@unibas.ch

Abstract

The origination and evolution of novelty is one of the most fascinating questions in evolutionary biology. However, how the underlying gene network are rewired from the ancestral gene network to produce evolutionary novelties is largely unknown. Besides, what are the mechanisms behind the different evolvability of novelties also needs to be investigated. The repeated evolution of innovative pigmentation patterns on the anal fin in East African cichlid fish is an ideal model to study these questions. One pattern is eggspots, the circular pigmentation pattern with a transparent outer ring that emerged once in the most species rich cichlid lineage, the haplochromines, exhibiting large varieties with different numbers, sizes and colours. While ancestral haplochromine species feature another fin pigment trait in form of blotch, which is reddish with ill-defined boundary. Anal fin pigmentation pattern was also independently evolved in the ectodine lineage, which possess similar blotch pattern as the haplochromine blotch. Unlike haplochromine eggspots, the ectodine blotch shows almost no variation among species. By sequencing the whole gene sequences of 40 previously identified ectodine blotch and haplochromine eggspots related candidate genes across the phylogeny, followed by data analysis including transcription factor binding prediction and positive selection detection, our study proposed three hypothesis of the gene network relationships among eggspots, the ectodine blotch and anal fin: 1) “eggspots>blotch=fin”; 2) “eggspots=blotch=fin”; 3) “eggspots=/>blotch>fin”, and our data mostly supported the first hypothesis, which means that eggspots, but not the ectodine blotch, might have a much more independent gene network from the ancestral anal fin gene network, thus providing a clue of the mechanism about the different evolvability between eggspots and the ectodine blotch. In addition, eggspots might have different roles as the ectodine blotch, and selection might be stronger on it. Besides, the ectodine blotch and the haplochromine blotch might have originated independently.

key words

novelty, gene network, evolvability, cichlid fish, eggspots, blotch

Introduction

Phenotypic novelties are important promoters of biodiversity as they provide new raw materials for speciation and adaptation and, ultimately, diversification (e.g. eyespots in nymphalid butterflies (Monteiro 2015), wing spots in the *Drosophila melanogaster* lineage (Arnoult et al. 2013), the neural central crest in vertebrates (Green et al. 2015) and bird beaks (Bhullar et al. 2015; Bright et al. 2016)). Previously, evolutionary novelty was defined as "*novel structures that are neither homologous to any structure in the ancestral species nor serially homologous to any part of the same organism*" (Müller and Wagner 1991). However, the boundary between homology and analogy is often difficult to define (Wagner 2014). While the phenotype, as product of genotype and the environment, is inherited indirectly from generation to generation, the underlying gene networks (gene-gene interactions) are inherited directly (Wagner 2014). Such gene-gene interactions could be much more important than individual genes during the emergence and evolution of innovations. Because interactions among genes in a network can have a buffering effect on external and internal perturbations, depending on internal parameters of gene network to allow a system to maintain its functions (biological robustness (Kitano 2004)). On the other hand, this buffering effect can bear cryptic variations for the emergence of evolutionary innovation in a changed environment or genetic background (Wagner 2005). Therefore, it is time to shift the focus from individual genes, natural selection centralism and frequency-derived phenotypic variation description, to the gene network for phenotypic evolution, ontogenetic development and the causation of evolutionary innovation (Pigliucci and Müller 2010).

To examine how gene networks are rewired during the evolution of phenotypic innovation, it is important to know how gene networks are built up at the first place. One aspect is to determine the relative role of *cis*- and *trans*-regulatory elements in the evolution of phenotypic innovation (Glassford et al. 2015; Wallbank et al. 2016). Open questions are, for example, 1) what is the relatively contribution of *cis*- and *trans*- regulation in phenotypic innovation? In some cases, *trans*-regulatory elements play an important role, such as repeated co-option of transcription factor (TF) *Optix* in the formation of red colour patterns in the wings of *Heliconius* (Martin et al. 2014). On the other hand, *cis*-regulatory elements can play a role, such as in the repeated modification of spots in *Drosophila* (Prud'homme et al. 2006). The question is, thus, whether changes in *cis*- and *trans*-regulatory elements have similar or different effects on evolvability, thereby driving innovation (Wittkopp et al. 2004). A related question is whether there is a general principle or whether the rela-

tive contribution of *cis*- and *trans*- regulation to innovation is system specific (Ordway et al. 2014; Camino et al. 2015; Hu et al. 2015). 2) How does the network build up by *cis*- and *trans*- regulatory elements? In some cases, very few nucleotide mutations are sufficient for building up new gene interactions, such as the distinct combinations of binding sites of CREs in genes *yellow* and *tan* leading to different response of the *trans*-landscape to exhibit contrasting expression of male-specific abdominal dimorphic pigmentation in *Drosophila* (Camino et al. 2015). Also, transposable element (TE) insertion can provide ready-to-use novel TF binding sites thus incorporating new regulatory changes into an existing gene network (Feschotte 2008). 3) Which part of the gene network (internal nodes, eg. master regulatory genes, or terminal nodes, such as effector genes) is more readily rewired? For example, the divergent expression of TF d*istalless* can control gene expression of *yellow, ebony* and *tan* for the divergent spot morphologies, but the whole body coloration of *Drosophila* resulted from independent changes in multiple loci, rather than changes in a single "master regulator" (Ordway et al. 2014). Besides, changes of terminal effector genes might not need the whole gene network to be rewired (e.g. point mutation in *MC1R* (Hoekstra et al. 2006)).

In addition to know how gene networks underlying evolutionary innovation are built up, another open question is how do gene networks underlying novel traits emerge. Do they largely rely on a pre-existing gene network that is only slightly modified, or do they originate *de novo*? And to what extend a gene network is "innovative" compared to the ancestral gene network? How will it affect the evolvability of a trait? The so-called "Christmas tree model" (Wagner and Lynch 2008) suggests that the ancestral *trans*-regulatory landscape was re-used for *Drosophila* wing pigment pattern formation. According to this model, different *trans*- regulatory elements bind to *cis*- regulatory elements, just like Christmas tree decoration. In this case, the gene network of wing pigment pattern is still tightly connected to the ancestral wing developmental gene network, which might limit its evolvability. In nymphalidae butterfly eyespots, on the other hand, different TF insertion form their own cross-regulatory network, which appears to be quasi-independent to the ancestral wing developmental network (Shirai et al. 2012). This independence could be causally linked to the diversity of eyespots patterns in butterflies (Wagner 2014). Therefore, the extent of independence of a gene network might be causally linked to the evolvability of a trait (Wagner 2014). To test this hypothesis, it is important to examine a system, which includes different potential of phenotypic novelties evolvability but with close related genetic background.

Here, the origin of an evolutionary innovative pigmentation patterns on the anal fin in East African cichlid fish is an ideal model, considering that this system has the advantages that the innovation evolved relatively recently (Salzburger et al. 2005; Santos et al. 2014; Meyer et al. 2015), so that the noise of individual evolutionary history and stochasticity effects can be reduced (Faria et al.

2014); in addition, the anal fin pigmentation patterns have evolved repeatedly in several East African cichlid lineages but with different evolvability. Perhaps the most famous example of fin pigment patterns in cichlids are the so-called 'eggspots' on the anal fins of the haplochromines, the most species-rich lineage in East African cichlids (Salzburger 2009; Santos et al. 2014). Eggspots are circular pigmentation patterns with a transparent outer ring (Santos et al. 2014) (Fig. 1). The different haplochromine species show a great diversity of eggspots phenotypes, differing in colours, sizes and numbers. It has been suggested that eggspots might mimic real eggs to attract females (Wickler 1962), and be involved in male-male competition (Theis et al. 2012; Theis et al. 2015) and species recognition (Couldridge 2002). Ancestral haplochromine species feature another fin pigmentation trait in form of blotch, which is reddish with ill-defined boundary (Santos et al.) (Chapter 2). Anal fin pigmentation patterns have independently evolved in another cichlid lineage, the ectodines from Lake Tanganyika (Salzburger et al. 2007). The anal fin blotch of one genus, *Callochromis*, is similar to the blotch of ancestral haplochromines (Fig. 1). The ectodine blotch was also suggested to be related to female attraction (Fryer and Iles 1972), although is less investigated. Noticeably, unlike haplochromine eggspots, the ectodine blotch shows almost no variation among species.

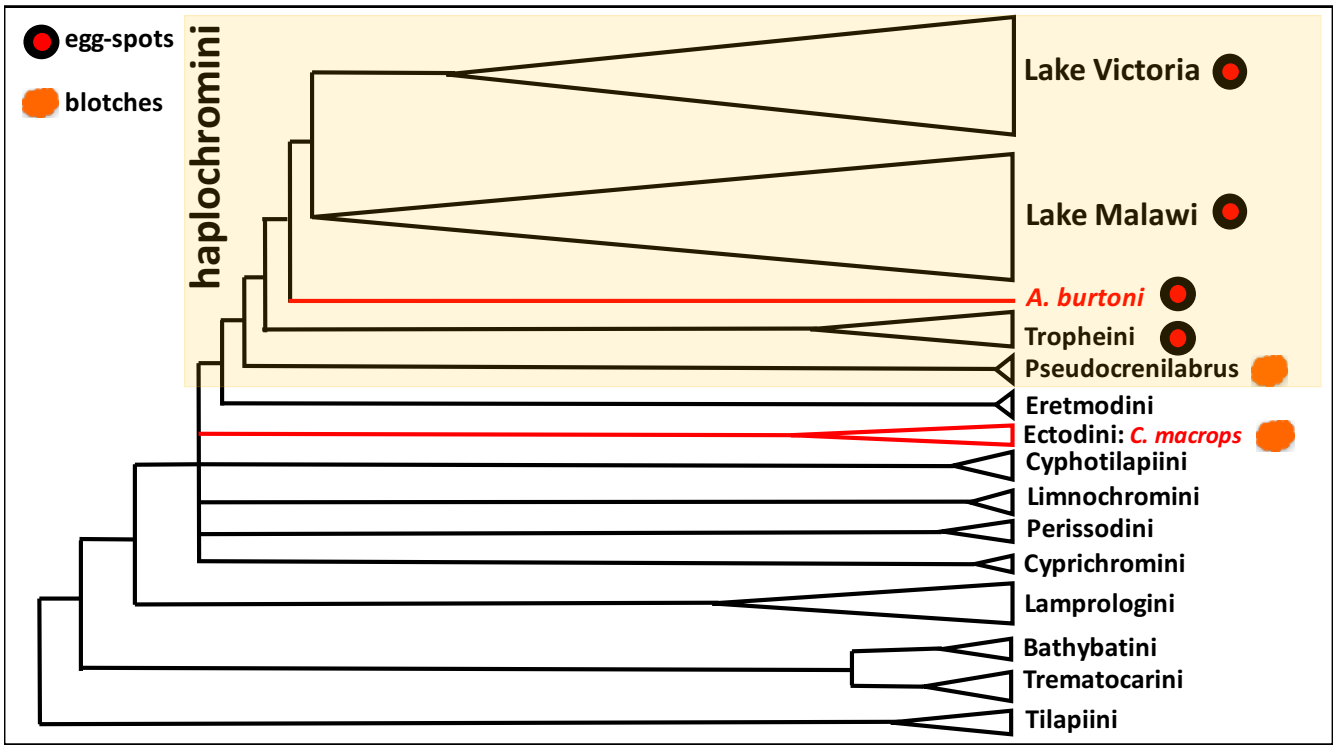

Fig. 1 Convergent evolution of anal fin pigmentation patterns in East African cichlid fish. Schematic molecular phylogeny of the East African cichlid fishes based on combined evidence from Salzburger et al., 2005; Meyer et al., 2015; Takahashi and Sota, 2016. Triangle symbol represented species richness based on studies from Salzburger et al., 2005 and https://en.wikipedia.org. Names on the right side indicated rivers or tribes.

Several studies have focused on the genetic basis of haplochromine eggspots. For example, Santos et al. (Santos et al. 2014) have shown that a *cis*-regulatory change in the form of a TE insertion in the upstream region of gene *fhl2b* appears to be causally related to the eggspots phenotype. Experiments with transgenic zebrafish suggested that this *cis*-regulatory change drives expression in a certain type of pigment cells, iridophore (Santos et al. 2014). However, since the *trans*-landscape might be different between zebrafish and cichlids, it is still unclear how *fhl2b* affect the emergence of eggspots and whether *cis*- or *trans*- of *fhl2b* is responsible for the formation of eggspots. Another study found that coding sequence change in a xanthophore related gene, *csf1ra,* was highly expressed in two independently evolved fin pigment patterns (haplochromine eggspots and ectodine tassel) (Salzburger et al. 2007). The same study also identified positive selection in the ancestral branch of Haplochromine lineage, which might causally link haplochromine eggspots to the radiation of this lineage. Recently, based on differential gene expression profiles of 46 candidate genes among haplorchomine eggspots, haplochromine blotch and ectodine blotch, Santos et al., (Santos et al.) (Chapter 2) suggested that these three types of anal fin pigmentation patterns might not share a common genetic basis. However, based on a thorough comparative transcriptomic data analysis between eggspots and ectodine blotch (Gu et al., in preperation) (Chapter 3) showed that at least 15.7 % of the ectodine blotch related candidate genes showed similar expression patterns as eggspots, suggesting that they could, at least partly, share a common gene network. However, what are the roles of these shared and unshared genes in these innovative pigment patterns remains to be tested.

Therefore, we mainly address the following questions in this study: 1) which changes rewired the gene network during the evolution of the anal fin pigment pattern? For example, are they from pre-existing gene regulatory network or *de novo* originated? 2) What are the mechanisms behind the different evolvability of these anal fin pigment traits (eg. eggspots are with large variations, but blotches are not)? To answer these questions, we first investigated 40 sequences evolution of the ectodine blotch related candidate genes based on Gu et al. (in preperation) (Chapter 3). To this end, we characterised the whole gene region of 40 blotch related candidate genes including upstream and coding region across the phylogeny, and we compared sequence similarity among species with blotch/eggspots/without anal fin pigmentation patterns across the phylogeny. We then examined the sequences further by applying methods for conserved noncoding element (CNE) detection, TF binding prediction and positive selection analyses. Our hypothesis are that 1) eggspots, but not the ectodine blotch, might have a much more independent gene network from the ancestral anal fin gene network, thus providing a clue of the mechanism about its higher evolvability. 2) Eggspots could have different roles as ectodine blotch, and selection might have a stronger role in eggspots.

Methods and Materials

1. Gene characterization across the phylogeny

To examine how the 40 ectodine blotch related candidate genes from Gu et al., (in preparation) (Chapter 3) evolved during evolution, we tried to amplify the whole gene region including upstream regions using long polymerase chain reaction (PCR) amplification and Ion Torrent next generation sequencing technology. We tried to amplify these genes in one species of each tribe (totally 16 species) according to the phylogenetic tree built by (Meyer et al. 2015). Together with four available cichlid genomes (*A. burtoni, N. brichardi, P. nyererei, O. niloticus*) we could thus study sequence evolution in a total 20 species representing the phylogenetic spectrum of cichlids in East Africa (Fig.1).

Laboratory strains of *C. macrops, A. burtoni, Haplochromis cf. stappersii, Labidochromis careulus, Cyphotilapia frontosa, Cyprichromis leptosoma and Pseudocrenilabrus philander* were kept at the University of Basel (Switzerland) under standard conditions (12h light/12h dark; 25°C) and permit 1010H issued by the cantonal veterinary office Basel. Before dissecting tissues for DNA and RNA extraction, all specimens (males) were euthanized with MS222 (Sigma-Aldrich, USA) following an approved procedure (permit no. 2317 issued by the cantonal veterinary office Basel). Individuals of all other specimens (*Lobochilotes labiatus, Eretmodus cyanostictus, Ophthalmotilapia ventralis, Grammatotria lemairii, Xenotilapia spiloptera, Greenwoodochromis christyi, Perissodus microlepis, Lepidiolamprologus elongatus, Trematocara nigrifrons, Boulengerochromis microlepis*) were collected at our field-site in the South of Lake Tanganyika and processed in the field following our standard operating procedure (Muschick et al. 2012). Field work was covered by permits issued by the Department of Fisheries, Republic of Zambia.

Genomic DNA was extracted with the DNeasy Blood & Tissue kit (Qiagen) following the manufacturer's protocol. Isolation of RNA was performed according to the TRIzol protocol (Invitrogen, USA) after incubating the dissected tissue in 750µl of TRIzol at 4 °C overnight. Tissue samples were then homogenized with a Bead-Beater (FastPrep-24; MP Biomedicals, France). Subsequent DNase treatment was performed with the DNA-Free kit (Ambion, place). RNA quantity and quality was determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA); cDNA was synthetized using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA).

PCR primers were designed using Primer premier 5 (Lalitha 2000) to amplify the whole gene region including upstream, exons and introns if the total length was less than 10kb. When the length was longer than 10kb, we designed more than one pair of PCR primers to amplify different parts of the gene. Primer sequences are available in Supplementary Table 1. High-Fidelity PCR

Master Mix with HF buffer (BioLabs) and a touch-down annealing process was used for the amplification (Supplementary Table 1). PCR products were visualized with electrophoresis in a 1% agarose gel using GelRed (Biotium, USA) under the following run conditions: 100 V, 30 min. Prior to library construction, we purified the correct size fragments from the gel using GenEluteTM PCR Clean-up kit (Sigma). To sequence all these long amplicons, PGM Ion Torrent (www.lifetechnologies.com) with protocal *Prepare Amplicon Libraries Requring Fragmentation Using the Ion Xpress™ Plus Fragment Library Kit* (Life Technologies) was used. We barcoded each species for all the genes and the final pooled barcoding library was sequenced using CHIP 316 (Life Technologies) with 400bp length. We used FASTQ Quality Trimmer by sliding window in Galaxy (https://usegalaxy.org/) to trim low quality reads in both 5' and 3' end, window size 2, Step size 1, and the minimum quality score was set to 20. After quality trimming, we used Filter FASTQ to filter the short reads with minimum size is 40bp. The software Geneious (http://www.geneious.com, (Kearse et al. 2012)) was used to do *de novo* assemble to produce consensus sequences for each gene in each species. Medium Sensitivity/Fast in Custom Sensitivity with default parameters was used for the *de novo* assembly, and the maximum gap size was set to 20,000 bp to expand reads across introns. We extracted the targeting corresponding gene sequences in tilapia from Ensemble as a reference (Supplementary Table 2), and annotated the assembled consensus sequences by mapping them to the reference using Geneious v 8.1.3 (http://www.geneious.com, (Kearse et al. 2012)). Afterwards, we mapped the original reads to the annotated consensus sequences to recheck the assembled sequences.

In case of smaller gaps or low coverage produced by Ion Torrent sequencing, we fixed them with Sanger sequencing. Primers for Sanger sequencing were designed with software Primer premier 5 (Lalitha 2000) (Supplementary Table 1). High-Fidelity PCR Master Mix with HF buffer (BioLabs) and a touch-down annealing process was used for the amplification (Supplementary Table 1). PCR products were visualized with electrophoresis in a 1.5% agarose gel using GelRed (Biotium, USA). PCR products were enzymatically cleaned with ExoSAP-IT (Affymetrix, USA) and sequenced with BigDye 3.1 Ready reaction mix (Applied Biosystems) after BigDye Xterminator purification (Applied Biosystems) on an AB3130*xl* Genetic Analyzer. Sequences were corrected, trimmed and aligned manually in CODONCODE ALIGNER (http://www.codoncode.com/aligner/, CodonCode Corporation).

2. Adaptive sequence evolution detection

To see whether adaptive sequence evolution happened in specific lineages, we tested for positive evolution for each amplified gene. Considering phylogenetic and evolutionary inference

can be severely misled if recombination is not accounted for, we screened the sequence alignment using GARD implemented within Hyphy before running positive selection tests (Kosakovsky Pond et al. 2006). The unrooted phylogenetic tree used in PAML (Yang 1997; Yang 2007) was modified according to the phylogenetic tree (since the samples of Limnochromis abeelei and Bathybates graueri are not available, we replaced them with species *G. christyi* and *T. nigrifrons* in the same tribe, respectively) obtained based on a genomic multi-marker approach (Meyer et al. 2015). All the model comparisons in PAML (Yang 1997; Yang 2007) were fixed with branch length (fix_blength=2) derived under M0 model in PAML (Yang 1997; Yang 2007). Alignment gaps and ambiguity characters (eg. Y for T or C) were not removed (Cleandata=0). Codon frequencies were approximated using the F3x4 calculation. Different starting parameters were used to run each model multiple times to avoid local optima.

The branch-site model was used to test positive selection affecting a few sites along particular lineages (foreground branches). Comparisons were made between the modified model A (model=2 NSsites=2) with corresponding null model with ω_2=1 fixed (fix_omega=1 and omega=1). A likelihood ratio test (LRT) was then used to calculate a chi-square approximation, and p/2 value was used considering mixture distribution (see PAML manual). The Bayes empirical Bayes (BEB) was used to identify which sites are under positive selection. The foreground labeled branches see Supplementary Table 3.

3. Conserved noncoding elements (CNE) detection and TF binding prediction

To see whether there are eggspots or blotch specific CNEs, we extracted the available corresponding genome sequences from Ensembl database of cichlids (*A. burtoni, N. brichardi, P. nyererei, O. niloticus*), and analyzed them together with the sequences from our Ion Torrent sequencing. All the sequences were aligned using software Geneious v 8.1.3 (http://www.geneious.com, (Kearse et al. 2012)) and mapped onto the phylogeny (Supplementary Fig.1). To see whether the lineage specific CNEs are transposable elements, RepeatMasker web server (http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker) was used. To detect the possible TF binding sites in these CNEs, and to see whether common TFs are shared among these CNEs, TF binding analysis was predicted with software MatInspector within genomatix software suite (https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html) using default parameters with zebrafish as the reference.

Results

1. Gene characterization across the phylogenetic tree

Ion Torrent sequencing produced ~ 120,000 raw reads and after trimming and filtering, ~100,000 reads were left (exact reads numbers see Supplementary Table 4). Raw reads are available from the Sequence Read Archive (SRA) at NCBI under the accession number SRA…. In total, 40 genes were successfully sequenced for downstream data analysis. For details about the lengths of the sequenced individual genes, the taxa sequenced, missing data, etc. see Supplementary Fig. 1.

Among these 40 genes, 16 genes showed similar expression pattern (*tpd52, fhl2b, 13024, 13044, 10298, 7947, 15173, 26203, fhl2a, cecr, crip1, gpnmb, pnp4a, trpm1a, trpm1b, apod*) in both eggspots and blotch; Two genes showed opposite (*23698*, *pnp5a*) expression pattern in eggspots and blotch (Gu et al., in preparation) (Chapter 3). Four genes (*fhl2b, 13024, 26203, pnp4a*) showed specific CNE pattern segregating with eggspots phenotype (Fig.2), and all of them showed similar expression pattern in both eggspots and blotch (Gu et al., in preperation) (Chapter 3). For example, in accordance with Santos et al., (2014), gene *fhl2b* showed eggspots specific SINE insertion. Furthermore, our study found that there is a LINE insertion in species without anal fin pigmentation in the corresponding same region as SINE in eggspots species (Fig. 2). Interestingly, species *P. philander* with blotch which is homologous to eggspots only has part of SINE, and the sequence in *C. macrops* with blotch only showed part of the LINE sequence. While the basel lineages (*T. nigrifrons, B. microlepis, O. niloticus*) showed neither eggspots specific SINE nor non-eggspots specific LINE insertion in the upstream region (Fig. 2). Another gene, *26203*, which is an unannotated teleosts-specific gene according to Ensembl database also showed LINE (3.5kb) insertion in the upstream region (<2kb from exon1). Instead of the whole LINE region, only several nucleotide mutations within the LINE (less than 2kb from exon1) were segregating with eggspots (Fig. 2) and *P. philander* with the blotch, while no specific pattern for *C. macrops* with the blotch and other species without anal fin pigmentation (Fig. 2). Gene *13024* was also found specific TE insertion in the upstream region (less than 5kb from exon 1) segregating with eggspots. While in gene *pnp4a*, an eggspots specific CNE deletion instead of insertion was found in intron 1 (ca. 200bp) (Fig. 2). Noticeably, all these genes showed no ectodine blotch specific CNE pattern. Genes *tmem, cecr, zdhhc14* might also show eggspots specific CNE pattern, but need further sequences confirmation from species *L. labiatus* and *P. philander* (Supplementary Fig. 2).

Six genes (*26203, crip1, trpm1a, 23711, 23698, zdhhc14*) showed eggspots specific non-synonymous amino acid substitution (*26203*, Alanine-eggspots Serine at nucleotide position 112; *crip1*, Leucine, Methionine, Valine- eggspots Glycine at nucleotide position 254; *trpm1a*, Proline-eggspots Alanine at nucleotide position 67; *23711*, Aspartic Acid-eggspots Asparagine at nucleotide postion 1531; *23698*, Glycine,Aspartic Acid, Arginine, Alanine-eggspots Serine at nucleotide position 40; *zdhhc14*, Alanine-eggspots Valine at nucleotide position 1451) (Supplementary Fig. 1),

and three of them (*26203, crip1, trpm1a*) showed similar expression pattern in both eggspots and ectodine blotch (Gu et al., in preperation) (Chapter 3). Genes *prtfdc* (Proline-eggspots Threonine/Alanine at nucleotide position 529) and *pnp5a* (Aspartic Acid-eggspots Glutamic Acid at nucleotide position 465) might also have eggspots specific non-synonymous amino acid substitution, but need further confirmation by fixing the gap of missing data from *L. labiatus* and *P. philander* (Supplementary Fig. 1). Noticeably, gene *23698* and *pnp5a* showed opposite expression patterns between eggspots and the blotch, with highly expression in the blotch, but low in eggspots (Gu et al., in preperation) (Chapter 3). Besides, seven genes (*15173, gatsl*, *pmelb, pnp4a, trpm1a, myoc, vtna*) showed ectodine blotch specific non-synonymous amino acid substitution (Supplementary Table 2). In addition, many genes showed haplochromine blotch specific substitution (*vegfc, 7947, slc23a1, 13044, 15173, 23698, 23711, cecr, crip1, gatsl, gpnmb, mc1r, pnp4a, prickle, rab27bb, rgs, ttyh, zdhhc14, myoc, apod, vtna*) (Supplementary Table 2).

Noticeably, it could be that these eggspots segregations including upstream region and coding region are due to phylogenetic signals (Meyer et al., 2015) after the divergence between *Pseudocrenilabrus* and the common ancestor of other haplochromines. However, since only a very small portion of genes exhibited the eggspots segregating pattern compared to a large portion of genes showed similar pattern in haplochromine lineage including *P. philander*, the chance that they are produced only by phylogenetic signal is low. Besides, most sequences between *P. philander* with the blotch and *L. labiatus* with eggspots are much similar, which could further support that these two species are clustered together in the phylogeny, so that *L. labiatus* could make a control to rule out the phylogenetic effect.

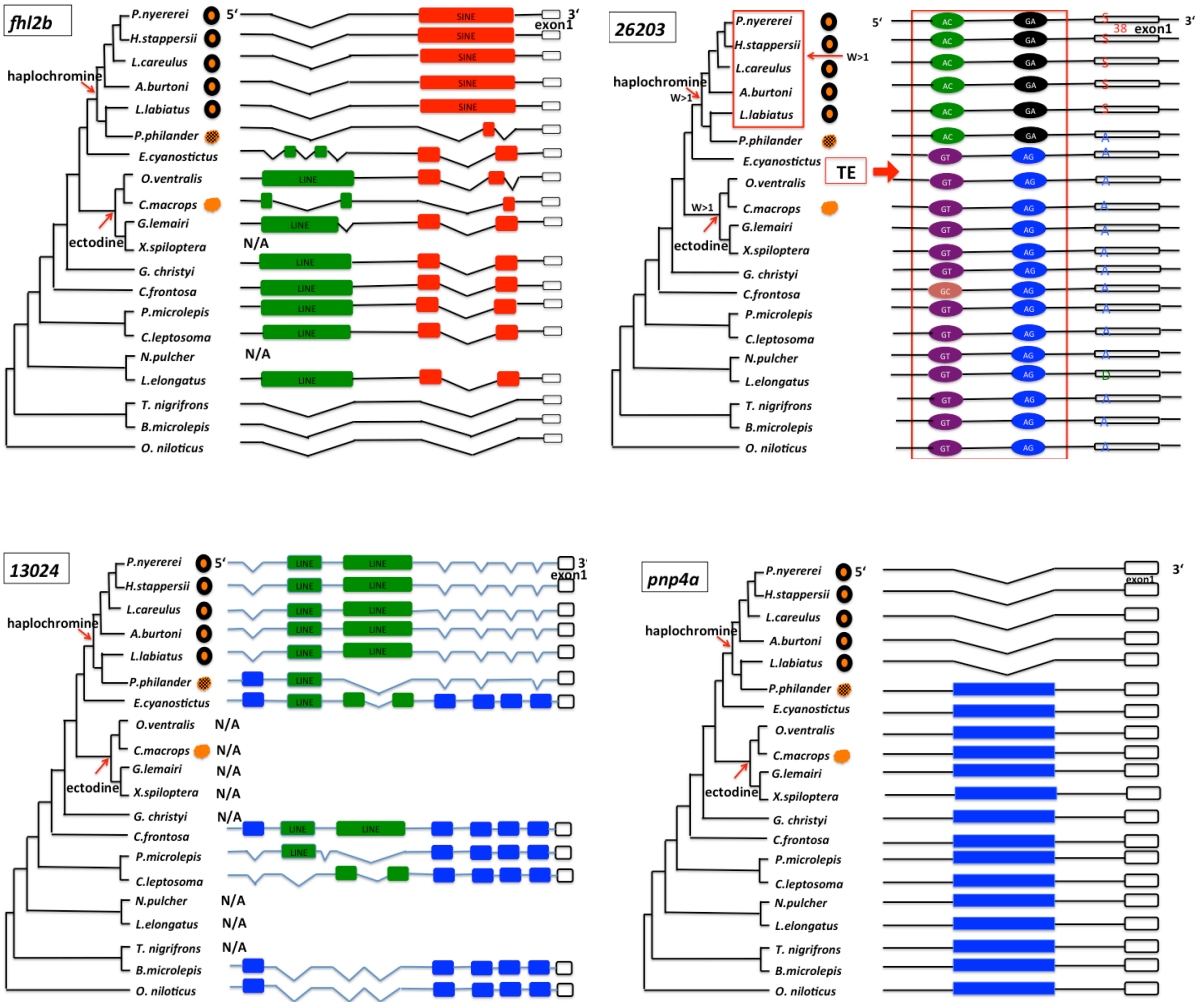

Fig. 2 Genes (*fhl2b, 26203, 13024, pnp4a*) with eggspots specific CNE pattern.

2. Adaptive sequence evolution

Genes *15173, prtfdc, zdhhc14,* and *26203* were found to have evolved under positive selection when species with eggspots were labeled as foreground branch, and their eggspots specific substitution sites were also found under positive selection. Two genes *2365* and *impdh* were found under positive selection in the branch of *C. macrops*, coinciding with the ectodine blotch. One gene *vtna* showed positive selection in *P. philander,* coinciding with the haplochromine blotch (Supplementary Fig. 1). Genes *prtfdc, 13044, 15173, 26203, trpm1b, zdhhc14* were found under positive selection in haplochromine branch. Details see Supplementary Fig. 1 and Supplementary Table 3.

3. TF binding prediction

Considering upstream CNE might have regulatory roles as *cis*-regulatory element in the genes, we predicted the TF binding sites in the eggspots specific CNE regions in genes *fhl2b, 13024, 26203 and pnp4a* mentioned above (Supplementary Table 5). Interestingly, all these eggspots specific CNEs showed eggspots specific TF binding sites (Supplementary Table 5) with many common shared TF binding sties, such as estrogen and hormone related TFs, homeobox, etc. No-

ticeably, in the upstream TE insertion in gene *26203*, the two eggspots specific nucleotide substitution were suggested to have specific binding affinities to Hmx2/Nkx5-2 homeodomain TF (Hmx2) and Glucocorticoid modulatory element binding protein 2 (GMEB2) (Supplementary Table 5). *P. philander* has similar sequences as the species with eggspots, but with one different nucleotide mutation (C to A) in the corresponding region. Therefore, it also showed TF binding site for Hmx2 as eggspots, but not for GMEB2, but for its specific Hypoxia response elements binding site. And Glucocorticoid related TF binding site was also found in eggspots specific SINE in gene *fhl2b*. Interestingly, the repressive binding sites for glucocorticoid (glucocorticoid receptor (IR2)) was found in the non-eggspots LINE in gene *fhl2b* (Supplementary Table 5).

Discussion

Although eggspots and blotches are both innovative pigmentation patterns, their evolvability and variability are quite different. The eggspots phenotype emerged once in the most species rich lineage of cichlids, the haplochromines, with the different species exhibiting quite different numbers, colours and sizes of eggspots. However, only several species in genus *Callochromis* in the independently evolved ectodine lineage exhibit blotch patterns, with much less phenotypic variation. What are the mechanisms behind their different evolvability? Our study suggested that the intrinsic developmental gene network and extrinsic factors, such as selection for eggspots might be responsible for this discrepancy.

Intrinsic factor for the different evolvability between eggspots and the ectodine blotch

Our results showed that 22.5% (9/40) blotch related candidate genes showed eggspots specific sequence pattern, especially for CNE sequences in the upstream region (*fhl2b, 13024, 26203*) and intron region (*pnp4a*), but also included the coding region (*26203, 23711, 23698, crip1, trpm1a, zdhhc14*) (Supplementary Fig. 1). Among them, six genes (*fhl2b, 13024, 26203, pnp4a, crip, trpm1a*) were both highly expressed in eggspots and blotch, and gene *23698* showed opposite expression in these two innovative pigmentation patterns (Gu et al., in preperation) (Chapter 3). These genes will be good eggspots candidate genes for further study. Noticeably, all the four genes (*fhl2b, 13024, 26203* and *pnp4a*) with eggspots specific CNE pattern we found here were both highly expressed in eggspots and the blotch (Fig. 1, Gu et al., in preperation) (Chapter 3). Interestingly, no blotch specific CNE pattern was found among these 40 candidate genes.

Based on these results, we proposed the first hypothesis to explain the different evolvability between eggspots and the blotch, the “eggspots>blotch=fin” hypothesis. This hypothesis suggests that the gene network underlying the eggspots phenotype could be much more independent than

blotch when compare to the ancestral anal fin gene network. The evidence comes from the results that 10% (4/40) genes we test here showed eggspots specific CNE patterns, including TE insertion (*fhl2b, 13024, 26203*) and CNE deletion (*pnp4a*), and all of them were found their own specific TF binding sites based on prediction (Supplementary Table 5). Besides, the eggspots CNE shared common TFs binding sites, such as Glucocorticoid related TF, thyroid hormone related TF, homeobox, etc (Supplementary Table 4). It has been shown that thyroid hormone could play a role as a morphogen for stripe pattern formation in zebrafish (McMenamin et al. 2014). The roles of steroid and homeobox TFs in eggspots formation will be interesting to study in the future. Besides, TE insertion could play a role as "ready to use" element to build new gene network (Feschotte 2008), such as in the case in mammalian evolution (Lynch et al. 2011). The eggspots specific SINE in *fhl2b* has already been shown to be related to iridophore in zebrafish (Santos et al. 2014). The common TFs of eggspots specific CNE among genes *26203*, *fhl2b* and *13024* might have functions as up-regulation master genes to control several effector genes, similar as the role of TF *Distalless* in controlling downstream effector genes (Arnoult et al. 2013). Similarly, point mutation within coding region could also be involved in gene network by changing binding opportunities. On the other hand, no CNE specific to the ectodine blotch was found. Instead, it seems that sequences of blotch in *C. macrops* are much more similar with species without anal fin pigmentation. Therefore, new independent gene network might not have built up for the blotch phenotype. And indeed, Gu et al., (in preperation) (Chapter 3) found that several blotch were related candidate genes related to fin formation.

However, we cannot rule out the other two possibilities. One is that 2) "eggspots=blotch=fin", which means that both eggspots and the blotch are still tightly linked to the ancestral anal fin gene network. For example, although the eggspots specific CNEs were suggested to bind to specific TFs based on TF binding prediction, they still share large amounts of TFs with species without anal fin pigmentation patterns. And the substitutions within coding region might be related to the terminal effector genes, such as in pigmentation related genes, *trpm1a* (melanophore) (Braasch et al. 2009). Therefore, the core gene network in eggspots might still depend on the ancestral anal fin gene network. The other possibility is that 3) "eggspots=/>blotch>fin", which means that both eggspots and blotch are relatively independent to the ancestral anal fin gene network. The evidence comes from the observation that seven genes (*pmelb, 15173, gatsl, pnp4a, trpm1a, zdhhc14, myoc*) out of these 40 genes showed blotch specific amino acid substitutions (Supplementary Table 3).

The question whether or not the gene networks are independent is important is because that it could affect the evolvability of the innovative trait (Wagner 2014). For example, the 'Christmas

tree model' (Wagner and Lynch 2008) predicts that the origination of wing spot in *Drosophila* occurred by making use of the pre-existing ancestral *trans*-regulatory landscape, which means that the wing spot is tightly integrated into the other part of the wing. In this case, any change of the spot might affect the development of other aspects of the wing, which limit its evolvability and variation (Wagner 2014). In contrast, the relatively independent cross-regulatory network in the eyespots of nymphalid butterflies might be more flexible without direct influencing of the ancestral wing (Shirai et al. 2012), so that it can evolve different sizes, shapes and positions on the wing (Wagner 2014). If our hypothesis of 'spot>blotch=fin' is correct, then this might explain the different evolvability between eggspots and the ectodine blotch by acquiring developmental individuality. Further functional experiment will be helpful to have a deep understanding of it.

The effect of downstream selection in eggspots and the ectodine blotch

Intrinsic developmental mechanisms are important for the origin of innovation (Bright et al. 2016). However, the maintenance of innovations in a clade might also be related to selection. Six genes (*prtfdc, 15173, 26203, gatsl, rab27a* and *zdhhc14*) were under positive selection in the branches of species with eggspots, and among them there are two genes (*15173, 26203*) which were highly expressed in eggspots and the ectodine blotch. Noticeably, the eggspot specific amino acid substitution in gene *26203 and zdhhc14* were found to be under positive selection (Supplementary Table 2), which might causally link eggspots to its radiation. Since eggspots were supposed to be related to female attraction (Wickler 1962) and male-male competition (Theis et al. 2012; Theis et al. 2015), their formation might have fitness advantages, such as to increase fertilization rate in the water environment and/or act as a sexual advertisement. Alternatively, the mutations might be responsible for other traits, and only indirectly link to eggspots formation. In this case, the selection of eggspots would be the by-product of other traits. Previously, Salzburger et al. (Salzburger et al. 2005) have shown that *csf1ra* experienced a period of positive selection in the ancestral branch of haplochromine lineage, and suggested that this could causally link eggspots as the key innovation to the adaptive radiation of this species-rich lineage. We also found that six genes (*prtfdc, 13044, 15173, 26203, trpm1b, zdhhc14*) and among which, two genes (*trpm1b and zdhhc14*) with haplochromine specific substitutions were under positive selection in the haplochromine lineage (Supplementary Table 3, Supplementary Fig. 1). Noticeably, only three genes were under positive selection in *C. macrops* (Supplementary Table 3). Actually, the role of anal fin pigmentation was only tested in *A. burtoni* and *A. calliptera,* but not in other species (Egger et al. 2011; Theis et al. 2012; Theis et al. 2015). Therefore, anal fin pigmentation in different species and environments might have different roles and thus might evolve under different selection regimes. And although the ectodine blotch was suggested to be related to female attraction (Fryer and Iles 1972), less studies

were done. And the haplochromine blotch, which shows similar patterns as the ectodine blotch, was shown to not relate to female attraction, but to have a role as male sexual signals (carotenoids-based color) (Egger et al. 2011). Therefore, the roles of anal fin pigmentation in more cichlid fish needs to be investigated.

Clues for genetic basis of two blotches——the ectodine blotch and the haplochromine blotch

Although both the haplochromine blotch and the ectodine blotch have similar patterns (reddish color with ill-identified boundary) (Fig. 1), they might originate from different genetic basis based on our results, which also support the conclusion from Santos et al., (submitted) (Chapter 2). Evidence comes from 1) no specific similar sequence patterns of candidate genes were found for both the ectodine blotch and the haplochromine blotch. 2) According to the phylogeny built by (Meyer et al. 2015), the haplochromine blotch is homologous with eggspots. Indeed, many genes showed haplochromine specific pattern including CNEs and coding region (Supplementary Fig.1 and 2; Supplementary Table 2). For example, it has been shown for *fhl2b* that the eggspots specific SINE insertion has a role related to iridophore (Santos et al., submitted) and we found that species with non-anal fin pigmentation showed a specific LINE insertion. While followed by sequence alignment, the upstream region of *P. philander* only has a small portion overlapping with eggspots specific SINE. And the upstream of *C. macrops was* partially overlapped with the non-anal fin pigmentation LINE (Fig. 2). One possibility is that the eggspots SINE and non-anal fin pigmentation LINE could be antagonistic to each other, and the formation of the blotch in haplochromine could be due to part of eggspots specific enhancer missing, while the origination of the ectodine blotch could be due to part of ancestral silencer missing (Fig. 2), but further functional experiments are needed. A similar case was found in the loss of male-specific pigmentation in *Drosophila*, which is due to the inactivating mutations of a silencer element (Johnson et al. 2015). Other genes in our study also showed similar upstream patterns and amino acid substitutions between eggspots and haplochromine blotch (Supplementary Fig. 1, Supplementary Table 2). However, we cannot rule out the possibility that it is only because of phylogenetic signal, and haplochromine blotch might have evolved independently as Santos et al. suggested (submitted) (Chapter 2). For example, the TE insertion of gene *26203* showed haplochromine blotch specific nucleotide mutation, which showed its own TF binding sites for Hypoxia response elements based on prediction. And many genes showed haplochromine blotch specific amino acid substitutions (Supplementary Table 2). Further gene expression profile for the haplochromine blotch is needed.

Conclusion

Based on next generation sequencing and a series of analysis, our study revealed several eggspots candidate genes (*fhl2b, 26203, 13024, trpm1a, crip1, zdhhc14, 23711, 23698*) for down-stream analysis. Furthermore, we propose three hypothesis of the gene network relationships among eggspots, the ectodine blotch and anal fin: 1) "eggspots>blotch=fin"; 2) "eggspots=blotch=fin"; 3) "eggspots=/>blotch>fin". It could be that eggspots, but not the ectodine blotch, might have a much more independent gene network from the ancestral anal fin gene network, thus providing a clue of the mechanism about the different evolvability between lineages with egg spots and the ectodine blotch. Besides, eggspots could have a different role as the ectodine blotch, and selection might have a stronger role in eggspots. In addition, the ectodine blotch and haplochromine blotch might have originated independently. Further functional experiment, CHIPseq and gene expression profile in a more thoroughly species including species with and without anal fin pigmentation patterns are needed.

Acknowledgements

We would like to acknowledge the sampling help from Attila Rüegg, Adrian Indermaur, Yuri Kläfiger and Bernd Egger; Many thanks to Brigitte Aeschbach for the lab management and Lukas Zimmermann for the IT support. This project was fund by University of Basel, Switzerland, the Swiss National Science Foundation (SNSF) and the European Research Council (ERC).

Author Contributions

LG and WS designed the experiment and wrote the manuscript. LG did long pcr amplification for the 40 genes; the Ion Torrent library construction with the help from ZM and NB, and all the data analysis. NS did sanger sequencing for coding region of genes *pnp4a and myoc*, exon 2 and 3 of gene *2365*, exon 1 of gene *gpnmb*, exon1, 2, 3 of gene *7947 and* exon1 of gene *rab27a.*

References

Arnoult L, Su KFY, Manoel D, Minervino C, Magriña J, Gompel N, Prud'homme B. 2013. Emergence and diversification of fly pigmentation through evolution of a gene regulatory module. Science. 339:1423–1426.

Bhullar B-AS, Morris ZS, Sefton EM, Tok A, Tokita M, Namkoong B, Camacho J, Burnham DA, Abzhanov A. 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. 69:1665–1677.

Braasch I, Liedtke D, Volff J-N, Schartl M. 2009. Pigmentary function and evolution of tyrp1 gene duplicates in fish. Pigment Cell Melanoma Res. 22:839–850.

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.

Camino EM, Butts JC, Ordway A, Vellky JE, Rebeiz M, Williams TM. 2015. The evolutionary

origination and diversification of a dimorphic gene regulatory network through parallel innovations in cis and trans. PLoS Genet. 11:e1005136.

Couldridge VCK. 2002. Experimental manipulation of male eggspots demonstrates female preference for one large spot in Pseudotvopheus lombavdoi. J. Fish Biol. 60:726–730.

Egger B, Klaefiger Y, Theis A, Salzburger W. 2011. A sensory bias has triggered the evolution of egg-spots in cichlid fishes. PLoS One. 6:e25601.

Faria R, Renaut S, Galindo J, Pinho C, Melo-Ferreira J, Melo M, Jones F, Salzburger W, Schluter D, Butlin R. 2014. Advances in Ecological Speciation: an integrative approach. Mol. Ecol. 23:513–521.

Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. Nat. Rev. Genet. 9:397–405.

Fryer G, Iles T. 1972. The cichlid fihses: their biology and evolution. edinburgh, uk.

Glassford WJ, Johnson WC, Dall NR, Smith SJ, Liu Y, Boll W, Noll M, Rebeiz M. 2015. Co-option of an Ancestral Hox-Regulated Network Underlies a Recently Evolved Morphological Novelty. Dev. Cell. 34:520–531.

Green SA, Simoes-Costa M, Bronner ME. 2015. Evolution of vertebrates as viewed from the crest. Nature. 520:474–482.

Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 313:101–104.

Hu P, Liu M, Zhang D, Wang J, Niu H, Liu Y, Wu Z, Han B, Zhai W, Shen Y, et al. 2015. Global identification of the genetic networks and cis-regulatory elements of the cold response in zebrafish. Nucleic Acids Res. 43:9198–9213.

Johnson WC, Ordway AJ, Watada M, Pruitt JN, Williams TM, Rebeiz M. 2015. Genetic Changes to a Transcriptional Silencer Element Confers Phenotypic Diversity within and between Drosophila Species. PLoS Genet. 11:e1005279.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kitano H. 2004. Biological robustness. Nat. Rev. Genet. 5:826–837.

Kosakovsky Pond LS, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics. 22:3096–3098.

Lalitha S. 2000. Primer Premier 5. Biotech Software&Internet Rep. 1.

Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 43:1154–1159.

Martin A, McCulloch KJ, Patel NH, Briscoe AD, Gilbert LE, Reed RD. 2014. Multiple recent co-options of Optix associated with novel traits in adaptive butterfly wing radiations. Evodevo. 5:7.

McMenamin SK, Bain EJ, McCann AE, Patterson LB, Eom DS, Waller ZP, Hamill JC, Kuhlman JA, Eisen JS, Parichy DM. 2014. Thyroid hormone-dependent adult pigment cell lineage and pattern in zebrafish. Science. 345:1358–1361.

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.

Monteiro A. 2015. Origin, development, and evolution of butterfly eyespots. Annu. Rev. Entomol. 60:253–271.

Müller GB, Wagner GP. 1991. Novelty in Evolution: Restructuring the Concept. Annu. Rev. Ecol. Syst. 22:229–256.

Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22:2362–2368.

Ordway AJ, Hancuch KN, Johnson W, Wiliams TM, Rebeiz M. 2014. The expansion of body coloration involves coordinated evolution in cis and trans within the pigmentation regulatory network of Drosophila prostipennis. Dev. Biol. 392:431–440.

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

Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh S-D, True JR, Carroll SB. 2006. Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. Nature. 440:1050–1053.

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.

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

Salzburger W. 2009. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18:169–185.

Santos ME, Baldo L, Gu L, Boileau N, Musilova Z, Salzburger W. The molecular basis of independently evolved anal fin pigment patterns in cichlid fish (submitted).

Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, Belting H-G, Affolter M, Salzburger W. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:5149.

Shirai LT, Saenko S V, Keller RA, Jerónimo MA, Brakefield PM, Descimon H, Wahlberg N, Beldade P. 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.

Takahashi T, Sota T. 2016. A robust phylogeny among major lineages of the East African cichlids. Mol. Phylogenet. Evol. 100: 234-42.

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.

Theis A, Salzburger W, Egger B. 2012. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. PLoS One. 7:e29878.

Wagner A. 2005. Robustness, evolvability, and neutrality. FEBS Lett. 579:1772–1778.

Wagner GP, Lynch VJ. 2008. The gene regulatory logic of transcription factor evolution. Trends Ecol. Evol. 23:377–385.

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

Wallbank RWR, Baxter SW, Pardo-Diaz C, Hanly JJ, Martin SH, Mallet J, Dasmahapatra KK, Salazar C, Joron M, Nadeau N, et al. 2016. Evolutionary Novelty in a Butterfly Wing Pattern through Enhancer Shuffling. PLoS Biol. 14:e1002353.

Wickler W. 1962. “Egg-dummies” as Natural Releasers in Mouth-breeding Cichlids. Nature [Internet] 194:1092–1093.

Wittkopp PJ, Haerum BK, Clark AG. 2004. Evolutionary changes in cis and trans gene regulation. Nature. 430:85–88.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Supplementary Table 1 Primers for long pcr amplification for Ion Torrent sequencing and pcr amplification for sanger sequencing

gene	primerF	primerR	Tm
pricke_up	CAAAAAGTGTCACCATGCCG	GCAGGAGAATGAGGGCGAG	65-55
prickle_pro	CGCCCTCATTCTCCTGCCTGTTTC	CTCCCTCTGTTCCGCCATCTCTGC	65-55
PRICKLE_2_of_2	GTCGGTAGAAATGTCTGCTGTTCCTTG	GTACCATCAGTTGATCTCCATCAGCGT	65-55
prickle_up	CAAAAAGTGTCACCATGCCG	GCAGGAGAATGAGGGCGAG	60-50
10298_up	GCAAAGCATGTGATAGACCTGTTCCCCT	GCCAAAGTGTCGGCGAGCAGAGC	65-55
10298_OV	CAGTTTGAGTTTGTATTCTCTTAGCAGC	TTCTGATGAAAAAATGGTGAAAAGC	65-55
10298_ov2	CAGTTTGAGTTTGTATTCTCTTAGCAGC	TCTGATGAAAAAATGGTGAAAAGC	65-55
10298_OV_special	CTTGTATTCTCTTAGCAGCAGACTTG	TAGTTTGTGGTCATTTTTACAGATTTC	65-55
10298_ov_primer	CAGTTTGAGTTTGTATTCTCTTAGCAGC	TTCTGATGAAAAAATGGTGAAAAGC	65-55
10298_pro	GAGCAGACAGTCCTTCATTATTTCCT	CTTTTAACCCCTACCACCGCC	65-55
10298	GATAGCCCTGCGGGTCTCATCTGAAT	GTGACGGTGGGTAAAGCGGGTAAGC	65-55
cecr_new	GACGAAGGCACAGACACAGATCAAAATAT	CTTTCAACCCTCCGATGCCCAC	65-55
cecr_intron	TGACGAAGGCACAGACACGGATC	TGACATCAGCACAGTGGCAGGGT	65-55
cecrnew1	TATGAGGGTGACTGAAGGGATG	TTGGCAGAGATGAAGGCGTC	60-50
cecr_newnew	GTTGGACGGGAGGACAGCGGGAAG	CAAACAGGGACGGGTCATCGGCAC	70-60
cecrnew1	TATGAGGGTGACTGAAGGGATG	TTGGCAGAGATGAAGGCGTC	60-50
cecr_pro	TTCGTGTCTTTCATGGACTATCTTCTC	GTCAGTCAGGATCTAATATCACTTGGC	60-50
cecr1_1of2	GGTTTGCTCTGCTGCTCCCATTCAT	GGATGCCACTTCTATTCACCTCAAGTATGT	65-55
cecr_intron	TGACGAAGGCACAGACACGGATC	TGACATCAGCACAGTGGCAGGGT	60-50
fhl2b_up_long	CTCCACTCACAACCTTTTTACTATG	GTGTCTACTGAGCGCTGTGGT	65-55
fhl2b_pp_primer2	TACTTCAAACCTTCAAAAGACCCTCAAT	GGCAGCTTGGTAGCAGTCGAGG	65-55
fhl2b_OV_primer2	TTGTTGTTCGCTTTTTATTTCTGT	GAGCTGCCTTGTATCATTTCTTG	65-55
fhl2b_pro	CTTCCGCCTCCACGCC	GATTCAAATCATGCAAAGTTAGGTG	60-50
fhl2b_A1_hb	GCTGTTGTTAAATGCTTTTGTTGTTCGCT	GTTCTTGTTCCACGTATGGATGTTTTCTGAG	65-55
fhl2b_A1_On	GCTTTTGTTAAATGTTTCTGTTGTTCGCT	GGTCATTCACATAAAACGCAGTATCTTCAG	60-50
trpm1b_intron	TCCTAAAAAGAAAAATTATCCATCGAGCAGC	GACCAGTACAAAAGCTGCAATGGGATAATG	65-55
trpm1b_2	GGTGGTGACTGTGGTCCCGT	TTGTCCTCCTGAGTTGAAAAATCC	65-55

trpm1b_87	TAAATGAAGGAGAAAGTTGGAT	GCAATGATACTCAGGTAGGGT	65-55
trpm1b_new	GCGCAATAACATGAGAAATGACAGAGC	GGTGTTGAGAGTGCACATACTTGAGGC	70-60
13024_new	TACAAGCATGAAGAAAAGTTAACGCAGT	CCAAGAAGACCAATAGAAGGCCAAAG	60-50
13024_c_primer2	CACAGGTAATCTTTGCACTCTGACACT	GAAACAAAAACAAAAAAAGAGAAGTGAAT	60-50
13024_pro	GCTGCTGATTCTGTTGTAATTCGGTTC	CTGAGTAAGCGGTGGTGCTGTGAC	65-55
13024	CATGCTGTCTAAATCTGACCTGCCAATAAC	CTAGCTTCCCTTCAGAAGACAAGAAACCTG	65-55
13024_haplolack	GCTTTTGGAGCCTCTTCATGTAGTC	GTGTTCGTGTCTTTATCAGCCCTACT	60-50
13024_LINE	CCCCTAAACAGTGTTGCAGT	GTGTGTGTTCTCCTTTTGATGTT	60-50
13024_part3	TTTAATTGTGTAAGTAAGATGCTGCAGAG	CCATTACTTTCCTTTTCTGGAGCCAT	60-50
13024_up	CAGTTTGGAGCTAATGATTTTAAATGTGTGC	TGAAGAAGCAGCAAATCTACTGTATGTGGTG	60-50
zdhh_up	CAAAGCAGCTAGGAACACAAG	TTAATAAGGAGTCGGAAAAAAG	65-55
zdhhec	GTTCTGGTTGTATACATGTTCTG	GGAAAAAAATTATCGGCC	60-50
zdhh_ex10	GTCTTATGGTGTTGAATGTCTCT	GGTGTGTTCGTTTTTATCGT	60-50
zdhh_pro1	TTGTCTTTTTTCCGACTCCTTATT	GTCTGTCTGACTCGCATTGATC	60-50
zdhh_pro2	CCACCCACTGTCCTGTGATTG	CAGATGTTTTACTGAGGTGTGTTCGT	60-50
zdhhc14_2_of_2	GGGGTCTGGTGGCTTGCTGTGAAT	CGTCAGCCCATCAGTCGAGTTTGTG	65-55
trpm1a_up	TATGAAAAGTAAAGCCTTAAATT	CTTCTCTGTCCTCTCACGC	60-50
trpm1a1new	GTGTTTCTCGGACATTCGGTTTCAT	TTCAGGAGGTTCAGGTTTGCTTTTTC	70-60
trpm1a_ex1	GAAGCAGAATCATCCTCAT	TGAACACACCTCGTCTCC	60-50
trpm1a_pro1	CATTCGGTTTCATGTTTTTCTT	TGCCCTTTTATATATTTCTCATTT	60-50
trpm1a1_pro_new	TTGTTACTGTGAAGAAGGAGAAGC	TTTATGATGGACAGCTAGAATTGG	60-50
TRPM1a_2	GCTTCAGACATCCTTTCGTTTGCTCAC	TAATGCGTCCAATAAAGACTCATCTCCTGT	60-50
TRPM1a_1	TAATACCAGCAAAATCACTTCTAATGACGCC	CCAGGTAGCCACATTAACATTAAACTGAGGAAC	65-55
trpm1a_pro2	GAGTGACAAGAAGCACGGGAAGAGG	GCCTCTCTGGACTCTGTTGGACCC	65-55
tmem_up	GCGTGAAAGCAAAAAGAAAAGC	TTAGGCAACACCACAGGCAACT	65-55
tmem	CCTCCGTTACGCATGAGCTGAAGC	CCGCATCGTCGTCCTTTCCTTTCT	70-60
TMEM235_1_of_2	GGAAGTTGGACGCTAGGATAAGGGATG	GAACACCATAGGCCCACTGTAGAGGAT	65-55
TMEM_Hb	TTTGGCTCCTAGTCCCTAAACCTAAATGTAC	TAAATCCGCCTCCTTACTTGGACCTC	65-55
st8_up	GAATGTTTTCATGTTTACCAAGAT	CACCACCTCCAATCCCAC	65-55

st8new	GTGGAGCTGTCTGGGATGTGTTG	TTCCCTGAGATTAGGCAGTGATTTC	70-60
st8	GCTACATACAAGGAATGAAAACGCTC	TATGTCGGTAATGGTCTCTGGAGTCAC	65-55
ST8SIA_Hb	GTGTATCGTATGTATGCCAAGTGACGGGTT	TTGAACTGAGCTGATGTATGTCGGTAATGGT	65-55
st8sia_1_of_2	GCGAGGACTCCTGTGAGGACATGAG	CTCCTTCCCTGAGATTAGGCAGTGATTTC	65-55
rgs_up	CTGGGACAAAGAGAAGGAAAGAAG	TTAAAGGTAAGGAGTCACGAATGGT	65-55
rgs_upnew1	GAAACTTGAATTTTTTGATTAACAG	CACACATAAAGACAGCAGTAGAGT	60-50
rgs_upnew2	CAGCAGGAGATAGCATGGCGAAAT	TTGATAGGTGCGATTGTGCCGTG	70-60
rgs	CCGTGTGGTTGCCTCTCTGTTGTC	CCTATCATTCCCAGTGTGAGCCCTAAC	70-60
RGS4_Hb	TAACACTATTTTGAAAGCACATAAGCACCAT	CCAATCTGTGACCCTAAACCTAAAACTAACT	60-50
RGS4	CTAGGGAAATGCAAGCTGCGGAAAG	GACTCTACTCCTACAAACAAGGGAAATCGG	65-55
pmel_2_of_2	CCTCATTAAGGTTGCATTCCTCCCAT	CCTGGTGCTTGTGACACGCATAAC	60-50
pmelb_up	GCATTTATTTATGTTCAGTTTGTT	TGCATTGATTTTACTTTTTTCTT	65-55
pmelb	CCGTTTATAAAGCCAGTTTTCTGTTGCCT	GCATAACCACTTGTATTTAGCTGTGTGAGGC	70-60
pmelb_pro	CAGCTGGAGGTGTGACTGTAATC	CGCATGTCTAAGGTATAATTCTGG	60-50
plcb_intron	TTTCCACTTTGTATCAATGTAAC	CGTCAACATGGACTAAAATCT	60-50
plcb_up	CTTGTGACTGAAATCAAATAATG	CACTAAGTACCACTCTGACGTTC	60-50
plcb_pro1	GACCCATACACAACGAGTGGCTTTCAT	GCAAAGGCTCCAGTGGTATAAGGGCTAG	65-55
plcb_pro2	CCAAAGGTCCGAAATGTGTT	CTGACTGGATGGCATCAAGAG	60-50
plcb_pro3	GCCTGGAGAAATCATACTCAAGACC	TAAAAGTGAAAAAACAACAACAAAACATT	60-50
plcb_up	CTTGTGACTGAAATCAAATAATG	CACTAAGTACCACTCTGACGTTC	60-50
plcb_intron	TTTCCACTTTGTATCAATGTAAC	CGTCAACATGGACTAAAATCT	60-50
PLCB2	CAGCTACTGAAGGTGTACTTTAATTTATGGTTTGT	CTCTCATTAAGCAATCTTCTACTATTTCATATCCG	60-50
PLCB2_Hb_1	CCACTTCTTGTTAGTCTTAGCAGCCATCTTAGC	CACTTGGTGCGGTATTTCTTCTTAGGGTCT	65-55
PLCB2_Hb_2	GAAGTTTGACCCTTTCTGTGACAGGATTGAC	GCCCTGATTATAGGTTATGTAACAGTTTGAAAGC	60-50
ttyh_intron	CCAGAGTAGCATTATTGGATTGAGTTC	TTGATAGTGAGAACAGAGTGGGTGG	60-50
ttyh_pro1	CAACATCTCTGAAGGAAGCCCGT	CAACCCTAAGCCTCCCAAACAAC	60-50
ttyh_pro2	GCTGTTGGAGTGGGTTTCTATGGG	GCAGGCACTTGTTTGGGGTTTATT	60-50
ttyh_pro3	GATTCTGTGATTTATGTGATTCTCCCTAT	GAGTTATACCGTGACGACAGCCTT	60-50
ttyh_2_of_2	CAGTGATGAACGACTTTGCCTACCCTCT	CAACTACCTCATATTGCCTGCTTTAATGTCTC	65-55

TTYH_Hb_2	TTGTCCTCGTGGTCTTCTATTCATTCTTTAT	CAACTACCTCATATTGCCTGCTTTAATGTCTC	60-50
TTYH_Hb_TE	CCTACAGAGTAAACGACCAGTAATGAATGC	CAGCCCTGATTTTCTAGTGAGATTATTTTG	65-55
TTYH_Hb_1	CTCAGGGCTGGAGGTCAACGGGATAC	GGAACAGGCTTGCTGGAAACCTTAGTGC	65-55
pnp2_up1	TAGTTCAAAAAAAATACTTCTCAAATG	GTACAGTGTCTGATCCCCGTG	60-50
pnp2_up2	TCAACTTTCTGAGCGTACAGCATC	CAATGGGAATGTGTCTTGAGGTAG	60-50
pnp2_up3	CCTCTTTTCCTCCTTGTAAACCTT	GTGAACATTTCTCTTTAGATACAGCAT	60-50
pnp2_up4	TGACATGCTTTTAATAACTTTTCTT	TTGTGGTCCACTTGTTCTCG	60-50
pnp2_pam	GGTGTGTCCGCTCAGCCAATG	GATTCTGTGTCTGTGTGTCCAAGTGAG	70-60
pnp2_ex6	TACTGGCTGGTTTATGGCT	CAACTGAGTGGGATTTTCTTC	60-50
pnp2_pro2	GATCTGCCTGACCTCTGCCCT	GGACTTTACATTCATTCGTTACATTTACTC	60-50
pnp2new	CTCGGTGCATTTTCCTCTAGGCTTTCTTCTG	GAGTGCGTTTTCAAAGATTCTGTGTCTGTGTGTC	65-55
PNP_2_of_2	CATGCAAAGTAGACCAAGATCAGACAGCG	GGTAACTATAATCCTTTCACCTTTTGACCTCCG	65-55
pnp_up	TCAAGATGCTAAAGTAGTGCG	TTTGGTTAAGGTCTGGACAG	60-50
pnp	CACTCGCACCGTCTCTTCGTC	CAATGTCCGTTGTTGGTATTCTGC	70-60
pnpnew	CAAGTAGGTAAGGCAAAAACAAAGTAATGGGCAG	TGAAAAACATACAAAGATGCTTGAGTCACAGGC	65-55
PNP	CCTCTTTCCTCGTTGCGCTTCTTCAT	CACAGCGGTTATTGGTATGTAAAGTTTTAGGAT	65-55
pnpnew	TGATTCTTTCATGTTTGGGTT	CTGTCATGGAATTGAGGGAGT	60-50
myoc_pam	TTGGCATCCTAAACACTTCCTCTCAG	TCAAAATGCAATCCAGAAAACAAAAAC	60-50
myoc_up	GCACGCTTCAGAAACCACC	GGAGACTAATCCAAACCCCATAT	60-50
myoc_ex2	CAGGGACCTTACAACTTTGAG	CTGCATCCACACACCATACT	60-50
myoc_pro	CACGAGCAAAAAGAAGAAACACG	GCCTTGACAAAGCACCTGAAATG	60-50
myoc_exon3	GATCTCTGCATCCCACGTCTTCC	CAGCCTTGACAAAGCACCTGAAATG	65-55
MYOC_2_of_2	GAGGCAGATGGTCAAGAGCTGGTGG	CGGCTGAGGACATGACTGCTTGGTT	65-55
13027_pam	TGATGAATAAAGCGTGCCGTCTGAT	TGCTGACGTTTACTGACAGGACATTTC	70-60
13027_up	GCCACTGAGGGACGGAGGG	TTAAAACAGGAAATGACCAAAAGAAAACT	65-55
13027	CATTTTCTCCTCATTATCAAGATGGCTGGT	CCTCATGTTTGGATGAACATTTGCTGTAAC	65-55
slc_up1	GATATGGGTTGGCAAGAGAC	TTAAAGTAATTATATGTTCAAGGCT	60-50
slc_up2	TTGAATGGCTGATTTGGGC	CCTCCTGTGTGTATTTAGTCTGTGTG	60-50
slc_up3	TCGTGGTGGTCAGAGGGC	GGTATGTGGTTGAAAAGTTGGC	60-50

slc_up4	CATCACAAAAGGGGAAACAC	CAAACTCTTCAAACTAGACAGGC	60-50
slc_up5	TCACAAATTCATGGTAGCTGCCG	GCAGTTAGTTTTCTTACATCTTCTGGACAC	70-60
slc7_pro1	GTTATAAGGGCAGCCCGTAAAGCG	CCTCACGAAGACTCAGCACAGCAAG	65-55
slc7_pro2	CGATAGGAAGGGATAAGAAAAGGAAAAG	TAAGACGAAAACGGGAAGTATTGACG	60-50
slc7a10_2_of_2	CTGAAGAACTGACTCCGATGAATAATGAAACAT	CAACAGATCAACAGTGAGCATCATGTGAGTAATT	65-55
slc25_pro1	GACATTCCCCCGTCTCGTCACCTAC	TACCTGCTATTCCAGCCGACACCAT	65-55
slc25_pro2	TATAACTTGTTTTGCCTCATTCCAG	GGTACTTGTTGCCTTCTTCGC	60-50
slc25a36_2_of_2	GCGGGTGATTGGGAATATCATGTCTGTG	CCAACCTCCTCCACCTGCTGGCTAAG	65-55
16584_p1	CATATTTTTCTTTTTACTTTTTGGGACCAGC	CCAACCGCAAATACAGACACAGTCAGAG	70-60
16584_p2	TATCTTTGGGTCTTGTTCACGAGTG	TTAAGAAGGTTGGCATGTCATTGTC	70-60
16584_p3	CGTGCAGGCTCTTTTAGTTATATT	CAGGTGAGGTGTTGTTTTGTTAC	60-50
16584p3new	CAGTATACCACAGCAGGAACAGGAATT	CATAGCACCGTACTCACCAAGATAATC	70-60
16584_ex10	TAGAGAAGCTGGGCCG	TTGATAACCACACCAAAATAAT	60-50
16584_ex4	GTGTATCAGGGGATGGGGTT	GAGTTGGGAATGAGGATGTGTT	60-50
16584_ex3	TCACGAGTGAGGGGAGC	TGAAGACGAATGTTTTACCTG	60-50
16584p3new	CAGTATACCACAGCAGGAACAGGAATT	CATAGCACCGTACTCACCAAGATAATC	60-50
16584_p2	GTGAGATAGCACCAAAGAACTTGCGAGG	GGAATGAATTGGCATTGAAGTTTGAGTTAAGG	65-55
16584_p1	CTGCTTTCTTCCATTTGTGCAACATCTCT	CTGGATGAAAGTGGGTACAGCAATCACTG	65-55
2365_p1	CTCCACCTTCCTCTCCTCCTTCCTG	GCTTCGTTGTTGCAGCATCCACT	70-60
2365_p2	CCTGTTGTCATGCATTGTTGGCTGTATT	TATGTAGGTCCACTTCCTGCCCTTTTG	70-60
2365p1new	CCTCTTTCCTTATCCTCCACCTTCC	TGCAATACCAGCTACTTTTTCTGTCTCC	70-60
2365	GTCTCCATCTCCCACTTCAAGCAGCTC	CAGCGTCTACTCTGTTATTATCTTATTCCACTTTC	65-55
7947	TCGTTCTTACCTCTGATCTCTCTCCTGTT	TAAAAAGCATTAAATCTTATCTCGAGACGTG	70-60
7947new2	GCTTGTTGAAGGTTGCTTTGGATC	CTGATTACGTTTGAGGCGGGTC	70-60
7947new1	CGTTCTTACCTCTGATCTCTCTCCTGT	GCAAACTGCAGCTGTCCCTCT	70-60
7947	CTGGGACTTCTCCTCTACAGCCTGGTTG	GACTGCAAACTCTTTCAGTTTAACTCAATTCACT	60-50
15173_p1	TAAGGAACCAGGCAAATTGGAGTCAC	CTTCACGAAGAAATGCAGGACAGCC	70-60
15173_p2	CATGAGGGAAAAATGAGAAAACC	TTTAGAAACAAAACAATGCAGGG	65-55
15173_ex1	CGCAGCCATAAGTCAAGT	CATCAGAGTAGCATATAAAGAAAT	60-50

15173_ex67	CATACGATGCATTGCTTAAC	GGCCGAAATTCCTCC	60-50
15173_ex8	TACTACAGGAAAAACAATGCT	TAGAAACAAAACAATGCAGG	60-50
15173	GACTAAATTGCTATTTCTTTGACAGGTGTTGTTC	TAAAGTGGAGACTGACACTGGTTTTAATTCTTAAT	60-50
23711_p1	CACGGCTCAAACTCTACAAATG	TATCGTGGACTCTCTGGAAACC	60-50
23711_p2	CAGCATTCACATTTGTCTTGGAGT	GCAGCTGATCTGACAGTTCCTTG	60-50
23711	GTTAGTGGTCAGAGGTTCGGGTGATGTGTG	GCAGTGACAATTTGCACGAAGCAGCT	65-55
tfec	CAGTTGTTCTTCTTCTGTAAGGATTTGC	GCTCGGAGCCATCACTCAGTCAT	70-60
TFEC	GATAGCCAGACGGAGCAAACAAACAGC	CGACAAGAACTACACGCTTAATGCACAGG	65-55
rab27a	CTTGGGATTTGGTGTGTGGGTCTGTG	CTGACTGGTTGTTGTCCTGTATCGCCTC	70-60
rab27anew	CTCTACACAGTATCACGAGGAAGAAGTT	GGATAAAGTAAACAGAGGAAGGAATG	60-60
rab27a_pro	CTTTTTTCTGGCTTGGTTTTTATCAC	GTAAACAGAGGAAGGAATGTGGTGAG	60-50
impdh	GCTGATTGAAGTGGCTTTGTTTTGGTC	CATTTGGTCACTGTAAGGTTGGGAAGGT	70-60
impdh	GCTGATTGAAGTGGCTTTGTTTTGGTC	CATTTGGTCACTGTAAGGTTGGGAAGGT	65-55
IMPDH1_1_of_2	CTAATGCTGATGTTTAGAAAGCCTAATCTGTACTT	GTAAATACACTTGGCAGTATGTATGCAATTAACG	60-50
IMPDH_Hb_TE	GATGGACGAGTTTAGACCAGTTTATCAGGAGG	TAAAACAGCGATGTCCCGTGAAAGCAA	65-55
IMPDH_Hb	TGTGCGGTTTGTTTGGCATTAGATTAGTT	TAACATTTGGTCACTGTAAGGTTGGGAAGG	60-50
gpnmb	GGACGAGGAGAAAGAGGGACACGCT	CTGGAGCACCATCAAACAATTAAAACAAGC	70-60
gpnmbnew	TTCCCACACAAGCACTCATTATCAG	TACAGGGGGTTTCTCTCAGTCAGC	70-60
gpnmb_ex4	CAGCCAATGGGCAGGTTC	GACGGTGTTGTCGGTAGAGAGG	60-50
GPNMB	CTGGGTGCTACAGTATAAGTGGTGGTCGG	AACAGCCATTGCTTGAGCCTAATTCATT	60-50
prtfdc_newp1	TTGTTCAGAAGGAAGTAAGCT	GAAAGAAAAAGATTAGAGAACG	60-50
prtfdc_newp2	TTGCTTTGCTTGAACAGTATT	GCAATGGGAGAGAAATATGTG	60-50
prtfdc_7_f	TCCCGAGATCACAGAAACAT	GGCACCTTGAACACACACTT	65-55
prtfdc_newp1	TTGTTCAGAAGGAAGTAAGCT	GAAAGAAAAAGATTAGAGAACG	60-50
prtfdc_newp2	TTGCTTTGCTTGAACAGTATT	GCAATGGGAGAGAAATATGTG	60-50
prtfdc_pro	TGTGTTACCGCTTTTGTCTTGTTGAG	CAATGGGAGAGAAATATGTGCTGCT	60-50
PRTFDC1_2_of_2	CGCCTGAAGGAGACCTGGGATAAGC	GGTACGAGGATGCCAACGTCCATAAGC	65-55
PRTFDC1_2_of_2	GCCTGAAGGAGACCTGGGATAAGC	CGAGGATGCCAACGTCCATAAGC	65-55
23698_exon16	CATTCCTTCTGTGCTCACCATC	TGGCTTCAGTTACAAAACCTTCAT	60-50

23698_exon79	GCTGTAACAGAGAGGAAGTGATT	GAGATGAAAACAAAGAGAACTACC	60-50
23698_16_2	GCAACCATGCAGGACCTCAACG	GTCAGCCTCCACGGTCAAATGC	65-55
23698_pro	TGTTCCACTTGTCTTTCATC	TTAGCATTTTCTGTGTTTTG	60-50
23698	CCACCCAGCCCTGGTAGTCCCG	GAATCAAACACCATCTTTACCCATACTTAGAGCAC	70-60
13044_exon13	GCTTCCACTGCCTCCAC	GCCTATATACTTGCCATCCTT	60-50
13044_pro	CATAAATAGTCAAATGTTACAGCGGT	GTTACACGGGGAAAACTTACACAC	60-50
13044	GGCTTAGAGGCGAACAACTGACATAACAG	GTGGCTTTAAGATAATCCCATCAGAACTGC	65-55
26203	TACACAGTTTTTTTCCCCATCT	TGTGTCTCCTAATGTCAATGTCTC	60-50
26203_exon	CCTCTCCCGTCTTGTCTTTCTC	TGTGTCTCCTAATGTCAATGTCTC	60-50
26203	GAACCAGGAACAAACGCATTCAGGAGC	GCACCAGCAGAGCCAAGAGTGGAGTT	65-55
crip_pro1	GGACAACTAAAACACATCACTTTCACTCG	CGGCGTCTTTATATTTTGGGCTCG	60-50
crip_pro2	GGAAACCTCAGCAGGGTGTAATCAG	CGGACACACAAGTGGGGAAGAG	65-55
CRIP2_Hb_2	CTCGGTGTTCCCGCAGTAGCATCATC	CATCAACAATCGGCTCTGACATCCACAT	65-55
CRIP2_Hb_1	CCTCATCCCAATCCTGACATCAGCAAG	GAGTCACAGAGGCAGCAGAGGAAACACC	65-55
CRIP2	CATCTACTGAGACTCTTCAGTCTTACCCCTGTC	GAGGCTACAAATAATACTGTGGTTGATGGTC	65-55
gatsl_pro1	TTCTTTTTTACCTTTTAGTCCTTCCTC	CGAGCCGTGTGTCGAGTCAT	60-50
gatsl_pro2	CACATCCACTTCACAACCCATAGG	GGTAGTCGTCATCGCTTTCTCCAG	60-50
GATSL	GTCTCGGCCTGCCGTCTGAGTGTAT	CAGGAGGGTCAGGGAGTGGGAAGAAG	65-55
phldaexon	CACGAAGGGCAGGAGCAC	TTCAAACATCCAGAAGATAAACGC	65-55
PHLDA1	CTGAATCAGGGGCAGCTCTGTTTG	GGAGAATGGAGGCTCGGTAGGGAAG	65-55
mc1rexon	GCTCCAGAGACAGCTGATGAAGGC	GACAGAGGAAGAATGTTGTGTTGCG	65-55
MC1R	GGGAGGCAGCAGTAATGCTTTGGCT	GCAGGCCCTGTTACCAGCACTTATGT	65-55
tpd52_cds	TTGTGTGTGTTTCACACAGGTTTTCTAGAT	CCATCAAGTTGTTGGTGGGCGT	65-55
tpd52l1	CTAACCTCACTGTCCCCTCACCTGTCTT	CTTGCTCTGATCTTGCTGGGCTGTCT	65-55
tpd52_pro	TGTCATTAGGGTGTCGTCAGTT	TTTGAGTTAAAGTGCTGCGAC	60-50
tubb_2_of_2	GAATCCATTACCAGACACCACCATCG	GTCCCCTCTTATTTTCAATCTCCTCTTCAT	65-55
RAB27B_1_of_2	TGGAACAATGCAAGACATAAACCAGGAG	GGGAGGACAAGTAAGCATAAAGGGAGG	65-55
RAB27B_2_of_2	TGTGGACCCTAAGGTGTTTTCTGAGCT	GACATACAGTTGTTTCATTTTGCAGTTTTCTT	65-55
rab27ba_hb	GTGTTGCCATCTCCTTGTTCTGTTCCTC	CAGAGCCTATTCCCAGGGTGCATGTT	65-55

tagln3	CTCCCGCTTCCGTTCCGTTCTG	CGTGAGGTTGATTTGGTTACAGAGGATACT	65-55
10754	CTGAAGTTTGCAGTTGGCTCTGAAGTAAG	CAAGCAGTAGGGTTGAACAGTATGAAGC	65-55
fhl2_1_of_2	GAGTGCCTGGTTTGACCTTTCCACATGG	TTATCAACTCTGCATCATTCCCTGATCATAAAGC	65-55
fhl2_pp_primer2	TACTTCAAACCTTCAAAAGACCCTCAAT	GGCAGCTTGGTAGCAGTCGAGG	65-55
fhl2_ov_primer2	TTGTTGTTCGCTTTTTATTTCTGT	GAGCTGCCTTGTATCATTTCTTG	60-50
fhl2_up_long	CTCCACTCACAACCTTTTTACTATG	GTGTCTACTGAGCGCTGTGGT	60-50
ACSL6_P2	CAGTTCTTCCGCTCCCTGCCGACCT	CTACCGTTGGGCCAGTTTTGCTTTGCT	65-55
ACSL6_P1	TTAGTATAAACAGCGATCTGCCTGAGAAGTC	CATGGTAATGAGTCAGCAGCTTGGAGG	65-55
SYNGR_1_of_2	TCCCAAAGCGCACTCATACCTTCCAC	CCATGCTGCACTGCTACACCCGATC	65-55
apod	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
2365_ex1	CTTTCCTTATCCTCCACC	CATGGAAGCATCTACAATTC	56.6
2365_ex3	TGATCCCCAAGCATAGCAGC	GACTTCAATGCACACCTCTCTG	64.2
7947_ex1ex2	CCAGCCCTCTATTCTGTGC	CCATAGCTAATTTTAAGATGTGTTG	60
7947_ex3	CCTTTATCGTGAAATTCTTAAG	GCAAGGACATGGTACTCC	56.8
gpnmb_ex1	CTGACTGTAGAGAGCGGATGG	TGAGTTTCCTGAGTTTGGGGG	67.3
impdh_ex1	CCTGTGAGAAGGTCACCC	CACTCCACCCACAACTTCG	61.3
impdh_ex15_p1	GAGAGAAATGGAGGGTAAAGC	CGAGAAACTATGCCACGACG	61.1
impdh_ex15_p2	CCAGTGTCAACAGTAGATG	GCTATTTGTCACATGCTTC	54.3
myoc	TTCTCAAGGACCAGCTTGCC	GTTGTTGCATCAGTCCTGCC	66.1
myoc_ex1	CACGAGCAAAAAGAAGAAACACG	CAGCCTGTCTCTGCCTCTTTAGAC	69.7
myoc_ex2	GTGTGATTATTTTCAGGGTTAGAGC	TGCTCGATTCCAACAACTCC	63
myoc_ex3	CTCTGCATCCCACGTCTTCC	TAGCACAGCCTTGACAGAGC	63.8
pnp4a _ex1	GATGACGAGAACAAGTGGACCA	AGGGCAGAGGTCAGGCAGATC	69.7
pnp4a _ex2	GTGATCTGCCTGACCTCTGC	TGTCCAAACACCAGCTTCCC	65.6
pnp4a _ex3ex4	AGACATCCCCAACTTTCCGC	GCACACAGTACACTCCCTCC	62.8
pnp4a _ex4	TTTGCATGCAGGGACGTTTC	TACGCCAAACCTGACAGACG	66.5
pnp4a _ex5	TTGAGACGGTGATGCTGACC	CAAGTTGAGGTTTGGGCAGC	66.8
pnp4a _ex6	GCTTGAAGTCCTTTTGCTGC	GATTCTGTGTCTGTGTGTCCAAGTGAG	64.3

Supplementary Table 2 Point mutation for 40 blotch related candidate genes in different lineages

Ensembl	Gene	point mutation			
		egg spot	ectodini blotch	haplochromine blotch	haplochromine
ENSONIT00000002365	VEGFC	no	no	389	157,235
ENSONIT00000007947	retsat	606may, needllpp	no	176,471	no
ENSONIT00000010754	rab38	no	no	no	no
ENSONIT00000013024	slc23a1	no	no	2,26,499,555,583,	491
ENSONIT00000013044	PCBP3	no	no	69,88,246,372,399	no
ENSONIT00000015173	CRD2	no	76,291,	5,76,94,149,	38,587may
ENSONIT00000023698	keratin	14	no	14,57,65,75,148,193,194,196,231,242,340,	222
ENSONIT00000023711	slc2a5	511	no	110,	no
ENSONIT00000000025	cecr1b	no	no	29	no
ENSONIT00000001209	crip1	85	no	104	no
ENSONIT00000016736	gatsl2	no	147,	327,	no
ENSONIT00000005686	gpnmb	no	no	60,151,195,263,575,	no
ENSONIT00000014451	impdh1b	no	no	no	no
ENSONIT00000026719	mc1r	no	no	173	no
ENSONIT00000026350	phlda1	no	no	no	no
ENSONIT00000020824	pmel1b	no	5,71,161,271,304,443,	no	no
ENSONIT00000016474	pnp5a	155may	no	unknown	unknown
ENSONIT00000000731	pnp4a	no	106，144，	15，16，28，	291，

ENSONIT00000021803	prickle2a	no	no	356,710,711,734,	no
ENSONIT00000008139	PRTFDC1	177may	no	no	no
ENSONIT00000016030	rab27a	no	no	no	no
ENSONIT00000021335	rab27b	no	no	215	214
ENSONIT00000013824	rgs4	no	no	32,33,111,151,	no
ENSONIT00000003926	slc7a10a	no	no	no	no
ENSONIT00000018418	slc25a36b	no	no	no	73
ENSONIT00000021336	ST8SIA3	no	no	no	no
ENSONIT00000017829	tagln3b	no	no	no	no
ENSONIT00000025015	tmem235	no	no	no	no
ENSONIT00000019285	trpm1b	no	336,	no	470
ENSONIT00000004083	trpm1a	23	1254,	no	23,163,
ENSONIT00000006814	ttyh2l	no	no	86,140,419,	6
ENSONIT00000016436	tubb5	no	no	no	no
ENSONIT00000023686	zdhhc14	484	no	339,350,	47
ENSONIT00000023494	MYOC	no	115,245,408,	143,	36,404,
ENSONIT00000015512	fhl2a	no	no	no	no
ENSONIT00000017889	fhl2b	no	4,	no	no
ENSONIT00000023475	APOD	no	no	79	no
ENSONIT00000010298	vtna	no	59,106,134,144,383,	92,189,382.	no
ENSONIT00000026203	novel	38	no	no	no

ENSONIT00000000849	tpd52	no	no	no	no

Supplementary Table 3 Likelihood ratio test statistics of branch-site model comparisons

gene	foreground branch	model A	model B	p/2 value
2365	**c**	**-2176.96**	**-2178.47**	**0.04**
	hap	-2179.7	-2179.69	0.5
	pp	-2178.25	-2178.88	0.13
	spot	-2179.7	-2179.69	0.5
7947	c	-3093.23	-3093.22	0.5
	hap	-3092.31	-3092.3	0.5
	pp	-3093.03	-3093.02	0.5
	spot	-3091.68	-3091.67	0.5
10754	c	-1230.87	-1230.87	0.5
	hap	-1230.87	-1230.87	0.5
	pp	-1230.87	-1230.87	0.5
	spot	-1230.87	-1230.87	0.5
13024	hap	-3288.71	-3288.71	0.5
	pp	-3287.37	-3287.44	0.35
	spot	-3288.71	-3288.71	0.5
13044	c	-2035.13	-2035.12	0.5
	hap	**-2033.1**	**-2034.47**	**0.049**
	pp	-2034.47	-2034.69	0.25
	spot	-2033.78	-2034.28	0.5
15173	C	-3243.48	-3243.48	0.5
	hap	**-3239.97**	**-3243.48**	**0.004**
	pp	-3243.16	-3243.3	0.29
	spot	**-3240.41**	**-3243.48**	**0.005**
23698	hap	-3001.42	-3001.42	0.5

	pp	-2998.8	-2999.17	0.2
	spot	-3001.42	-3001.42	0.5
23711	c	-2816.78	-2816.78	0.5
	hap	-2816.78	-2816.78	0.5
	pp	-2816.78	-2816.78	0.5
	spot	-2816.78	-2816.78	0.5
cecr	c	-2600.17	-2600.17	0.5
	hap	-2600.17	-2600.17	0.5
	pp	-2600.17	-2600.17	0.5
	spot	-2600.17	-2600.17	0.5
crip	c	-1027.02	-1027.02	0.5
	hap	-1027.02	-1027.02	0.5
	pp	-1025.38	-1025.38	0.5
	spot	-1027.02	-1027.02	0.5
gatsl	c	-1635.41	-1635.54	0.3
	hap	-1635.7	-1635.7	0.5
	pp	-1634.9	-1634.9	0.5
	spot	-1635.7	-1635.7	0.5
gpnmb	c	-3595.94	-3595.94	0.5
	hap	-3595.94	-3595.94	0.41
	pp	-3595.8	-3595.8	0.5
	spot	-3595.94	-3595.94	0.5
impdh	**c**	**-2473.91**	**-2473.91**	**0.005**
	hap	-2473.91	-2473.91	0.5
	hap_c	-2473.91	-2473.91	0.5
	pp	-2473.91	-2473.91	0.5

	spot	-2473.91	-2473.91	0.5
mc1r	hap	-1594.01	-1594.01	0.5
	pp	-1594.07	-1594.07	0.5
	spot	-1593.78	-1593.78	0.5
phlda	spot	-590.11	-590.11	0.5
pmelb	C	-3365.63	-3365.63	0.5
	spot	-3364.27	-3365.2	0.08
pnp	c	-1443.32	-1443.32	0.5
	spot	-1443.32	-1443.64	0.21
pmelb	C	-3365.64	-3365.63	0.5
	spot	-3364.27	-3365.2	0.08
pnp	c	-1443.33	-1443.32	0.5
	spot	-1443.32	-1443.64	0.21
	hap	-4325.1	-4325.1	0.5
	pp	-4324.81	-4324.81	0.5
	spot	-4325.1	-4325.31	0.25
	hap	**-1170.68**	**-1179.29**	**<0.01**
	pp	-1179.29	-1179.29	0.5
	spot	**-1176.8**	**-1179.29**	**0.015**
rab27a	c	-1069.52	-1069.52	0.5
	hap	-1069.45	-1069.49	0.38
	pp	-1069.52	-1069.52	0.5
	spot	-1069.17	-1069.3	0.3
rab27bb	c	-1251.73	-1251.73	0.5
	hap	-1251.73	-1251.73	0.5
	pp	-1251.62	-1251.92	0.22

	spot	-1251.73	-1251.73	0.5
rgs	c	-1299.16	-1299.16	0.5
	hap	-1299.16	-1299.16	0.5
	pp	-1299.07	-1299.07	0.5
	spot	-1299.16	-1299.16	0.5
slc7a	c	-2471.35	-2471.35	0.5
	hap	-2471.35	-2471.35	0.5
	pp	-2471.35	-2471.35	0.5
	spot	-2471.35	-2471.35	0.5
slc25a	C	-1259.62	-1259.62	0.5
	hap	-1258.7	-1258.89	0.26
	pp	-1259.62	-1259.62	0.5
	spot	-1258.27	-1258.27	0.5
st8	hap	-1761.78	-1761.78	0.5
	pp	-1761.78	-1761.78	0.5
	spot	-1761.77	-1761.77	0.5
tagln	C	-965.42	-965.42	0.5
	hap	-965.42	-965.42	0.5
	pp	-965.42	-965.42	0.5
	spot	-965.42	-965.42	0.5
tmem	c	-1048.51	-1048.51	0.5
	hap	-1048.51	-1048.51	0.5
	pp	-1048.51	-1048.51	0.5
	spot	-1048.5	-1048.5	0.5
trpm1a	C	-6784.66	-6784.67	0.44
	ecto	-6784.61	-6784.61	0.5

	hap	-6783.76	-6784.62	0.09
	pp	-6784.63	-6784.72	0.33
	spot	-6787.33	-6784.73	0.5
trpm1b	C	-6836.3	-6836.3	0.5
	hap	**-6832.59**	**-6836.01**	**0.005**
	pp	-6836.3	-6836.3	0.5
	pp_c	-6836.3	-6836.32	0.42
	spot	-6835.86	-6835.98	0.3
ttyh	C	-2717.25	-2717.25	0.5
	hap	-2717.25	-2717.25	0.5
	pp	-2716.04	-2716.01	0.5
	spot	-2717.25	-2717.25	0.5
tubb	C	-2080.37	-2080.37	0.5
	hap	-2080.37	-2080.37	0.5
	pp	-2080.37	-2080.37	0.5
	spot	-2080.37	-2080.37	0.5
	hap	**-2732.38**	**-2736.36**	**0.002**
	pp	-2736.38	-2736.38	0.5
	spot	**-2731.48**	**-2736.38**	**0.001**
apod	hb	-979.26	-979.26	0.5
	pp	-979.28	-979.28	0.5
26203	spot	-1838.71	-1842.464	0.003
	hap	-1840.62	-1842.61	0.02
	pp	-1842.61	-1842.61	0.5
	C	-1842.61	-1842.61	0.5
10298	c	**-3432.03**	**-3434.82**	**0.0090965**

hap	-3436.02	-3436.02	0.5
pp	-3436.02	-3436.02	0.5

Supplementary Table 4 Output reads details of ion torrent sequencing

species	first round		second round		third round	
	raw	after	raw	after	raw	after
C	117816	107418	163569	150973	143288	138932
LE	104282	94696	143769	134441	123027	120001
LL	115607	106665	151523	142418	110868	108295
HC	114346	103187	142447	133648	157639	153428
OV	124957	113405	190089	176161	112440	109299
GW	150280	135700	144392	135099	104437	100574
PM	145680	129459	7598	7062	104229	101625
GL	143184	128688	142061	131441	113157	109900
Tn	131065	119994	143904	136165	112000	109165
EC	109280	97951	141706	132584	106873	103790
CL	104421	95775	146155	134935	140690	136273
PP	115199	104486	150146	141761	136342	133253
Cy	137608	124022	18441	17053	140830	137431
L	113560	103831	155085	142175	105889	103161
XS	133516	119129	157793	146987	36835	35996
BM	116517	106278	126137	119742	138956	134685

Note: C: *C. macrops*, LE: *L. elongates*; LL: *L. labiatus*; HC: *H. cf. stappersii*; OV: *O. ventralis*; GW: *Greenwoodochromis*; PM: *P. microlepis*; GL: *G. lemairi*; Tn: *T. nigrifrons*; EC: *E. cyanostictus*; CL: *C. leptosome*; PP: *P. philander*; Cy: *C. frontosa*; L: *L. careulus*; XS: *X. spiloptera*; BM: *B. microlepis*.

Supplementary Table 5 Eggspots and non-eggspots specific CNE TF binding sites prediction

26203_spot_GA mutation	26203_philander_mutation	26203_non_eggspots
Hmx2/Nkx5-2 homeodomain transcription factor	Hmx2/Nkx5-2 homeodomain transcription factor	Hepatocyte nuclear factor 3 (alpha, beta) (FOXA1, FOXA2)
Glucocorticoid modulatory element binding protein 2	Hypoxia response elements, binding sites for HIF1alpha/ARNT heterodimers	Tumor suppressor p53 (3' half site)
Tal-1 beta/HEB heterodimer		Hepatocyte nuclear factor 4 gamma (NR2A2), DR1 sites
Atonal homolog 1, HATH1, MATH-1		Tal-1beta/HEB heterodimer
GATA-binding protein 5		Atonal homolog 1, HATH1, MATH-1
Retina-derived POU-domain factor-1, dimeric binding site		Transcription factor E2a (E12/E47) (secondary DNA binding preference)
		GATA-binding factor 2
		Mix1 homeobox-like 1

fhl2b_specific	fhl2b_non_spot_specific
LTSM elements with 6 bp spacer	v-Myb, variant of AMV v-myb
Leucine rich repeat (in FLII) interacting protein 1	SRY-related HMG-box gene 4, dimeric binding sites
Insulator protein CTCF (CCCTC-binding factor)	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)
EGR1, early growth response 1	Homeobox B13
Lens-specific Maf/MafA-sites	Sine oculis homeobox homolog 4
Binding sites for homodimers of large Maf-proteins	Autoimmune regulator
LBP-1c (leader-binding protein-1c), LSF (late SV40 factor), CP2, SEF (SAA3 enhancer factor)	Basic leucine zipper transcription factor, ATF-like 3
Smad3 transcription factor involved in TGF-beta signaling	cAMP-responsive element binding protein 1
Zinc finger protein insulinoma-associated 1 (IA-1) functions as a transcriptional repressor	Initiator (INR) and downstream promoter element (DPE) with strictly maintained spacing
KRAB-containing zinc finger protein 300	GLI-Kruppel family member GLI3
Pleomorphic adenoma gene (PLAG) 1, a developmentally regulated C2H2 zinc finger	GA repeat binding protein, beta 1)

protein

Ras-responsive element binding protein 1

GATA binding factor

Pleiomorphic adenoma gene-like 1 (secondary DNA binding preference)

Sp4 transcription factor (secondary DNA binding preference)

Ikaros 1, potential regulator of lymphocyte differentiation

Mammalian transcriptional repressor RBP-Jkappa/CBF1

Regulatory factor X 5

Paired box protein Pax-9

Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs, IR3 sites

Estrogen-related receptor gamma, homodimer DR4 binding site

Retinoic acid receptor / retinoid X receptor heterodimer, DR5 sites

Tumor protein p63

Iroquois homeobox 5

Tumor suppressor p53 (5' half site)

Interferon regulatory factor 4

Egr-2/Krox-20 early growth response gene product

Progesterone receptor binding site, IR3 sites

Constitutive androstane receptor / retinoid X receptor heterodimer, DR4 sites

PRDI (positive regulatory domain I element) binding factor 1

Upstream stimulating factor 1/2

H6 homeodomain HMX3/Nkx5.1 transcription factor

POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma

Binding site for a Pbx1/Meis1 heterodimer

Homeobox D13 / Hox-4I

Caudal type homeobox transcription factor 2

Mammalian C-type LTR TATA box

Spalt-like transcription factor 1

Homeo domain factor Nkx-2.5/Csx, tinman homolog low affinity sites

AT-binding transcription factor 1

Hepatic nuclear factor 4alpha, DR1 sites

NF-kappaB (p50)

Downstream Immunoglobulin Control Element, interacting factor: BEN (also termed Mus-TRD1 and WBSCR11)

Peroxisome proliferator-activated receptor gamma

V-myb avian myeloblastosis viral oncogene homolog-like 1 (AMYB)

Winged helix protein, involved in hair keratinization and thymus epithelium differentiation

SRY (sex determining region Y)-box 6
TGFB-induced factor homeobox 2-like, X-linked, dimeric binding site
Transcription factor E2a (E12/E47) (secondary DNA binding preference)
Neurogenin and NeuroD binding sites
Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin
Meis1a and Hoxa9 form heterodimeric binding complexes on target DNA

Zinc finger protein of the cerebellum (Zic3)

Atonal homolog 1, HATH1, MATH-1
PTF1 binding sites are bipartite with an E-box and a TC-box (RBP-J/L) spaced one helical turn apart
GC box elements

Yin and Yang 1 activator sites
Cyclin D-interacting myb-like protein, DMTF1 - cyclin D binding myb-like transcription factor 1

SAM pointed domain containing ets transcription factor
Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12
T-box transcription factor TBX15, dimeric binding site

GATA-binding factor 1
Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome

X gene core promoter element 1

E2F transcription factor 1
Kruppel-like factor 7 (ubiquitous, UKLF)
Repressive binding sites for glucocorticoid receptor (IR2)
PAX6 paired domain and homeodomain are required for binding to this site
TGFbeta-inducible early gene (TIEG) / Early growth response gene alpha (EGRalpha)
Hypermethylated in cancer 1 (secondary DNA binding preference)

Myc-interacting Zn finger protein 1, zinc finger and BTB domain containing 17 (ZBTB17)
Hypoxia induced factor-1 (HIF-1)
Smad3 transcription factor involved in TGF-beta signaling factor PU.1

ERBA-related gene-2, homodimer DR1 binding site
Carbohydrate response element binding protein (CHREBP) and Max-like protein X (Mlx) bind as heterodimers to glucose-responsive promoters
ZF5 POZ domain zinc finger, zinc finger protein 161 (secondary DNA binding preference)
E2F, involved in cell cycle regulation, interacts with Rb p107 protein

Member of the RSRF (related to serum response factor) protein family from Xenopus laevis

Proximal sequence element (PSE) of RNA polymerase II-transcribed snRNA genes

Krueppel-like factor 2 (lung) (LKLF)

Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y

E74-like factor 1

STAT6: signal transducer and activator of transcription 6

Tumor suppressor p53

Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)

ETS family member FLI

PR domain zinc finger protein 14

Farnesoid X - activated receptor (RXR/FXR dimer), IR1 sites

Runt-related transcription factor 2 / CBFA1 (core-binding factor, runt domain, alpha subunit 1)

Activator protein 4

MYC-MAX binding sites

Zinc finger protein 410

LIM-homeodomain transcription factor

Brn-2, POU-III protein class

Transcriptional repressor CDP

Nuclear factor Y (Y-box binding factor)

Y box binding protein 1, has a preference for binding ssDNA

MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 2

E2F transcription factor 6

Glial cells missing homolog 1, chorion-specific transcription factor GCMa

Nuclear respiratory factor 1 (NRF1)

Zinc finger and SCAN domain containing 10 (ZNF206, Zfp206)

Histone H4 transcription factor, MIZF, dimeric binding site

Zinc finger transcription factor, Zic family member 2 (odd-paired homolog, Drosophila)

Glial cells missing homolog 1

E2F transcription factor 3 (secondary DNA binding preference)

Glial cells missing homolog 1 (secondary DNA binding preference)

Metal transcription factor 1, MRE

Forkhead box H1 (Foxh1)

Zinc finger and SCAN domain containing 4

Se-Cys tRNA gene transcription activating factor

Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)

Sex determining region Y

Brn-3, POU-IV protein class

Hox-1.3, vertebrate homeobox protein

Paired box 7 homeodomain-binding motif

Liver enriched Cut - Homeodomain transcription factor HNF6 (ONECUT1)

SRY-box containing gene 3, dimeric binding sites

Member of the vertebrate HOX - cluster of homeobox factors

Plant TATA box

TATA-binding protein, general transcription factor that interacts with other factors to form the preinitiation complex at promoters

Muscle TATA box

SRY-related HMG-box gene 7, dimeric binding sites

Regulatory factor X 3

Sox-5

Zinc finger protein 217

SRY (sex determining region Y)-box 9 homodimer

Cysteine-serine-rich nuclear protein 1 (AXUD1, AXIN1 up-regulated 1)

Homeobox transcription factor Nanog

Avian C-type LTR CCAAT box

Zinc finger transcription factor GLI1

Stimulating protein 1, ubiquitous zinc finger transcription factor

Zinc finger, BED-type containing 4; GC-box binding sites

Doublesex and mab-3 related transcription factor 4

Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA

Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF

Caudal type homeo box 1

Homeobox B8 / Hox-2delta

Homeobox C9 / Hox-3beta

POU class 4 homeobox 3 (POU4F3), BRN3C

Pancreatic and intestinal lim-homeodomain factor

NK6 homeobox 1

HMGA family of architectural transcription factors (HMGA1, HMGA2)

POU class 6 homeobox 1 (POU6F1)

Homeobox C8 / Hox-3alpha

Octamer-binding transcription factor-1, POU class 2 homeobox 1 (POU2F1)

Cart-1 (cartilage homeoprotein 1)

Zebrafish PAX2 paired domain protein

SRY (sex-determining region Y) box 9, dimeric binding sites

PAX6 paired domain binding site

Odd-skipped related 2
Promyelocytic leukemia zink finger (TF with nine Krueppel-like zink fingers)
V-myb avian myeloblastosis viral oncogene homolog-like 1 (AMYB) (secondary DNA binding preference)
Doublesex and mab-3 related transcription factor 3

Myogenic bHLH protein myogenin (myf4)

Achaete-scute complex homolog-like 2
Transcription factor AP-2, alpha
Drosophila initiator motifs

Pre-B-cell leukemia homeobox 3
TG-interacting factor belonging to TALE class of homeodomain factors
Calcium-reponse factor

MAX gene associated, dimeric binding site
Complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 2

Transcription factor yin yang 1
T-box transcription factor TBX21, dimeric binding site
Nascent polypeptide-associated complex subunit alpha 1

Erythroid krueppel like factor (EKLF)

Homeobox A1
Muscle segment homeo box 2, homologue of Drosophila (HOX 8)

Homeobox A3

Octamer-binding factor 1
Ecotropic viral integration site 1 encoded factor, amino-terminal zinc finger domain
Alternative splicing variant of FOXP1, activated in ESCs
Pit1, GHF-1 pituitary specific pou domain transcription factor
Jumonji, AT rich interactive domain 1B
Myocyte-specific enhancer factor 2
TCF/LEF-1, involved in the Wnt signal transduction pathway

SRY-box containing gene 3
Spi-C transcription factor (Spi-1/PU.1 related)
Lactotransferrin and delta-lactoferrin, growth-inhibiting protein 12

Tumor suppressor p53 (3' half site)
Zinc finger protein 282 (HTLV-I U5 repressive element-binding protein 1)

Amino acid response element, ATF4 binding site
Transcriptional repressor B lymphocyte-induced maturation protein-1 (Blimp-1, prdm1)
MyT1 zinc finger transcription factor involved in primary neurogenesis

Aryl hydrocarbon receptor nuclear translocator-like, homodimer	Hepatic nuclear factor 1
NK2 homeobox 6, Csx2	Zinc finger protein 652 (ZNF652)
Prospero homeobox protein 1, dimeric binding site	Homeodomain transcription factor Otx2 (homolog of Drosophila orthodenticle)
Thyroid hormone receptor, beta (ER5 - everted repeat, spacer 5)	T-cell specific HMG-box transcription factor 7
Estrogen-related receptor alpha	Hematopoietically expressed homeobox, proline-rich homeodomain protein
Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LRH-1), Nr5a2	
Estrogen response elements (ER alpha), IR3 sites	

13024_Hb_LINE

Snail family of zinc finger transcription factors binding to E-box motifs
Twist subfamily of class B bHLH transcription factors
Brachyury gene, mesoderm developmental factor
Metal induced transcription factor
Cyclin D binding myb-like transcription factor
Glucocorticoid modulatory element binding proteins
GC-Box factors SP1/GC
Activator-, mediator- and TBP-dependent core promoter element for RNA polymerase II transcription from TATA-less promoters
Estrogen-related receptors
PAX-2/5/8 binding sites
SWI/SNF related nucleophosphoproteins with a RING finger DNA binding motif
Growth factor independence transcriptional repressor
Glucocorticoid responsive and related elements
MAF and AP1 related factors

Myoblast determining factors
Human and murine ETS1 factors
Testis-specific bHLH-Zip transcription factors
Spalt-like transcription factor 2
Neuron-specific olfactory factor
GHF-1 pituitary specific pou domain transcription factor
NK6 homeobox transcription factors
Paralog hox genes 1-8 from the four hox clusters A, B, C, D
Special AT-rich sequence binding protein
C2H2 zinc finger protein PLZF
Brn-5 POU domain factors
Lim homeodomain factors
DM domain-containing transcription factors
Abdominal-B type homeodomain transcription factors
Bicoid-like homeodomain transcription factors
PAX-2 binding sites
HOX - MEIS1 heterodimers
v-ERB and RAR-related orphan receptor alpha
AP1, Activating protein 1
HOX - PBX complexes
Yeast TATA binding protein factor
Interferon regulatory factors
Neuron-restrictive silencer factor
"Negative" glucocoticoid response elements
Localized tandem sequence motif
Retinoblastoma-binding proteins with demethylase activity
Cellular and viral myb-like transcriptional regulators
Hepatic Nuclear Factor 1
X-box binding factors

PREB core-binding element
General transcription factor IID, GTF2D
Krueppel-like C2H2 zinc finger factors
hypermethylated in cancer
Hypoxia inducible factor, bHLH/PAS protein family
Prospero-related homeobox
Octamer binding protein
Chorion-specific transcription factors with a GCM
DNA binding domain
AT rich interactive domain factor
C2H2 zinc finger transcription factors 5
Autoimmune regulatory element binding factor
Human muscle-specific Mt binding site
Brn POU domain factors
Nuclear receptor subfamily 2 factors
LEF1/TCF
Ccaat/Enhancer Binding Protein
C2H2 zinc finger transcription factors 13
CP2-erythrocyte Factor related to drosophila Elf1
Positive regulatory domain I binding factor
Heat shock factors
PAX-4/PAX-6 paired domain binding sites
E2F-myc activator/cell cycle regulator
C2H2 zinc finger transcription factors 12
Vertebrate caudal related homeodomain protein
C2H2 zinc finger transcription factors 14
Fork head domain factors
Homeodomain transcription factors
EVI1-myleoid transforming protein
Homeobox transcription factors
MEF2, myocyte-specific enhancer binding factor
AT-binding transcription factor
Vertebrate TATA binding protein factor
MYT1 C2HC zinc finger protein

CCAAT binding factors
Activator/repressor binding to transcription initiation site
Nuclear factor of activated T-cells
NKX homeodomain factors
HTLV-I U5 repressive element-binding protein 1
POZ domain zinc finger expressed in B-Cells
Signal transducer and activator of transcription
Krueppel like transcription factors
TALE homeodomain class recognizing TG motifs
Zinc finger protein ZNF35
C2H2 zinc finger transcription factors 2
GLI zinc finger family
EGR/nerve growth factor induced protein C & related factors
Cart-1 (cartilage homeoprotein 1)
Cell cycle regulators: Cell cycle homology element
SOX/SRY-sex/testis determinig and related HMG box factors
Activator protein 2
Olfactory associated zinc finger protein
PAX-9 binding sites
Winged helix binding sites
Heterodimers between bZIP family members
cAMP-responsive element binding proteins
PAR/bZIP family
Ubiquitous GLI - Krueppel like zinc finger involved in cell cycle regulation
CLOX and CLOX homology (CDP) factors
PAX-3 binding sites
Pleomorphic adenoma gene
C2H2 zinc finger transcription factors 1
RXR heterodimer binding sites

Peroxisome proliferator-activated receptor
Two-handed zinc finger homeodomain transcription factors
Homolog to deformed epidermal autoregulatory factor-1 from D. melanogaster
Vertebrate SMAD family of transcription factors
C2H2 zinc finger transcription factors 4
E-box binding factors
CTCF and BORIS gene family, transcriptional regulators with 11 highly conserved zinc finger domains
Cell cycle regulators: Cell cycle dependent element
Serum response element binding factor
Double homeobox factors
snRNA-activating protein complex
GATA binding factors
pnp4a_noneggspots_specific
Podocyte-expressed 1 (POD1)
Pax-3 paired domain protein
Zinc finger protein 652 (ZNF652)
Pax-6 paired domain binding site
Cut-like homeobox 1, dimeric binding site
NMP4 (nuclear matrix protein 4) / CIZ (Cas-interacting zinc finger protein)
H6 homeodomain HMX3/Nkx5.1 transcription factor
NK2 homeobox 6, Csx2
GATA-binding factor 1
Hepatocyte nuclear factor 1 beta (HNF1B)
AT-binding transcription factor 1
AT rich interactive domain 5A (MRF1-like)
HESX homeobox 1, dimeric binding site

Special AT-rich sequence-binding protein 1, predominantly expressed in thymocytes, binds to matrix attachment regions (MARs)
TG-interacting factor belonging to TALE class of homeodomain factors
Meis1b and Hoxa9 form heterodimeric binding complexes on target DNA
THAP domain containing, apoptosis associated protein
Myelin regulatory factor
Non-palindromic nuclear factor I binding sites
MEL1 (MDS1/EVI1-like gene 1) DNA-binding domain 2
Composed binding site for Oct4, Sox2, Nanog, Tcf3 (Tcf711) and Sall4b in pluripotent cells
TEA domain family member 4, TEF-3
HMG box-containing protein 1
POU class 3 homeobox 3 (POU3F3), OTF8
Pit1, GHF-1 pituitary specific pou domain transcription factor
Yin and Yang 1 activator sites
Fork head related activator-2 (FOXF2)
E2F, involved in cell cycle regulation, interacts with Rb p107 protein
Aryl hydrocarbon / Arnt heterodimers, fixed core
CCAAT/enhancer binding protein (C/EBP), epsilon
Thyrotrophic embryonic factor / hepatic leukemia factor
CCAAT/enhancer binding protein (C/EBP), gamma
Tax/CREB complex
Hepatic leukemia factor
RB/E2F-1/DP-1 heterotrimeric complex
Octamer-binding factor 1
Jumonji, AT rich interactive domain 2 (JMJ)
LTSM elements with 7 bp spacer

Double homeobox protein 4
Homeobox protein engrailed (en-1)
Meis homeobox 1
Binding site for a Pbx1/Meis1 heterodimer
HMGA family of architectural transcription factors (HMGA1, HMGA2)
Growth factor independence 1
Sterol regulatory element binding protein 1 and 2
TGFB-induced factor homeobox 2-like, X-linked, dimeric binding site
Bipartite binding site of VDR/RXR heterodimers, DR1 sites
Constitutive androstane receptor / retinoid X receptor heterodimer, DR4 sites
Pdx1 (IDX1/IPF1) pancreatic and intestinal homeodomain TF
Even-skipped homeobox 1
ISL LIM homeobox 2
Homeobox B3 / Hox 2-gamma
Octamer-binding transcription factor-1, POU class 2 homeobox 1 (POU2F1)
Zebrafish PAX2 paired domain protein
Insulin promoter factor 1, pancreatic and duodenal homeobox 1 (Pdx1)
Transcription factor Jun-B
cAMP-responsive element binding protein 2
Thyrotrophic embryonic factor
Avian C-type LTR TATA box

Chapter 5

Expansion via duplication of a multiple-ligand transporter related gene family in cluster in teleost fish

Langyu Gu and Walter Salzburger

Chapter 5

Expansion via duplication of a multiple-ligand transporter related gene family in cluster in teleost fish

Langyu Gu, Walter Salzburger*

Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

*Correspondence:
Walter Salzburger, email: walter.salzburger@unibas.ch

Abstract

Gene duplications and the subsequent divergence of gene duplicates play an important role in the evolution of gene functions. Compared to individual gene duplication events, duplicated genes in clusters attract much more scientific attentions. An important question is about how and why do the genes keep in cluster during long evolutionary time. Here, we report a gene family, apolipoprotein D (ApoD), which only has one orthologous gene in tetrapods, but has expanded in teleost fishes via the teleost specific whole genome duplication (TS-WGD), lineage-specific gene duplications and genome rearrangement. Besides, these duplicated genes are located in two gene clusters in teleosts. Previous studies have shown that the *ApoD* gene in human might play a role as a multiple-ligand transporter (Weech et al. 1991; Rassart et al. 2000) and the gene has been suggested to be important in homeostasis and housekeeping functions (Weech et al. 1991). Our detailed expression profile analysis in cichlid fish, zebrafish and stickleback revealed that orthologous genes in the same physical order along their respective clusters exhibited similar tissue expression patterns. Combined with the results of positive selection detection in coding region, several models (eg. neofunctionalization, subfunctionalization, dosage effects) were proposed to explain the evolution of these duplicates, especially the new functions for the two putative novelties (lower pharyngeal jaw and anal fin pigmentation) in cichlid fish. The mechanisms behind the maintenance of the duplicates in cluster were discussed based on the functional and structural point of view, especially for the role of repeat elements. Taken together, the ApoD gene family provides an ideal model to study the evolution of new gene functions, gene duplication, gene cluster maintenance and their relationship with teleosts radiation.

Key words

gene duplication, gene cluster, TS-WGD, neofunctionalization, novelty, teleost fish

Introduction

Gene and genome duplication play an important role in speciation and adaptation by providing new genetic raw materials (Ohno 1970). The gene copies emerging from duplication events (including whole genome duplications) can undergo different fates, and a number of models have been proposed as to what might happen after duplication (Innan and Kondrashov 2010). In many instances, one of the duplicates becomes silenced via the accumulation of deleterious mutations (i.e. pesudogenization or nonfunctionalization (Ohno 1970)). Alternatively, the original pre-duplication function might be subdivided between the duplicates (i.e. subfunctionalization) (Force et al. 1999), or one of the duplicates might gain a new function (i.e. neofunctionalization) (Rastogi and Liberles 2005). Since the probability to accumulate beneficial substitutions is relatively low, examples for neofunctionalization are sparse. There are, nevertheless, examples for neofunctionalization, such as the expansion of repetitive regions in a duplicated pancreatic trypsinogen-like gene leading to a functional antifreeze glycoproteins in Antarctic notothenioid fish (Chen et al. 1997), or the alteration of a duplicated *opsin* gene leading to the evolution of trichromatic vision in primates (Dulai et al. 1999). Another selective advantage of gene duplication can occur due to the increased number of gene copies themselves (gene dosage effects) (Tang and Amon 2013), e.g. selection favors extra dosage of the transporters of glucose under nutrition limiting environment in yeasts (Lin and Li 2011).

Teleost fish, the most species-rich group of vertebrates, are characterized by a whole genome duplication event that occurred in their common ancestor (Jaillon et al. 2004; Kasahara et al. 2007; Schartl et al. 2013). This teleost specific whole genome duplication (TS-WGD) is responsible for the larger number of genes in teleost genomes compared to other vertebrate lineages and has been implicated with the spectacular diversity of this clade (Ohno 1970; Venkatesh 2003; Postlethwait et al. 2004). Noticeably, however, several teleost orders still contain few species, such as Elopomorpha and Osteoglossomorpha at the base of the teleost tree, challenging the view that the TS-WGD is directly responsible for speciation (Glasauer and Neuhauss 2014). It has thus been proposed that there was a time delay between the TS-WGD and taxonomic diversification (i.e. *time-lag model*) (Schranz et al. 2012). However, what are the factors responsible for this delay still a matter of debate. In addition, the relative importance of TS-WGD and lineage specific duplication on the abundance of duplicated genes in teleost is not clear. For example, by a thorough investigation of 37 gene families in fish, Robinson-Rechavi et al. (Robinson-Rechavi et al. 2001) found that only seven families were following a whole genome duplication pattern, while 30% families appear to have arisen independently in different lineages after TS-WGD. Therefore, more

evidence is required to establish a causal link between TS-WGD, gene abundance and teleost diversity (Venkatesh 2003).

Gene duplication resulting in so-called gene clusters (genes from same gene family physically closely linked in one chromosome (Garcia-Fernàndez 2005)), such as Hox gene clusters (Carroll 1995) and globin gene clusters (Proudfoot et al. 1980), have attracted considerable scientific attentions – not only because of cluster organization, but also because of the regulation mechanism of clustered genes in development, such as the spatial and temporal colinearity (Sproul et al. 2005; Gaunt 2015). The main questions are why and how are the duplicates maintained in a cluster instead of spreading randomly across the genome over evolutionary time scales? One possible explanation is the existence of global (i.e. affecting all genes of the cluster) enhancer elements located in the flanking upstream region of the entire gene cluster, as has been observed in the HoxD gene cluster in mice affecting hindbrain patterning (Spitz et al. 2003), and in the β-globin gene cluster in human (Levings and Bungert 2002). Another mechanism for gene cluster maintenance involves long and short range interactions between global regulatory elements and local promoters, respectively (Levings and Bungert 2002; Gaunt 2015; Deschamps 2016). An alternative proposal is that spatial collinearity of Hox gene cluster is to maximise physical segregation between active and inactive genes under the global regulation within the cluster (Gaunt 2015), such as "posterior" Hox proteins antagonise the function of more "anterior" ones (Narendra et al. 2015). However, there are still many open questions related to the evolution and maintenance of gene clusters to be answered: 1) what are the selective advantages for keeping genes in a cluster organization? 2) How are newly duplicated genes integrated into the gene cluster's pre-existing regulatory module? 3) Why the genes in certain gene family are frequently expanded and kept during evolution? 4) Is the position of a gene in a cluster relatively more important than gene sequence identity itself for performing the gene's function? For example, the competition between HoxD promoters for digit enhancer was found to depend on the position in the gene cluster instead of gene identity (Kmita et al. 2002). While the transcription of β-globin clusters was found to vary in a gene-specific manner (Dillon et al. 1997; Tanimoto et al. 1999). Therefore, it seems that the relative importance of gene position and gene sequence identity in the cluster in performing gene functions is context-dependent.

Here, we first report the expansion of a gene family, apolipoprotein D (ApoD) in teleosts. ApoD genes belong to the lipocalin superfamily of lipid transport proteins (Ayrault Jarrier et al. 1963; Rassart et al. 2000). *ApoD* gene has been suggested to function as a multiple-ligand, multifunctional transporter (Weech et al. 1991; Rassart et al. 2000) and to be important in the homeostasis and housekeeping function of most organs (Weech et al. 1991). Tetrapods possess a

single copy of *ApoD*, which is expressed in multiple tissues, most notably in brain and testis (see e.g. (Drayna et al. 1986; Provost et al. 1991; Rassart et al. 2000)), and was suggested to be involved in the central and peripheral nervous systems (Rassart et al. 2000). *ApoD* was also found as a transcriptional target of the p53 genes in human (Sasaki et al. 2009). While there is only one orthologous gene in tetrapods, teleost genomes show varying numbers of duplicates, which are located in two clusters based on Ensembl database (http://www.ensembl.org/). One orthologous gene of the ApoD family was previously found highly expressed in an innovative pigmentation patttern on the anal fin of cichlid fish, the blotch pattern in *Callochromis macrops* (Gu et al., in preperation) (Chapter 3). However, the functions of most duplicates in teleosts are still unknown. In this study, we used phylogenetic reconstruction, gene expression profiling detection, adaptive sequence evolution detection, repeat elements detection, and conserved noncoding elements (CNE) detection in the upstream flanking region of these clusters to examine the evolutionary history of ApoD gene family.

Materials and Methods

Samples

Laboratory strains of *C. macrops, Astatotilapia burtoni, Haplochromis cf stappersii, Labidochromis careulus, Cyphotilapia frontosa, Cyprichromis leptosoma, and Pseudocrenilabrus philander* were kept at the Zoological Institute of University of Basel (Switzerland) under standard conditions (12h light/12h dark; 25°C) and permit 1010H issued by the cantonal veterinary office Basel. Prior to tissue dissection, specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following an approved procedure (permit nr. 2317 issued by the cantonal veterinary office). Individuals of all other specimens *(Boulengerochromis microlepis, Bathybates graueri,* and *Perissodus microlepis)* were collected at our field-site in the South of Lake Tanganyika and processed in the field following our standard operating procedure (Muschick et al. 2012). Field work was covered by permits issued by the Department of Fisheries, Republic of Zambia. Zebrafish lab strains were provided by Markus Affolter, Biozentrum, University of Basel. Stickleback samples were provided by Dario Moser collected from Nideraach and Romanshorn in Switzerland.

DNA and RNA extraction

Genomic DNA was extracted following the DNeasy Blood & Tissue kit's standard protocol (Qiagen); isolation of RNA was performed according to the TRIzol protocol (Invitrogen, USA) after incubating the dissected tissues in 750ul of TRIzol at 4°C overnight. The tissues were then homogenized with a Bead-Beater (FastPrep-24; MP Biomedicals, France). Subsequent DNase

treatment was performed with DNA-free™ Kit (Ambion, life technologies). RNA quantity and quality was determined with a NanoDrop 1000 spectrophotometer (Thermo Scientific, USA). cDNA was produced using the High Capacity RNA-to-cDNA kit (Applied Biosystems, USA).

Gene sequencing

Long-range PCR primers were designed based on local alignments of the five available cichlid genomes from Ensembl using Primer premier 5 (Lalitha 2000) to amplify the coding gene region of each duplicated gene in cichlids (Supplementary Table 1). For PCR amplification, we used the Phusion® High-Fidelity PCR Master Mix with HF buffer (BioLabs) in a touch down annealing process (Supplementary Table 1). PCR products were then visualized via gel electrophoresis in a 1% agarose gel using GelRed (Biotium, USA) under the conditions 100 V and 30 min. To sequence these long amplicons, we used PGM Ion Torrent (www.lifetechnologies.com) with next generation DNA sequencing. Prior to library construction, we purified the PCR fragments from the gel using the GenEluteTM PCR Clean-up kit (Sigma). Genomic DNA was first fragmented followed by adapter ligation and nick-repair using Ion Plus Fragment Library Kit (Life Technologies). We barcoded each PCR fragment individually using Ion Xpress™ Barcode Adapters Kits (Life Technologies); pools consisting of 1410 individual PCR fragments (Gu et al., in preperation) (Chapter 4) were then sequenced on three chips on the Ion Torrent X platform using the sequencing 400 kit and ION 316 TM Chip Kit V2 (Life Technologies). The target read length was 400bp.

Assembly of DNA fragments

We first used the FASTQ Quality Trimmer in Galaxy to trim low quality reads in both 5' and 3' ends by applying a sliding window analysis (window size 2, step size 1, minimum quality score: 20). After quality trimming, we used Filter FASTQ to filter out short reads applying a cut-off size of 40bp. The software Geneious v8.1.3 (http://www.geneious.com, (Kearse et al. 2012)) was used for *de novo* assemblies and annotation for each gene in each species. Briefly, we applied the Medium Sensitivity/Fast in Custom Sensitivity with default parameters to do the *de novo* assembley and set the maximum gap size to 20,000 bp to expand reads across introns. The corresponding gene sequences in Tilapia from Ensembl Release 84 (Flicek et al. 2014) were downloaded and used as reference (Table 1), and the assembled consensus sequences were annotated by mapping them against this reference. Afterwards, we mapped the original reads to the annotated consensus sequences to re-check the assembled sequences.

In silico screening

Orthologs and paralogs of duplicated genes from the ApoD gene family were extracted from Ensembl Release 84 (Flicek et al. 2014) and NCBI database for human, coelacanth, spotted gar, zebrafish, fugu, medaka, stickleback and four available cichlid species (*Astatotilapia burtoni, Pundamilia nyererei, Neolamprologus brichardi, Oreochromis niloticus*) (Table 1). Furthermore, to determine gene copy numbers in the respective genomes, we used the paralogue coding sequences in each species as query in a tblastx search against the available corresponding Ensembl genomes using default parameters. For all unannotated positive hits we extracted a region spanning ca. 2 kb and predicted the coding region using online tool geneid 1.2 (http://genome.crg.es/software/geneid/geneid.html) to identify exon and intron boundaries and to annotate them in the extracted scaffold using Geneious v8.1.3. We then extracted the predicted coding sequence region and subjected it to BLAST searches to retrieve the corresponding cDNA sequences. All positive hits with a "max score" greater than 200 were re-mapped to the corresponding predicted gene to recheck and, in case of inaccurate predicted exon-intron boundary, modify the predicted annotation. Furthermore, we checked the neighboring genes to the newly predicted gene to examine whether synteny is as same as in other orthologous genes in all species. Besides, to check whether there are unannotated genes or gene losses in Amphioxus and Lamprey, we also used coding sequence of genes from human and gar and their neighbouring genes perform tblastx searches against the genomes of Amphioxus and Lamprey in Ensembl.

Phylogenetic reconstruction to infer gene duplication

Phylogenetic and evolutionary inference can be severely misled if recombination or gene conversion is not accounted for (Lapierre et al. 2016). Hence, we first screened the sequence alignment using GARD implemented within Hyphy (Kosakovsky Pond et al. 2006) to check gene conversion. To confirm the orthologs and paralogs of these duplicated ApoD genes, we constructed a maximum likelihood (ML) tree using PAUP 4.0 (Swofford 2002) with human as outgroup. First, we did the alignment using translated amino acids of all the duplicates with Mega v4.0 (Tamura et al. 2007) with default parameters. Then the corresponding nucleotide sequences were used for phylogeny reconstruction. The best-fitting model of nucleotide substitution was determined with the corrected Akaike Information criteria and likelihood ratio tests conducted in jModeltest v2.1.4 (Guindon and Gascuel 2003; Darriba et al. 2012).We then performed bootstrap analyses with 200 replicates.

Adaptive sequence evolution

To examine whether adaptive sequence evolution occurred following duplications, and how the duplicates evolved in different lineages, we calculated the ratio of non-synonymous to synonymous substitutions (ω or dN/dS) with a priori partitions (Supplementary Table 2) using codeml implemented in the PAML package (Yang 1997; Yang 2007). In these analyses, $0<\omega<1$ is consistent with purifying selection, $\omega=1$ suggests neutrality, and $\omega>1$ indicates positive selection. All the model comparisons in PAML were performed with fixed branch lengths (fix_blength=2) derived under M0 model in PAML. Alignment gaps and ambiguity characters (eg. Y for T or C) were not removed (Cleandata=0); codon frequencies were approximated using the F3x4 calculation. The branch-site model was used to test positive selection affecting a few sites along particular lineages (foreground branches). Comparisons are made between the modified model A (model=2 NSsites=2) with corresponding null model with $\omega_2=1$ fixed (fix_omega=1 and omega=1). A likelihood ratio test (LRT) was then used to calculate a chi-square approximation, and p/2 value was used considering mixture distribution (see PAML manual). The Bayes empirical Bayes (BEB) was used to identify which sites are under positive selection.

Repeat elements and conserved noncoding elements (CNE) detection

To check whether there are conserved noncoding elements (CNE) in the 5' upstream flanking region (800kb-1Mb) of teleost ApoD gene clusters, we extracted the available corresponding genome sequences data from Ensembl database of cichlids (*Maylandia zebra*, *Pundamilia nyererei, O. niloticus*), medaka, fugu and stickleback (Table 1). Comparative analyses were done with MVISTA (Mayor et al. 2000; Frazer et al. 2004) using the LAGAN alignment tool (Brudno et al. 2003) with *O. niloticus* as the reference. We applied the repeat masking option with fugu as reference. To compare the distribution of the repeat elements in the local region where the duplicated genes are located in different clusters and different lineages, the corresponding genome region were scanned using the RepeatMasker online server http://www.repeatmasker.org/cgi-bin/WEBRepeatMasker, using the cross_match method with zebrafish as the reference.

Gene expression in cichlid fish, stickleback and zebrafish

To detect the gene expression profile of different duplicates of the ApoD gene family, we performed quantitative polymerase chain reaction (qpcr) experiments for each member of the gene family in zebrafish (heart, brain, testis, skin, gill, eye, ovary), stickleback (testis, ovary, spleen, eye, brain, liver, gill, dorsal spine, bottom spine, plates) and a representative of cichlids (*A. burtoni*) (testis, ovary, spleen, eye, brain, eggspots, non-eggspots in the same anal fin, liver, lower pharyngeal jaw (LPJ), gill). The comparative cycle threshold method was used to calculate

differences in expression between the different samples using the housekeeping gene elongation factor 1 alpha *(elfa)* for zebrafish (McCurley and Callard 2008), ubiquitin (*ubc*) for stickleback (Hibbeler et al. 2008), and the ribosomal protein L7 (*rpl7*) for cichlids (Santos et al. 2014) as endogenous control. qPCR experiments were run on a StepOnePlusTM Real-Time PCR system (Applied Biosystems) using the SYBR Green master mix (Roche, Switzerland) with an annealing temperature of 58°C and following the manufacture's protocols. Primers were designed with the software GenScript Real-Time PCR (Taqman) Primer Design available at https://www.genscript.com/ssl-bin/app/primer (Supplementary Table 3).

Results

Ion torrent sequencing produced ~ 120,000 raw reads. After trimming and filtering, ~100,000 reads were left. Finally around 160 to 900 raw reads were produced for each PCR product (exact reads numbers see Supplementary Table 4). Raw reads are available from the Sequence Read Archive (SRA) at NCBI under the accession number SRA… Together with genes predicted from Enmsebl database and newly predicted genes from our *in silico* screening (one new gene in cichlid fish, two new genes in stickleback, See Table 1), different numbers of duplicates in two clusters were found in different teleosts (Fig. 1).

Table 1 Duplicated genes in ApoD gene family and their corresponding Ensembl numbers

species	gene name	Ensembl
tilapia	copyA1	ENSONIG00000018635
	copyA2	ENSONIG00000006468
	copyA2a	ENSONIG00000006461
	copyA2b	ENSONIG00000006464
	copyA2c	ENSONIG00000006466
	copyB1	predicted
	copyB2a	ENSONIG00000018634
	copyB2b	ENSONIG00000018633
stickleback	copyA1	ENSGACG00000017003
	copyA2SI	ENSGACG00000005360
	copyA2SII	predicted
	copyB1	ENSGACG00000000558
	copyB2a1	predicted
	copyB2a2	ENSGACG00000016996
	copyB2a3	ENSGACG00000016998
	copyB2b	ENSGACG00000016991
medaka	copyA1	ENSORLG00000017938
	copyA2m1	ENSORLG00000013123
	copyA2m2	ENSORLG00000013130
	copyA2m3	ENSORLG00000013153
	copyB2a	ENSORLG00000017942
	copyB2b	ENSORLG00000017946
fugu	copyA1	ENSTRUG00000004136
	copyA2	ENSTRUG00000008463
	copyB1	ENSTRUG00000017918
	copyB2a	ENSTRUG00000004180
	copyB2b	ENSTRUG00000004313
zebrafish	copyA1	ENSDARG00000060345
	copyA2	ENSDARG00000057437
	copyB2	ENSDARG00000060350
gar	copyA	ENSLOCG00000007072
	copyB	ENSLOCG00000007087
coelacanth	gene	ENSLACG00000022488
human	gene	ENSG00000189058

Evolutionary history of different duplicates in different lineages

Ensembl data mining and *in silico* screening showed that different numbers of duplicates with lineage specific duplication were located in two clusters in two chromosomes in teleosts (Fig. 1). Interestingly, orthologous genes are located in the same physical order along their respective clusters in different teleosts (Fig. 1), and the tandem lineage specific duplicated genes are always located next to each other. Besides, since gar diverged from the ancestor of teleost before TS-WGD, the two clusters most likely emerged via the TS-WGD.

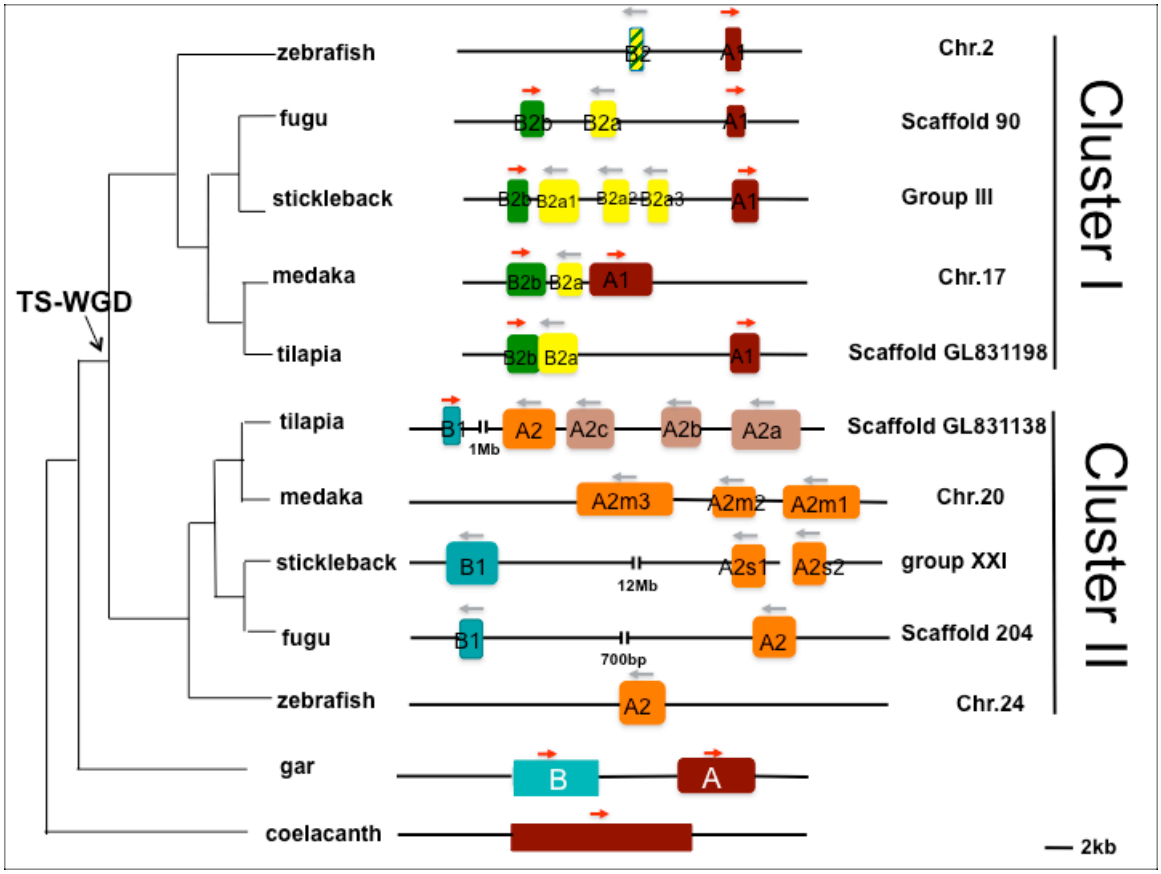

Fig.1 ApoD gene family clusters organization in teleosts and tetrapods. Each block represents a single gene copy. Genes with the same color in different species represent orthologous, while within same species represent sister paralogues. Red, orange and brick colors represent genes deriving from the same common ancestor. Blue, yellow and green colors represent genes coming from another common ancestor.

Phylogeny reconstruction based on coding sequence of the duplicates in different species suggested that there is only one copy in tetrapod, whereas a tandem duplication produced two copies (*copy A* and *B*) in one cluster in gar. In the course of the TS-WGD, two clusters with different numbers of duplicates emerged. *Copy A1* showed slow evolutionary rate (as indicated by short branch lengths), and, thus, more closely resembles the ancestral sequence. Its sister *copy A2* was further duplicated in a lineage-specific manner in cichlids and medaka to produce three additional copies (*A2a, A2b* and *A2c*), and copies (*A2m1, A2m2, A2m3*), repectively. The homolog to *copy B* in gar duplicated further in teleosts producing two paralogs, *copy B1* and *copy B2*. *Copy*

B2 duplicated further in fugu, stickleback, medaka and tilapia and resulting into duplicates *B2a* and *B2b*. In addition, lineage specific duplication of *B2a* occurred in stickleback (*copy B2a1, B2a2, B2a3*). Besides, duplicates (*copy A1- copy A2, A2a, A2b, A2c, A2m1, A2m2, A2m3*; *copy B1-copy B2, B2a, B2a1, B2a2, B2a3, B2b*) are located in two different clusters in teleosts, which could be mostly due to TS-WGD. Noticeably, *copy A* showed accelerated evolutionary rate (longer branch) compared to *copy B* clade in teleosts, and this is mostly due to the three cichlid lineage-specific duplicates (*copyA2a, A2b, A2c*) (Fig. 2).

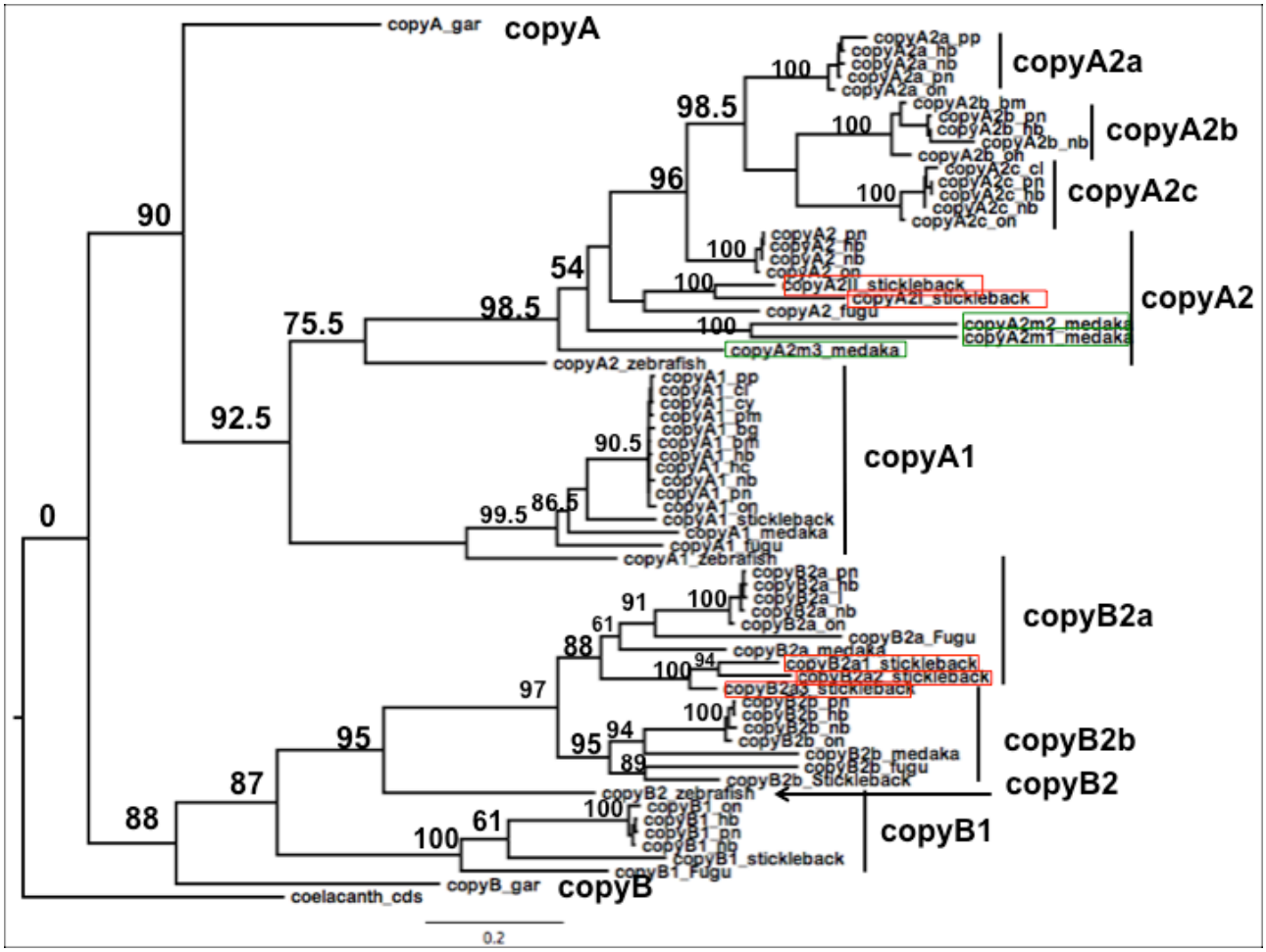

Fig.2 Maximum likelihood phylogenetic tree reconstruction to infer gene duplication. Red rectangular marks stickleback-lineage specific duplication. Green rectangular marks medaka-lineage specific duplication. Bootstrap value larger than 50 were marked on the branch.

Adaptive sequence evolution for duplicates in different lineages

PAML analyses, according to the branch-site model, significant positive selection (ω>1) was detected in several branches prior to duplication (e.g. ancestral branch of copy *B2a* and *B2b*), and the strongest signal for positive selection was found in the branches leading to the lineage-specific duplicates in cichlids, sticklebacks and medaka (Fig.3, Supplementary Table 2).

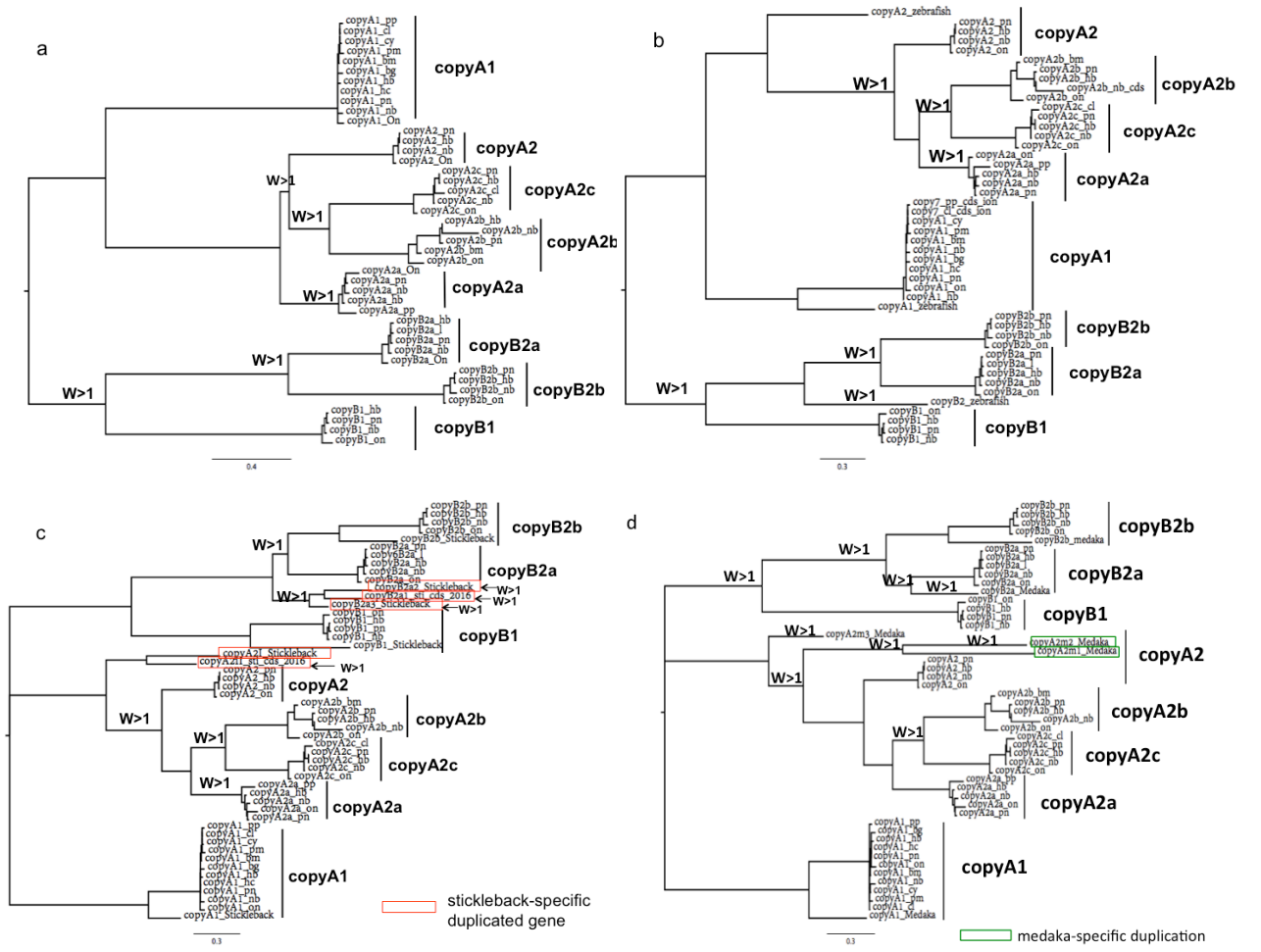

Fig.3 Positive selection detection in cichlid fish (a), zebrafish (b), stickleback (c) and medaka (d). Red rectangular marks stickleback-lineage specific duplication. Green rectangular marks medaka-lineage specific duplication. Branches with w>1 were marked.

CNE detection in gene desert and repeat elements detection in local region

A large number of CNEs were identified in the 5' upstream flanking region of the whole gene clusters (5' upstream 800kb to 1Mb) in teleosts. These include repeat elements such as short interspersed elements (SINE), long interspersed nuclear element (LINE), long terminal repeat (LTR), etc. (Fig. 4). Repeat elements in local region in clusters showed several patterns (Fig. 5, Supplementary Table 5): 1) Species with higher numbers of duplicates have higher numbers of repeat elements. 2) Clusters with lineage-specific tandem duplicates have higher repeat numbers than the other cluster in cichlid fish and stickleback. 3) In human, each exon is surrounded by repeat elements, especially *Alu* repeats. 4) The numbers of the repeat elements were much higher in cichlid fish than other teleosts.

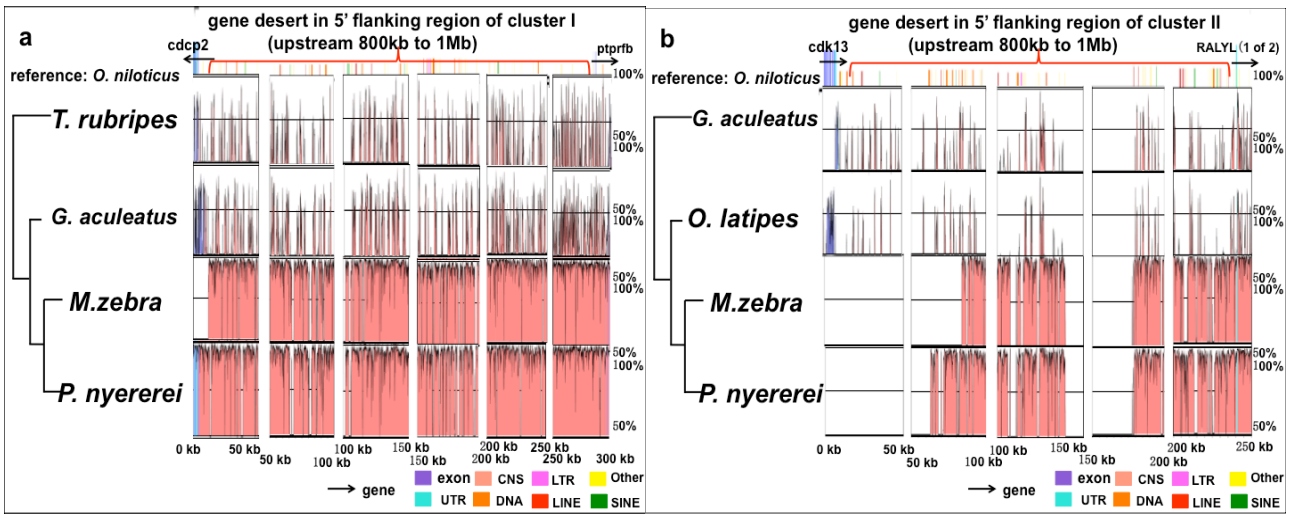

Fig.4 MVISTA plots for the comparison of the 5' homologous regions of upstream flanking gene desert region of two clusters in teleosts. Blue peaks indicate conseved coding reigons, pink peaks represent conserved noncoding regions.

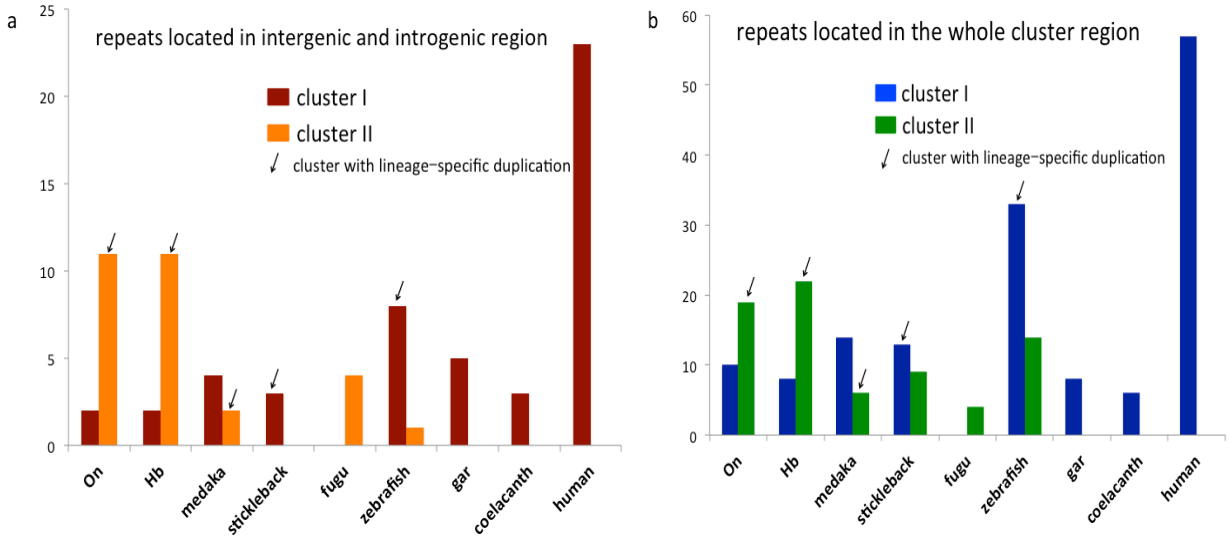

Fig. 5 Repeat elements distribution in clusters in different species. (a) distribution of repeat elements in intergenic region and introgenic region. (b) distribution of repeat elements in the whole cluster region, including 5' 10kb flanking region and 3' 10kb flanking reigon.

Gene expression in zebrafish, stickleback and cichlid fish

To see the function of each duplicate, we performed qpcr experiments for each duplicate in different tissues in zebrafish, stickleback and a cichlid (Fig. 6 and 7). The following general observations were made: 1) Orthologous genes in the same physical order along their respective clusters exhibited similar expression patterns. 2) The expression patterns became much more specific with increasing lineage specific duplicates. For example, *copy B2* in zebrafish was expressed in testis, skin and ovary. With lineage specific duplication, the expression of paralog B2a became specific in testis in cichlid fish. Although we do not know whether *copy B2* was also expressed in spleen in zebrafish (compared to *copy B2b* in cichlid fish which was expressed in ovary, testis and spleen), the function apparently became much more specific in spleen tissue in

stickleback with lineage specific duplications (*copy B2b, B2a1, B2a2, B2a3*). 3) Expression redundancy was found in lineage specific duplicates in cichlids and sticklebacks. For example, the four copies of B2 (*copy B2b, B2a1, B2a2, B2a3*) were all highly expressed in spleen tissue in stickleback. 4) Phylogenetic reconstructions showed that *copy B1* of cichlid fish clustered together with *copy B1* of stickleback. However, the *copy B1* in cichlid fish was expressed in liver instead of testis in stickleback. Considering *copy B1* in cichlid fish was involved in the inverted chromosome region as shown in Ensembl database (Fig.1), one possibility of the expression pattern change could be due to change of upstream *cis*-regulatory element, which needs further investigation. 5) Interestingly, it seems that new functions were obtained for duplicates copies *A2, A2a1, A2a2, A2a3* in cichlid fish, especially in the lineage-specific duplicates copies *A2a1, A2a2, A2a3*, which were highly expressed in LPJ. LPJ is a second set of tooth-bearing jaws and is functionally used to process food, which could be related to ecological speciation in cichlid fish (Muschick et al. 2011; Muschick et al. 2012).

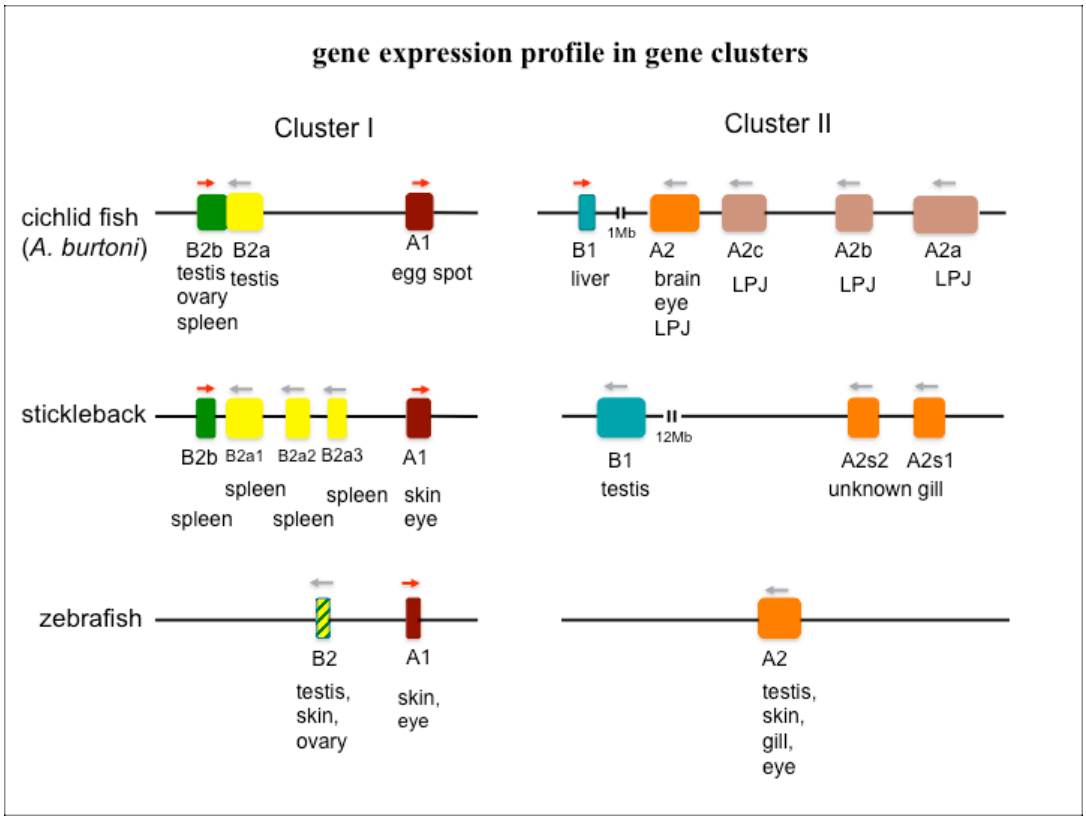

Fig. 6 Simplified digram of gene expression profile in gene clusters in cichlid fish, stickleback and zebrafish. Each block represents a single gene copy. Genes with the same color in different species represent orthologous, while within same species represent sister paralogues. Red, orange and brick colors represent genes coming from the same common ancestor. Blue, yellow and green colors represent genes coming from another common ancestor.

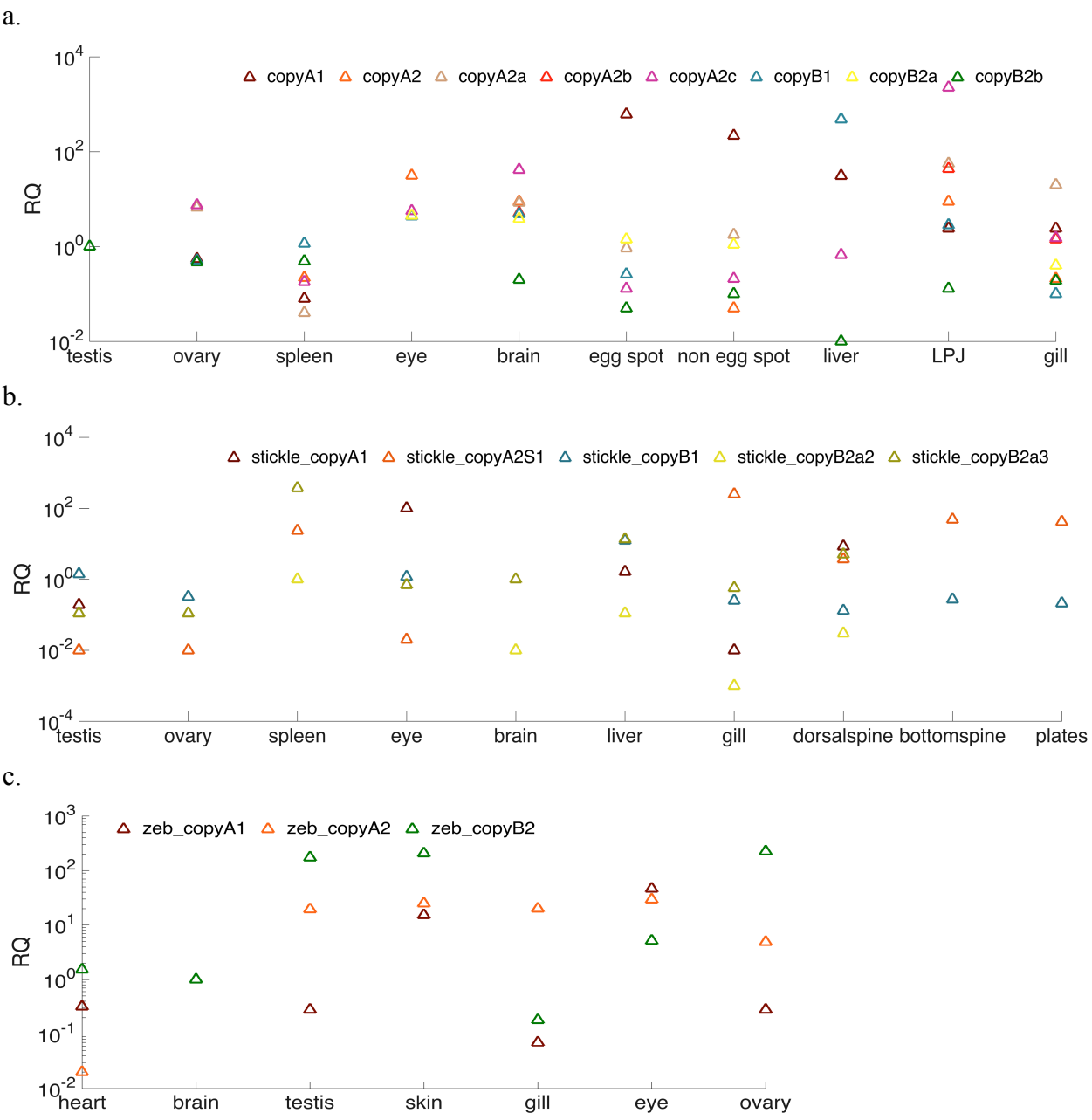

Fig. 7 Gene expression profile for duplicated genes in cichlid fish, stickleback and zebrafish. a. simplified expression profile for duplicated genes in cluster. Each block represents a single gene copy. b. Relative quantitation (RQ) plot of duplicated genes in cichlid fish. c. RQ plot of duplicated genes in stickleback. d. RQ plot of duplicated genes in zebrafish. LPJ: lower pharyngeal jaw.

Discussion

In this study, we focused on the evolution of ApoD genes in teleost fish, which evolved via gene duplication. To reconstruct the evolutionary history of the duplicates and to examine the mechanisms of the expansion of ApoD gene family in teleosts, we performed a series of analyses including phylogenetic reconstruction and syntenic data analysis. Based on these results, our hypothesis about the expansion of this gene family is in the following: there was an initial tandem

gene duplication in the common ancestor of gar and the teleosts producing two tandem duplicates in one gene cluster. Followed by TS-WGD, and subsequent genome rearrangements and lineage-specific tandem gene duplications and losses, different numbers of duplicates were produced, which are located in two clusters in different teleosts (Fig. 1). Based on gene expression profile analysis and positive selection detection in cichlid fish, zebrafish and stickleback, neofunctionalization, subfunctionalization and dosage effects appear to be responsible for the evolution of these duplicated genes. New functions related to two innovative traits (lower pharyngeal jaw and anal fin pigmentation pattern) were found in cichlid paralogs duplicates. The mechanisms behind the maintenance of the expansion of this gene family in clusters are discussed based on functional and structural point of view, especially for the role of repeat elements.

1. Evolution of the duplicated genes in ApoD gene family

Based on the prediction from Ohno's neofunctionalization hypothesis (Ohno 1970), asymmetry of divergence between old and new copies is expected. This can happen via accelerated rate of molecular evolution in the duplicated copy, with advantageous mutations possibly being under positive selection (ω>1). This pattern was found in the duplicate of *copy A1* and its paralogs in our study (Fig. 3). The original *copy A1* showed conserved expression in skin in teleosts, and has the slowest evolutionary rate (shortest branch), thus it resembles the ancestral sequence more closely (Fig. 3). On the other hand, positive selection (ω>1) and accelerated sequence evolution apparently occurred in the early stage after duplication (branches of *copy A1* paralogs) (Fig.3), which might be due to adaptive evolution immediately following the duplication. Indeed, gene expression profiles suggested that new functions evolved in these paralogs (testis, ovary, spleen, LPJ, brain, liver, gill) compare to the original *copy A1* (skin, eye) (Fig. 6).

Subfunctionalization might be responsible for the evolution of copy *B2* and its paralogs (Fig. 6). According to the prediction of duplication-degeneration-complementation (DDC) model (Innan and Kondrashov 2010), evolutionary rates of both copies were nearly symmetrical (similar branch length). Although the 'escape from adaptive conflict' (EAC) model (Innan and Kondrashov 2010) could also explain the symmetrical evolutionary rate pattern between paralogs, this model predicts that each duplicated gene should be under positive selection (ω>1), which is obviously not the case here (Fig. 3). Besides, the ancestral branch of copy *B2a* and *B2b* was found to be under under positive selection, which could be because of the adaptive advantages after functional specialization. For example, copy *B2* in zebrafish was expressed in testis, skin and ovary. The expression profile in its paralogs *copy B2a* and *B2b* became much more specific, with *B2a* mainly expressed in testis in

cichlid fish, and *B2a1, B2a2, B2a3, B2b* highly expressed in spleen (Fig. 6 and 7). Noticeably, here we do not know whether it was also expressed in spleen because we did not test it in zebrafish. But since its paralogs *B2b* in cichlid fish and stickleback were both expressed in spleen, the original expression profile of *copy B2* could also include spleen.

In addition, most lineage specific-duplicates in cichlid fish (copy *A2a, A2b, A2c*), stickleback (copy *A2s1, A2s2,* copy *B2a1, B2a2, B2a3*) and medaka (copy *A2m1, A2m2, A2m3*) appear to have evolved under positive selection (Fig. 3), and could thus have adaptive advantageous functions, e.g. related to a putative ecological related tissue, LPJ in cichlid fish (Muschick et al. 2011; Muschick et al. 2012) and immune defense in stickleback (Huang et al. 2016). The functions of lineage specific duplicates in medaka need further investigation. Since these lineage-specific duplicates could be at the early stage after duplication and in this regard, the ApoD gene family represents an ideal model system to explore the interplay between gene duplication and functional varieties. In addition, since functional redundancy was found in these lineage-specific duplicates (LPJ in cichlid fish, spleen in stickleback), dosage effects could also be responsible for the maintenance of these lineage specific duplicates, because an increase in expression of a gene through gene duplication can also be adaptive (Tang and Amon 2013). Noticeably, another possibility to keep gene redundancy is to create new regulatory hierarchies, such as the duplication of sex Determining gene *dmrt1* in medaka (Herpin et al. 2010).

2. Mechanisms of gene family expansion

Since the chance of beneficial mutations are relatively low, most of time duplicated genes will become pseudogenes or be lost during evolution (Ohno 1970). But this is not the case for duplicated genes in ApoD gene family, which were kept with different functions. Then the question is, why the genes in this family are so frequently expanded? On the one hand, from a functional point of view, it has been suggested that gene families related to transcription factors, steroids and molecular transporters often have a large numbers of genes due to duplication, such as the Hox gene family (Carroll 1995) or the β-globin gene family (Proudfoot et al. 1980). This might be partially related to their binding ability to regulate downstream target genes. ApoD in tetrapod was suggested to be related to multiple-ligand transporter, and is important in the homeostasis and housekeeping functions of many organs (Weech et al. 1991). Gene duplication in this gene family could be advantage for regulating downstream genes. Besides, duplication of genes that mediate the interaction between organism and environment, such as stress response genes (Kondrashov 2012), sensory genes (Horth 2007), transport genes (Ghimire-Rijal et al. 2014), etc. might be also

beneficial to be kept, which might be also the case for ApoD gene family, since it has been suggested to be functioned as a multi-ligand transporter (Weech et al. 1991; Rassart et al. 2000).

On the other hand, from a structural point of view, a large number of repeat elements were found in the local flanking regions of the two gene cluters in teleosts (upstream 10kb and downstream 10kb). These repeats include retrotransposons, DNA transposons and satellites (Fig. 4, Supplementary Table 5). It has been suggested that repeat elements are the main causes of gene family expansion and contraction, such as has been proposed for the rapid and dramatic expansion of the mouse *Abp* gene region (Janoušek et al. 2013). The increase of repeat elements densities may cause instability in the genome region since they are highly homologous sequences to promote recombination and genomic rearrangements. Indeed, we found that the cluster with higher numbers of duplicated genes always possesses higher numbers of repeats (Fig. 5). Our study promotes further questions to be answered: 1) When does these repeat elements invade different lineages? eg. before the duplication or after the duplication? 2) Why ApoD gene cluster regions are readily invaded by repetitive DNA? 3) What are the exact mechanism related to these elements to drive the ApoD gene duplication in different lineage? 4) Except the function to expand gene duplication, do these elements also play a role in regulating gene expression? 5) Is the duplication still continuing? For example, clusters in cichlid fish still harbor many repeat elements around the lineage specific duplicates (Fig.5). 6) How did these elements drive the inversion and genome rearrangement in the corresponding chromosome? 7) Why the distribution densities of the repeats are different in different lineages? For example, the number of repeats is much higher in cichlid fish than other teleosts (Fig. 5). It has been suggested that three waves of TE insertions occurred in cichlid genomes (Brawand et al. 2014), but haplochromine cichlids showed an increased efficiency in purging deleterious TE insertions more recently (Brawand et al. 2014). If it is the case, then the retention of these large numbers of repeat elements in ApoD family further suggest that they might play a crucial role and are therefore not purged out. In human, each exon is surrounded by repeat elements, especially *Alu* repeats (Supplementary Table 5). It has been suggested that *Alu* repeats play a role in genome recombination, gene expression, insertion mutations, etc. (Batzer and Deininger 2002). It will be interesting to see their roles in ApoD gene family evolution.

Another question is why are these duplicated genes are always located in clusters instead of spreading randomly in genomes during long evolutionary period? In addition, the sister duplicated genes are always located next to each other (Fig. 1). And orthologous genes in the same physical order along their respective clusters exhibited similar tissue expression patterns (Fig. 6). Several mechanisms were proposed to explain these phenomenons. 1) Global regulatory element in the

flanking region could be responsible for the cluster pattern (Montavon and Duboule 2013). During initial duplication stage, genes were relatively similar to each other, so that they might share common regulatory pathways and can be easily integrated into the cluster. Indeed, many CNEs were found in the flanking region of ApoD gene cluster, which might represent a gene desert (Ovcharenko et al. 2005) in our study (Fig. 4). This region could play a role as regulatory landscapes, but this would need further investigation. 2) Long range and short range regulatory elements interactions can affect gene expression in cluster (Gaunt 2015; Deschamps 2016). As mentioned above, many repeat elements were also found in the local region of the cluster (Fig. 5 Supplementary Table 5) in our study. These local repeats could play a role as regulatory element interacting with the long range CNE in the gene desert to affect gene expression of each duplicates, like Hox gene clusters (Gaunt 2015; Deschamps 2016). 3) The role of genome duplication cannot be ignored. Our results suggested that the numbers of duplicates are much larger after TS-WGD than before TS-WGD (Fig. 1). This phenomenon could be due to the invasion time of repeat element (eg. after TS-WGD) or could be due to homologous recombination in sister chromosome (Reams et al. 2014) after TS-WGD. Besides, compared to random gene duplications, cluster duplications can trigger potential regulatory innovations owing to the presence of an already "competent" locus architecture (Duboule 2007).

3. ApoD gene cluster and teleost speciation and adaptation, especially for cichlid fish

Interestingly, based on phylogenetic reconstruction, genes in the two clusters showed different expression patterns in a cluster manner in cichlid fish. The cluster possessing cichlid lineage-specific duplicated genes and that are expressed in LPJ showed accelerated evolutionary rate (longer branch) and were under positive selection ($\omega>1$). Based on the phylogenetic analyses, they are apparently a "young cluster". While the other cluster whose duplicates were related to ovary, testis and a pigmentation on the anal fin (Gu et al., in preperation) (Chapter 3) showed slower evolutionary rate (shorter branch), so that they mostly resembles the ancestral cluster; this is a signature of an "old cluster". Actually, LPJ and anal fin pigmentation patterns are two kinds of evolutionary innovations in cichlid fish (Salzburger, 2009), with LPJ related to ecological niche adaptation (Muschick et al. 2011; Muschick et al. 2012) and anal fin pigmentation related to sexual selection (Fryer and Iles 1972; Theis et al. 2012; Theis et al. 2015). In this case, it seems that functions related to natural selection and sexual selection were distributed in paralogs in different clusters. Actually, both LPJ and pigmentation are supposed to be derived from neutral crest cell (Green et al. 2015), so that it is not unexpected if they share common gene network.

In addition, the TS-WGD was suggested to be related to adaptive radiation in teleosts (Ohno 1970; Venkatesh 2003; Postlethwait et al. 2004). On the other hand, some teleost clades that experienced TS-WGD still possess poor species richness (Glasauer and Neuhauss 2014), challenging this view. Actually, the majority of the present teleost species originated quite late after TS-WGD, which is compatible with the "*time lag model*" (Schranz et al. 2012). Therefore, it seems that there is no direct link between TS-WGD and adaptive radiation. Instead, other factors such as ecological ones might be crucial. For example, although the radiation of Acanthomorpha (the most diverse group of teleosts) happened in the oceans, they are descended from freshwater ancestors (Chen et al. 2014). The transition from fresh to saltwater is a major adaptive step and once this adaptive function is taken, oceans provide a wide range of ecological niches to radiation. In our study, we showed that some duplicates of ApoD gene family were related to immune defence (spleen in stickleback), freshwater-marine water transition (plates in stickleback), and feeding strategy (LPJ in cichlid fish). Actually, TS-WGD could set the stage for adaptive radiation (Glasauer and Neuhauss 2014) by providing large raw genetic materials at the first place, and the following ecological related mutations accumulation might be the direct factors. In this case, newly duplicated ApoD genes in cichlid fish and stickleback could provide an ideal model to test the hypothesis. Besides, it is also supposed that lineage tandem specific duplication is much more important than TS-WGD. For example, previous studies showed that most gene families occurred independently in different lineages, and only a small fraction of gene families were found where duplications arose in a common ancestor (Robinson-Rechavi et al. 2001). In our study, we also found many lineage specific duplications with new functions, which seems to support this hypothesis. However, it seems that because of TS-WGD, one of the two clusters can evolve relatively freely as we showed above (accelerate evolutionary rate, Fig.3) and more duplicates were produced after TS-WGD. Besides, fixation of duplications are much more common in genome regions where rates of duplicating mutations are elevated due to the presence of already-fixed duplications such as "snow-ball" effect (Kondrashov and Kondrashov 2006). In this case, TS-WGD could prompt the chance of tandem local duplications. Therefore, although TS-WGD might be not directly link to teleosts radiation, it could be a crucial pre-request.

Conclusion

In this study, we first report the expansion of ApoD gene family via gene duplication in teleosts and that the genes are organized in two clusters. We then examined the gene expression profile of these duplicates in cichlid fish, stickleback and zebrafish. We found that subfunctionalization, neofunctionalization and gene dosage effect could explain the evolutionary

history of these duplicates. In addition, repeat elements in the upstream flanking gene desert region and local region could be responsible for the cluster maintenance. Besides, new functions for the two putative innovations (LPJ and anal fin pigmentation) in cichlid fish were found, which might be related to their adaptive radiation. Further functional experiment will be helpful to see the detail roles of these duplicates in the development of teleost, which is our ongoing project.

Acknowledgements

We thank Dario Moser for the sampling of stickleback and feeding help from Anja Frey, sampling of zebrafish from Dr. Heinz-Georg Belting of Prof. Markus Affolter group in Biozentrum, University of Basel. Many thanks for the help from Lukas Zimmerman for the IT support and Peter Fields for the help of syntenic analysis. And many thanks for the dissection help from Fabrizia Ronco, Adrian Indermaur and Attila Rüegg. Thank Liang Li for making expression profile figures. This project was supported by grant from the Swiss National Science Foundation (SNSF) and the European Research Council (ERC).

References

Ayrault Jarrier M, Levy G, Polonovski J. 1963. ETUDE DES ALPHA-LIPOPROT'EINES S'ERIQUES HUMAINES PAR. Bull. Soc. Chim. Biol. (Paris). 45:703–713.

Batzer MA, Deininger PL. 2002. Alu repeats and human genomic diversity. Nat. Rev. Genet. 3:370–379.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, Simakov O, Ng AY, Lim ZW, Bezault E, et al. 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 513:375–381.

Brudno M, Do CB, Cooper GM, Kim MF, Davydov E, Green ED, Sidow A, Batzoglou S. 2003. LAGAN and Multi-LAGAN: efficient tools for large-scale multiple alignment of genomic DNA. Genome Res. 13:721–731.

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. Nature. 376:479–485.

Chen L, DeVries AL, Cheng C-HC. 1997. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. 94:3811–3816.

Chen W-J, Santini F, Carnevale G, Chen J-N, Liu S-H, LavouÃ© Sã©, Mayden RL. 2014. New insights on early evolution of spiny-rayed fishes (Teleostei: Acanthomorpha). Front. Mar. Sci. 1.

Darriba D, Taboada GL, Doallo R, Posada D. 2012. jModelTest 2: more models, new heuristics and

parallel computing. Nat. Methods. 9:772.

Deschamps J. 2016. Birth and upgrowth of the Hox topological domains during evolution. Nat. Genet. 48:227–228.

Dillon N, Trimborn T, Strouboulis J, Fraser P, Grosveld F. 1997. The effect of distance on long-range chromatin interactions. Mol. Cell. 1:131–139.

Drayna D, Fielding C, McLean J, Baer B, Castro G, Chen E, Comstock L, Henzel W, Kohr W, Rhee L. 1986. Cloning and expression of human apolipoprotein D cDNA. J. Biol. Chem. 261:16535–16539.

Duboule D. 2007. The rise and fall of Hox gene clusters. Development. 134:2549–2560.

Dulai KS, von Dornum M, Mollon JD, Hunt DM. 1999. The Evolution of Trichromatic Color Vision by Opsin Gene Duplication in New World and Old World Primates. Genome Res. 9:629–638.

Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2014. Ensembl 2014. Nucleic Acids Res. 42:D749–D755.

Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. 1999. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 151:1531–1545.

Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. 2004. VISTA: computational tools for comparative genomics. Nucleic Acids Res. 32:W273–W279.

Fryer G, Iles T. 1972. The cichlid fihses: their biology and evolution. edinburgh, uk.

Garcia-Fernàndez J. 2005. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. 6:881–892.

Gaunt SJ. 2015. The significance of Hox gene collinearity. Int. J. Dev. Biol. 59:159–170.

Ghimire-Rijal S, Lu X, Myles DA, Cuneo MJ. 2014. Duplication of genes in an ATP-binding cassette transport system increases dynamic range while maintaining ligand specificity. J. Biol. Chem. 289:30090–30100.

Glasauer SMK, Neuhauss SCF. 2014. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics. 289:1045–1060.

Green SA, Simoes-Costa M, Bronner ME. 2015. Evolution of vertebrates as viewed from the crest. Nature. 520:474–482.

Guindon S, Gascuel O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52:696–704.

Herpin A, Braasch I, Kraeussling M, Schmidt C, Thoma EC, Nakamura S, Tanaka M, Schartl M. 2010. Transcriptional rewiring of the sex determining dmrt1 gene duplicate by transposable elements. PLoS Genet. 6:e1000844.

Horth L. 2007. Sensory genes and mate choice: evidence that duplications, mutations, and adaptive

evolution alter variation in mating cue genes and their receptors. Genomics. 90:159–175.

Huang Y, Chain FJJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, Samonte IE, Stoll M, Bornberg-Bauer E, Reusch TBH, et al. 2016. Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. Mol. Ecol. 25:943–958.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 11:97–108.

Jaillon O, Aury J-M, Brunet F, Petit J-L, Stange-Thomann N, Mauceli E, Bouneau L, Fischer C, Ozouf-Costaz C, Bernot A, et al. 2004. Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype. Nature. 431:946–957. Janoušek V, Karn RC, Laukaitis CM. 2013. The role of retrotransposons in gene family expansions: insights from the mouse Abp gene family. BMC Evol. Biol. 13:107.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, Ahsan B, Yamada T, Nagayasu Y, Doi K, Kasai Y, et al. 2007. The medaka draft genome and insights into vertebrate genome evolution. Nature. 447:714–719.

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, et al. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647–1649.

Kmita M, Fraudeau N, Hérault Y, Duboule D. 2002. Serial deletions and duplications suggest a mechanism for the collinearity of Hoxd genes in limbs. Nature. 420:145–150.

Kondrashov FA, Kondrashov AS. 2006. Role of selection in fixation of gene duplications. J. Theor. Biol. 239:141–151.

Kondrashov FA. 2012. Gene duplication as a mechanism of genomic adaptation to a changing environment. Proc. Biol. Sci. 279:5048–5057.

Kosakovsky Pond LS, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006. GARD: a genetic algorithm for recombination detection. Bioinformatics. 22:3096–3098.

Lalitha S. 2000. Primer Premier 5. Biotech Software&Internet Rep. 1.

Lapierre M, Blin C, Lambert A, Achaz G, Rocha EPC. 2016. The impact of selection, gene conversion, and biased sampling on the assessment of microbial demography. Mol. Biol. Evol.msw048.

Levings PP, Bungert J. 2002. The human β-globin locus control region. Eur. J. Biochem. 269:1589–1599.

Lin Z, Li W-H. 2011. Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. Mol. Biol. Evol. 28:131–142.

Mayor C, Brudno M, Schwartz JR, Poliakov A, Rubin EM, Frazer KA, Pachter LS, Dubchak I.

2000. VISTA : visualizing global DNA sequence alignments of arbitrary length. Bioinformatics. 16:1046–1047.

Montavon T, Duboule D. 2013. Chromatin organization and global regulation of Hox gene clusters. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 368:20120367.

Muschick M, Barluenga M, Salzburger W, Meyer A. 2011. Adaptive phenotypic plasticity in the Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol. 11:116.

Muschick M, Indermaur A, Salzburger W. 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22:2362–2368.

Narendra V, Rocha PP, An D, Raviram R, Skok JA, Mazzoni EO, Reinberg D. 2015. CTCF establishes discrete functional chromatin domains at the Hox clusters during differentiation. Science. 347:1017–1021.

Ohno S. 1970. Evolution by gene duplication. Springer Berlin Heidelb.

Ovcharenko I, Loots GG, Nobrega MA, Hardison RC, Miller W, Stubbs L. 2005. Evolution and functional classification of vertebrate gene deserts. Genome Res.15:137–145.

Postlethwait J, Amores A, Cresko W, Singer A, Yan Y-L. 2004. Subfunction partitioning, the teleost radiation and the annotation of the human genome. Trends Genet. 20:481–490.

Proudfoot NJ, Shander MH, Manley JL, Gefter ML, Maniatis T. 1980. Structure and in vitro transcription of human globin genes. Science. 209:1329–1336.

Provost PR, Villeneuve L, Weech PK, Milne RW, Marcel YL, Rassart E. 1991. Localization of the major sites of rabbit apolipoprotein D gene transcription by in situ hybridization. J. Lipid Res. 32:1959–1970.

Rassart E, Bedirian A, Do Carmo S, Guinard O, Sirois J, Terrisse L, Milne R. 2000. Apolipoprotein D. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 1482:185–198.

Rastogi S, Liberles DA. 2005. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. 5:28.

Reams AB, Kofoid E, Duleba N, Roth JR. 2014. Recombination and annealing pathways compete for substrates in making rrn duplications in Salmonella enterica. Genetics. 196:119–135.

Robinson-Rechavi M, Marchand O, Escriva H, Laudet V. 2001. An ancestral whole-genome duplication may not have been responsible for the abundance of duplicated fish genes. Curr. Biol. 11:R458–R459.

Sasaki Y, Negishi H, Koyama R, Anbo N, Ohori K, Idogawa M, Mita H, Toyota M, Imai K, Shinomura Y, et al. 2009. p53 family members regulate the expression of the apolipoprotein D gene. J. Biol. Chem. 284:872–883.

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff J-N,

Lesch K-P, et al. 2013. The genome of the platyfish, Xiphophorus maculatus, provides insights into evolutionary adaptation and several complex traits. Nat. Genet. 45:567–572.

Schranz ME, Mohammadin S, Edger PP. 2012. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. 15:147–153.

Spitz F, Gonzalez F, Duboule D. 2003. A Global Control Region Defines a Chromosomal Regulatory Landscape Containing the HoxD Cluster. Cell. 113:405–417.

Sproul D, Gilbert N, Bickmore WA. 2005. The role of chromatin structure in regulating the expression of clustered genes. Nat. Rev. Genet. 6:775–781.

Swofford D. 2002. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Assoc. Sunderland, Massachusetts.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. 24:1596–1599.

Tang Y-C, Amon A. 2013. Gene copy-number alterations: a cost-benefit analysis. Cell. 152:394–405.

Tanimoto K, Liu Q, Bungert J, Engel JD. 1999. Effects of altered gene order or orientation of the locus control region on human beta-globin gene expression in mice. Nature. 398:344–348.

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.

Theis A, Salzburger W, Egger B. 2012. The function of anal fin egg-spots in the cichlid fish Astatotilapia burtoni. PLoS One. 7:e29878.

Venkatesh B. 2003. Evolution and diversity of fish genomes. Curr. Opin. Genet. Dev. 13:588–592.

Weech P, Provost P, Tremblay N, Camato R, Milne R, Marcel Y, Rassart E. 1991. Apolipoprotein D—An atypical apolipoprotein. Prog. Lipid Res. 30:259–266.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555–556.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Supplementary Table 1 Primers for sequencing ApoD duplicates in cichlid fish and gene expression profile comparison in cichlid fish, stickleback and zebrafish.

species	gene	database	5’ primer	3’ primer	Tm
P. philander	copyA2a	pcr	CAGCCAAGGTGCTAAGTGTCATCCATAAC	TGTAAGACAAGCCAGTGTTTAATCAGTTCATC	65-55
A. burtoni		NCBI	/	/	/
N. brichardi		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
P. nyererei		NCBI	/	/	/
B. microlepis	copyA2b	pcr	CAGCCAAGGTGCTAAGTGTCATCCATAAC	TGTAAGACAAGCCAGTGTTTAATCAGTTCATC	65-55
P. nyererei		NCBI	/	/	/
A. burtoni		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
N. brichardi		NCBI	/	/	/
C. leptosoma		pcr	CAGCCAAGGTGCTAAGTGTCATCCATAAC	TGTAAGACAAGCCAGTGTTTAATCAGTTCATC	65-55
N. brichardi	copyA2c	NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
P. nyererei		NCBI	/	/	/
A. burtoni		NCBI	/	/	/
P. nyererei	copyA2	NCBI	/	/	/
A. burtoni		NCBI	/	/	/
N. brichardi		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
P. nyererei	copyB2b	NCBI	/	/	/
A. burtoni		NCBI	/	/	/

N. brichardi		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
P. nyererei		NCBI	/	/	/
A. burtoni		NCBI	/	/	/
L. careulus	copyB2a	pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
N. brichardi		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
P. philander		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
B. graueri		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
B. microlepis		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
C. leptosoma		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
A. burtoni		NCBI	/	/	
H. cf. stappersii	copyA1	pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
N. brichardi		NCBI	/	/	/
P. nyererei		NCBI	/	/	/
O. niloticus		Ensembl	/	/	/
C. frontosa		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
P. microlepis		pcr	CACCCTGTTGGCATCAAAGTTGGC	TGTCCCGGTGGTTCAGATATGTGC	65-55
O. niloticus		Ensembl	/	/	/
A. burtoni	copyB1	NCBI	/	/	/
N. brichardi		NCBI	/	/	/
P. nyererei		NCBI	/	/	/

Supplementary Table 2 Likelihood ratio test (LTR) statistics of branch-site model comparisons

foreground branch	branch_site model			sites under positve selection from Bayes Empirical Bayes (BEB) analysis
	modelA	modelB	p/2	
cichlid fish				
A2bA2c	-5672.9	-5673.69	0.1	
A2bA2c_branch	**-5676.08**	**-5679.02**	**0.01**	38N, 41L, 70K, 74I, 94E, **101V**, 103P, 116S, 117Y, 119T, 147L, 148R, 168T
B2aB2b	-5675.74	-5675.74	0.5	
B2aB2b_branch	**-5675.87**	**-5677.46**	**0.03**	11L, 24R, 63E, 71D, 73T, 121Y, 139S, 159A
B1B2aB2b	-5675.59	-5675.59	0.5	
B1B2aB2b_branch	**-5676.21**	**-5679.18**	**<0.01**	6S, 16P, 19S, 27A, 39L, 69R, 101V, 106Q, 117Y, 118F, 172A, 173K
A1B2aB2b	-5678.5	-5678.5	0.5	
A1B2aB2b_branch	-5676.35	-5676.43	0.34	
A1A2A2aA2bA2c	-5664.24	-5665.08	0.09	
A1A2A2aA2bA2c1_branch	-5678.37	-5678.61	0.25	
A1	-5678.1	-5678.31	0.26	
A1_branch	-5677.44	-5678.07	0.13	
A2A2aA2bA2c	**-5665.01**	**-5667.1**	**0.02**	17T, 38N, 64A, 82S, 86-, 90G, 92R, 96I, 114S, 116S, 128E, 135T, 136I, 148R, 153Y. 166P, 170H, 178D, 205G, 206N
A2A2aA2bA2c_branch	**-5677**	**-5679.34**	**0.015**	90G, 94E, 128E
B2b	-5679.25	-5679.25	0.5	
B2b_branch	-5678.7	-5678.75	0.38	
B2a	-5678.8	-5678.8	0.5	
B2a1_branch	-5679.25	-5679.25	0.5	
A2a	-5677.97	-5679.31	0.051	
A2a_branch	**-5676.6**	**-5679.35**	**0.01**	25F, 90G, 94E, **128E**
A2b	**-5675.01**	**-5677.58**	**0.01**	87-, 92R, 129T, 135T, 138Y, 148R, 153Y, 158L, 166P, 205G, 206N
A2b_branch	-5679.37	-5679.37	0.5	
A2c	-5676.29	-5676.37	0.35	
A2c_branch	-5678.45	-5678.85	0.19	
A2	-5678.47	-5679.07	0.14	
A2_branch	-5677.33	-5678.58	0.057	
B1	-5678.91	-5679.08	0.275	
B1_branch	-5679.11	-5679.29	0.27	

A2A2bA2c	-5671.38	-5672.33	0.08	
A2A2bA2c_branch	-5679	-5679.31	0.22	
A2aA2bA2c	**-5666.96**	**-5670.01**	**<0.01**	32N, 38N, 57A, 58R, 63E, 82S, 86-, 90G, 92R, 96I, 114S, 116S, 122S, 128E, 130D, **135T**, 136I, 146F, 148R, 153Y, 166P, 178D, 182D, 205G, 206N
A2aA2bA2c_branch	**-5673.31**	**-5678.82**	**<0.01**	25F, 41L, 90G, 94E, 101V, 103P, 114S, 119T, **128E**

stickleback

stiB2b	-8549.43	-8550.43	0.08	
cichlidB2b	-8549.01	-8549.25	0.25	
stiB2a2	**-8543.73**	**-8549.07**	**<0.01**	**40T, 41S, 42R**, 43F, 57L, 72A, 91T, 114G, 128G, 136P, 161L, 183V
stiB2a2B2a1	**-8540.58**	**-8544.58**	**<0.01**	27Y, 40T, 41S, 42R, 43F, 66E, 71P, 72A, 73S, **89D**, 91T, 93R, 98Q, 110E, 134F, 136P, 157T, 161L, 164Y, **165H**, 166V, 168Y, 169A, 183V, 184H, 201T, 208-, 212N, 214I
stiB2a2B2a1_branch	-8548.24	-8549.42	0.06	
stiB2a3	**-8547.69**	**-8550.41**	**<0.01**	69K, 97S, 113E, 207-
stiB2a1B2a2B2a3	**-8538.54**	**-8543.13**	**<0.01**	40T, 41S, 42R, 72A, 73S, 76R, **83N**, 89D, 110E, 123E, 133Y, **134F**, 160F, 161L, 165H, **166V, 168Y**, 169A, 175S, 183V, 201T
stiB2a1B2a2B2a3_branch	-8549.58	-8550.43	0.095	
cichlidB2aB2b	-8546.36	-8546.36	0.5	
cichlidB2aB2b_branch	-8550.48	-8550.48	0.5	
cichlidB2aB2bstiB2bB2a1B2a2B2a3	-8517.49	-8517.8	0.215	
cichlidB2aB2bstiB2bB2a1B2a2B2a3branch	**-8544.16**	**-8548.21**	**<0.01**	29L, 42R, 72A, 81E, 89D, 91T, 97S, 113E, 137Y, **155S**, 160F, 175S
stiB1	-8550.48	-8550.48	0.5	
newall	-8550.05	-8550.05	0.5	
newallbranch	-8550.42	-8550.48	0.365	
stiA2I	-8549.15	-8550.48	0.052	
stiA2II	**-8547.29**	**-8550.29**	**<0.01**	27Y, 28L, 74F, 96N, 98Q, 105-, 125A, 126K, 141W, 144E, 150Y, 180P, 184H
stiA2IA2II	-8550.48	-8550.48	0.5	
stiA2IA2II_branch	-8550.48	-8550.48	0.5	
cichlidA2A2aA2bA2cstiA2IA2II	**-8531.57**	**-8537.21**	**<0.01**	**100S**, 103-, 108R, 112I, 119P, 138S, 162R, **180P**, 184H, **211G**
cichlidA2A2aA2bA2cstiA2IA2II_branch	-8550.48	-8550.48	0.5	
stiA1	-8550.46	-8550.46	0.5	

cichlidA1stiA1	-8548.7	-8548.7	0.5	
cichlidA1stiA1_branch	-8549.7	-8549.7	0.5	
cichlidA2B2aB2bstiB2bB2a1B2a2B2a3	-8529.07	-8529.07	0.5	
cichlidA2B2aB2bstiB2bB2a1B2a2B2a3_branch	-8549.07	-8549.07	0.5	
stiB2a1	**-8544.24**	**-8546.47**	**<0.01**	69K, 71P, 77G, 89D, 92I, 134F, 157T, 161L, 164Y, 165H, 168Y, 169A, 175S, 201T, 203T

zebrafish

zebB2	**-7059.9**	**-7062.8**	**0.01**	23R, 49E, 57R, 61I, 76L, 159P, 164H, 167K, 177I, 179K
zebB2cichlidB2aB2b	-7052.29	-7052.29	0.5	
zebB2cichlidB2aB2b_branch	-7061.01	-7062.12	0.07	
cichlidB1B2aB2b	-7048.66	-7048.66	0.5	
cichlidB1B2aB2b_branch	**-7057.68**	**-7062.12**	**<0..01**	3-, 5S, 15P, 21T, **26A**, 34S, 39Q, 68R, 82-, 89R, 98V, 103Q, 114Y, 115F, **166A**, 167K, 178S
zebA2	-7062.97	-7062.97	0.5	
zebA2cichlidA2A2aA2bA2c	-7043.15	-7043.2	0.375	
zebA2cichlidA2A2aA2bA2c_branch	-7061.96	-7062.93	0.08	
zebA1	-7062.97	-7062.97	0.5	
zebA1cichlidA1	-7062.97	-7062.97	0.5	
zebA1cichlidA1_branch	-7059.92	-7060.64	0.115	

medaka

medaB2b	-8394.82	-8394.82	0.5	
medaB2bcichB2b	-8392.38	-8392.88	0.16	
medaB2bcichB2bbranch	-8391.97	-8393.19	0.059	
medaB2a	-8392.05	-8393.92	0.025	105R, 112N, 117K, 224D, 232Y
medaB2acichB2a	-8393.29	-8393.29	0.5	
medaB2acichB2abranch	-8392.55	-8394.02	0.043	52M, 142E, 144T, 189L, 196Y, 197A, 235C
medaB2aB2bcichB2aB2b	-8382.91	-8382.91	0.5	
medaB2aB2bcichB2aB2bbranch	-8390.76	-8392.78	0.02	58L, 71R, 118D, 120T, 166Y, 184S, 201A, 229T
medaB2aB2bcichB2aB2bB1	-8380.5	-8380.5	0.5	
medaB2aB2bcichB2aB2bB1branch	-8389.48	-8394.46	0.001	53S, **74A**, 116R, 137R, **146V**, 151Q, 162Y, 163F, 209E, 214A, 215K
medaA2m1A2m2A2m3	-8386.53	-8387.29	0.11	
medaA2m1A2m2	-8383.73	-8386.78	0.005	**76P**, 85N, 90L, 95E, 110E, 113Y, 114S, **124L**, 130-, 140L, 144T, 145A, 147V, 150P, 157G, 166Y, **169Y**, 214A, **238-**, 239I, 241-
medaA2m1A2m2branch	-8392.83	-8394.54	0.03	95E, 105R, 214A, 231T

medaA2m1	-8389.81	-8390.46	0.125	
medaA2m2	-8389.39	-8392.39	0.005	71R, 76P, 90L, **113Y**, 119G, 124L, 144T, 145A, 162Y, 238-, 239I
medaA2m1A2m2A2m3branch	-8393.9	-8394.47	0.14	
medaA2m3	-8392.85	-8394.74	0.025	187D, 205S, 209E
medaA2m1A2m2cichlidA2A2aA2ba2c	-8380.34	-8380.34	0.5	
medaA2m1A2m2cichlidA2A2aA2ba2c	-8388.75	-8394.82	<0.01	66S, 104A, 191I, 205S, 226S
medaA2m1A2m2A2m3cichlidA2A2aA2ba2c	-8376.37	-8376.37	0.5	
medaA2m1A2m2A2m3cichlidA2A2aA2ba2cbranch	-8393.83	-8394.82	0.08	
medaA1	-8394.82	-8394.82	0.5	
medaA1cichA1	-8394.82	-8394.82	0.5	
medaA1cichA1branch	-8391.87	-8393.11	0.06	
medaA1A2m1A2m2A2m3cichA2aA2bA2cA2A1	-8380.11	-8380.11	0.5	
medaA1A2m1A2m2A2m3cichA2aA2bA2cA2A1branch	-8394.2		0.3	
cichA2bA2c	-8382.5	-8384	0.04	
cichA2bA2cbranch	-8391.36	-8394.71	<0.01	121I, **146V**, 148P, 164T, 189L, 190R, 210T
cichA2aA2bA2c	-8376.93	-8381.8	<0.01	**129S, 131-**, 132-, 135G, **137R**, 141I, 159S, 161S, 167S, 173E, **180T**, 181I, 195Y, 208P, 220D, 236G, **237N**
cichA2aA2bA2cbranch	-8394.82	-8394.82	0.5	
cichlA2A2aA2bA2c	-8382.41	-8384.53	0.02	111A, 129S, 131-, 132-, 135G, 137R, 141I, 159S, 1661S, 173E**, 180T**, 195Y, 208P, 212H, 220D, 236G, 237N, 238-
cichlA2A2aA2bA2cbranch	-8394.82	-8394.82	0.5	

Supplementary Table 3 Qpcr primers for testing gene expression of ApoD duplicates in teleosts

primer	sequence 5'-3'
rpl7_qpcr_f	GGAGAAGTCCCTCGGCAAAT
rpl7_qpcr_r	GGCGGGCTTGAAGTTCTTTC
A2a_qpcr_f	GCTCAGCACTGAAGAAGCAG
A2a_qpcr_r	CATGGTTCCTTTCTGTTCACC
A2b_qpcr_f	CGATCTCCTGAAGGAACCAT
A2b_qcpr_r	TCTCCCAGGTACGGTTCAAG
A2c_qpcr_f	ACCTTCACCTGACATTCGC
A2c_qpcr_r	GACCAGCAGGAGAAGAAGGT
A2_qpcr_f	TTCTCCTTCTGATGCTCCCT
A2_qpcr_r	CCCAGGTACTGCTGAAGGTT
B2b_qpcr_f	CCAGGTGATTTCCCTCACTT
B2b_qpcr_r	CCAAGATACCTTGCAGCATC
B2a_qpcr_f	GATGCCAACAGGTACATTGG
B2a_qpcr_r	GAAGCTCGGTGTTCAGGAC
A1_qpcr_f	TCCTACGTCCTTCCCTACTCC
A1_qpcr_r	GAAGAGCCTGAGGATGTCAGT
ubc_qsti_f	AGACGGGCATAGCACTTGC
ubc_qsti_r	CAGGACAAGGAAGGCATCC
A2I_qsti_f	TGGTGTCCACACAGACCTTT
A2I_qsti_r	GCCCAGATACTGTGTGAGGTT
A1_qsti_f	AGCTGAAGAAGATCGAAGGG
A1_qsti_r	GGAGTAGGGCAGAACGTAGG
B1_qsti_f	ACCAAGCGACCTACAGTCCT
B1_qsti_r	CTTCCCATTGGACAGAAGC

B2a1_qsti_f	CAGCTGTTCAGGCCAACTT
B2a1_qsti_r	CCTCTTGATCCCGTACCACT
B2a2_qsti_f	GTGGTACGGGATCAAGAAGC
B2a2_qsti_r	ATCTTCAAGCAGCCCACTGT
B2a3_qsti_f	TTTGACGCTTCCAAGTACCTC
B2a3_qsti_r	AGGTTGTAGAGGGCAGTGCT
B2b_qsti_f	TGGGACCATAGACTCCATCA
B2b_qsti_r	GGAGCGTTCTCAAAGAAGGA
elfa_qzebra_f	CTTCTCAGGCTGACTGTGC
elfa_qzebra_r	CCGCTAGCATTACCCTCC
A1_qzebra_r	AGGAATAGACCAGAGCCGAA
A1_qzebra_f	ACGGCTGTGGTGGAAGATA
A2_qzebra_f	CGGCAATCATCCAGGATATAA
A2_qzebra_r	GGGCATAAGGTGTGAAATAGG
B2_qzebra_f	AGTCAGCGCTCAGTCCATC
B2_qzebra_r	CTGGAATGGTGAAGGGAACT

Supplementary Table 4 The number of raw reads and reads after trimming of Ion-Torrent next generation sequencing library construction for 16 cichlid fish species.

species	first round		second round		third round	
	raw data	after trimming_filter	raw data	after trimming_filter	raw data	after trimming_filter
C	117816	107418	163569	150973	143288	138932
LE	104282	94696	143769	134441	123027	120001
LL	115607	106665	151523	142418	110868	108295
HC	114346	103187	142447	133648	157639	153428
OV	124957	113405	190089	176161	112440	109299
GW	150280	135700	144392	135099	104437	100574
PM	145680	129459	7598	7062	104229	101625
GL	143184	128688	142061	131441	113157	109900
Tn	131065	119994	143904	136165	112000	109165
EC	109280	97951	141706	132584	106873	103790
CL	104421	95775	146155	134935	140690	136273
PP	115199	104486	150146	141761	136342	133253
Cy	137608	124022	18441	17053	140830	137431
L	113560	103831	155085	142175	105889	103161
XS	133516	119129	157793	146987	36835	35996
BM	116517	106278	126137	119742	138956	134685

Note: C: *C. macrops*, LE: *L. elongates*; LL: *L. labiatus*; HC: *H. cf. stappersii*; OV: *O. ventralis*; GW: *Greenwoodochromis*; PM: *P. microlepis*; GL: *G. lemairi*; Tn: *T. nigrifrons*; EC: *E. cyanostictus*; CL: *C. leptosome*; PP: *P. philander*; Cy: *C. frontosa*; L: *L. careulus*; XS: *X. spiloptera*; BM: *B. microlepis*.

Supplementary Table 5 Repeat elements in the two clusters in teleosts

Table 5A Repeat elements in the two clusters in tilapia.

cluster 1	5'_up10kb	intergenic 1	intergenic 2	3'down10kb
	LINE	0	repeat	satellite
	SINE		hAT/Charlie_transposon	hAT_transposon
	LINE			RC/Helitron_transposon
				LINE
				DNA/TcMar-Tcl_transpson

cluster 2	5'_up10kb	interexons	intergenic 1	intergenic 2	intergenic 3	3'down10kb
	Tcmar_tcl_transposon	Tcmar_tcl_transposon	0	repeat	0	SINE
	Tcmar_tcl_transposon	Tcmar_tcl_transposon		Gypsy_LTR		SINE
	SINE	repeat		LTR		LTR
				LINE		simple repeat
				LINE		simple repeat
				repeat		
				LINE		
				line		

Table 5B Repeat elements in the two clusters in *A. burtoni.*

cluster1	5'_up10kb	intergenic 1	intergenic 2	3'_down10kb
	LTR	0	hAT_charlie	repeat
	SINE		SINE	Helitron_transposon
	repeat			
	Helitron_transposon			

cluster 2	5'_up10kb	intergenic 1	inter exons	intergenic 2	intergenic 3	3'_down10kb
	LINE	hat_ac_transposon	hAT_AC_transposon	0	Tcmar_tcl_transposon	Kolobok_t2_transposon
	Tcmar_tcl_transposon	LINE	hAT_AC_transposon		hAT_charlie_transposon	LINE

	Tcmar_tcl_transposon	LINE						Helitron_transposon
	Tcmar_tcl_transposon	repeat						tcmar_tcl_transposon
	Tcmar_tcl_transposon	hAT_transposon						LINE
		hAT_transposon						tcmar_tcl_transposon
		helitron_transposon						

Table 5C Repeat elements in the two clusters in medaka.

cluster 1	5'_up10kb	intergenic1	intergenic2	3'_down10kb
	LINE	0	LTR	LINE
	LINE		SINE	repeat
	SINE		Gypsy	transposon
	hAT_transposon			PIF_Harbinger_transposon
				PIF_Harbinger_transposon

cluster 2	5'_up10kb	inergenic1	intergenic2	intergenic3	3'_down10kb
	kolobok_t2_transposon	LINE	0	maverick_transposon	0
	cmc_enspm_transposon				
	hAT_transposon				
	Cmc_enspm_transposon				

Table 5D Repeat elements in the two clusters in stickleback.

cluster 1	5'_up10kb	intergenic1	intergenic2	interexons	intergenic3	intergenic4	down10kb	down_10kb
	LTR	0	0	LINE	0	LTR	0	LINE
	Cmc_enspm					Maverick		hAT
								cmc_enspm
								hAT
								LINE
								cmc_enspm
								hAT_AC

				hAT_AC
cluster 2	5'_up10kb	intergeneci1	3'_down10kb	
	simple repeat	0	LTR	
	simple repeat		DANN/hAT	
	simple repeat			
	simple repeat			
	simple repeat			
	simple repeat			
	simple repeat			

Table 5E Repeat elements in the two clusters in fugu.

cluster 1	0			
cluster 2	5'_up10kb	intergenic1	intergenic2	3'_down10kb
	0	2	2	0
		LTR	LINE	
		SINE	repeat	

Table 5F Repeat elements in the two clusters in zebrafish

cluster 1	5'_up10kb	intergenic1	3'_down10kb
	DANN	satellite	repeat
	TcMAR	hAT	satellite
	hAT_charlie	satellite	repeat
	repeat	hAT	transposon
	Helitron	hAT	transposon
	repeat	hAT	transposon
	hAT	hAT	SINE
	LINE	hAT	transposon

	Kolobok_T2		transposon	
	hAT		transposon	
	repeat			
	hAT			
	LINE			
	repeat			
	hAT			

cluster 2	5'_up10kb	interexons	3'_down10kb
	hAT	transposon	SINE
	repeat		repeat
	transposon		transposon
	LTR_GYPSY		transposon
			repeat
			hAT
			satellite
			Helitron
			satellite

Table 5G Repeat elements in the two cluster in gar.

gar	5'_up10kb	interexons	intergenic	3'_down10kb
	SINE	satellite	SINE	SINE
	rRNA	SINE	SINE	
			SINE	

Table 5H Repeat elements in the two clusters in coelacanth

coelacanth	5'_up10kb	interexons	interexons	interexons	3'_down10kb
	SINE	hAT_AC		SINE	SINE
	SINE	SINE			

Table 5I Repeat elements in the two clusters in human

human	up10kb	interexons	interexons	interexons	interexons	down10kb
	SINE	SINE	LTR	LINE	LINE	SINE
	SINE	SINE	LTR	SINE	SINE	LINE
	SINE	SINE	SINE	Tcmar_tcl	SINE	SINE
	LINE	SINE	LINE	SINE		LINE
	SINE		SINE			LTR
	LINE		LINE			SINE
	SINE		SINE			tRNA
	LINE		LINE			LINE
	SINE		SINE			SINE
	LINE		LINE			LINE
	LTR		LTR			LTR
	SINE		LINE			SINE
	SINE					LTR
	hAT					LINE
	SINE					SINE
	LINE					LINE
	LTR					SINE

Chapter 6

Discussion and further perspectives

Chapter 6
Discussion

The main goal of this doctoral thesis was to understand the emergence and evolution of two innovative pigmentation patterns in East African cichlid fishes, haplochromine eggspots and ectodine blotch. By making use of comparative transcriptomic analysis based on newly generated RNA sequencing data from the ectodine cichlid *Callochromis macrops* and available data of *A. burtoni* (Santos et al.), I found that 15.7 % (43/274) of the investigated genes showed similar expression patterns between haplochromine eggspots and the ectodine blotch, suggesting that they might share common gene network (**Chapter 2 and Chapter 3**). By making use of Ion Torrent next generation sequencing technology and further data analysis, we sequenced 40 ectodine blotch related candidate genes (16 genes also showed highly expression in eggspots, two genes showed opposite expression in eggspots) across the phylogenetic tree of East African cichlids, including whole gene regions and upstream regions. The analysis revealed that at least four (*fhl2b, 26203, pnp4a, 13024*) genes exhibited specific CNS in the upstream and intron region in haplochromine species with eggspots, as well as specific amino acid substitutions (genes *26203, crip1, trpm1a, 23711, 23698, zdhh*). The sequences of *C. macrops*, on the other hand, showed patterns similar to species without anal fin pigmentation patterns. This suggests that eggspots, but not the ectodine blotch, might have a much more independent gene network from the ancestral anal fin gene network, thus providing a clue of the mechanism about the different evolvability between lineages with eggspots and the ectodine blotch (discussed in **Chapter 4)**. In addition, a multi-ligand transporter related gene family, apolipoprotein D (ApoD), was found to have expanded in teleosts, and in cichlids in particular, via gene duplication, with the genes being located in two clusters in teleost genomes. One member of this gene family was found to be highly expressed in the ectodine blotch. Interestingly, although most genes showed conserved homologous expression pattern in distant related teleosts, duplicated genes with new functions evolved in a lineage specific manner, especially in cichlid fish, and were expressed in two novelties, lower pharyngeal jaw and anal fin pigmentation (**Chapter 5**). The main findings are discussed in below:

Do eggspots and the ectodine blotch share a common gene network?

By a thorough RNA-seq experimental designed for the ectodine blotch in *C. macrops* and re-analyzing the existing trancriptomic data for eggspots in *A. burtoni*, we identified 274 blotch related candidate genes and 812 eggspots candidate genes. 15.7% of (43/ 274) blotch-related candidate genes showed similar expression pattern in haplochromine eggspots (**Chapter 3**), suggesting that they might share at least parts of a common gene network, which is contrast with the

conclusion from Santos et al, (Santos et al.) (**Chapter 2**). The possible explanation of these differences could be that we used the whole gene exrepssion profile derived from comparative transcriptomic data of the ectodine blotch instead of using several individual candidate genes. Besides, compared to eggspots which have different numbers, the ectodine blotch is much easier to control for positional effect to get candidate genes. In addition, some shared genes showed different expression patterns. For example, *col8a1b* related to collagen and *steap4* related to immunity showed high expression levels in eggspots, but low ones in the ectodine blotch. Besides, 15 genes showed high expression levels in the ectodine blotch, but low ones in eggspots. What are the roles of these shared genes with differential expression patterns in these two different novelties will be interesting to study in the future. This exemplifies the importance of focusing on entire gene networks instead of single candidate genes.

Two hypothesis were proposed for the evolution of these common shared and unshared genes: 1) The gene network formed by the shared genes might mean that eggspots and the ectodine blotch are homologous at a deep level; and the unshared genes were co-opted independently in two lineages. If this gene network is responsible for the anal fin pigmentation patterns, then their common ancestor might have innovative anal fin pigmentation, which was lost secondarily in most species in ectodine lineage. 2) It also could be that the shared genes were independently co-opted into two different networks. The roles of the shared and unshared genes in these two anal fin pigmentation patterns could be that 1) the common shared genes were responsible for the basic pigmentation pattern formation, while the unshared genes were responsible for the differences between eggspots and the ectodine blotch. 2) The shared genes could also be responsible for the differences between eggspots and the ectodine blotch, since we did not test different developmental points here. 3) The unshared genes could also be responsible for the basic pattern formation, for example, if they are different effector genes but with same phenotypic results. The roles of these shared and unshared genes in the gene network rewiring between eggspots and the ectodine blotch needs further investigation.

To what extent a novelty is innovative?

To further examine how the gene network related to the innovative anal fin pigmentation pattern in cichlids is rewired, we characterized the entire gene regions of 40 candidate genes, including the upstream region of the ectodine blotch related candidate genes across the phylogeny and including species with and without the innovative pigmentation (**Chapter 3**). Our study proposed three hypothesis of the gene network relationships among eggspots, the ectodine blotch

and anal fin: 1) “eggspots>blotch=fin”; 2) “eggspots=blotch=fin”; 3) “eggspots=/>blotch>fin”, and our data mostly support the first hypothesis, which means that eggspots, but not the ectodine blotch, might have a much more independent gene network from the ancestral anal fin gene network, thus providing a clue of the mechanism about its higher evolvability. Evidence comes from the observation that at least four genes showed eggspots specific conservative non-coding sequences (CNS) patterns (*fhl2b* [see also (Santos et al. 2014), *13024, 26203, pnp4a*]. These CNS showed eggspots specific TF binding sites based on the prediction. In addition, several common TF binding sites are shared among these CNS, such as Glucocorticoid related TFs, thyroid hormone related TFs, homeobox, etc, which might belong to the core gene network of eggspots. Besides, many genes also showed eggspots specific amino acid substitutions within their coding regions (*26203, crip1, trpm1a, 23711, 23698, zdhhc14*). These genes could also play a role in gene network rewiring by affecting protein-protein interactions, or might be the effector genes, such as pigmentation related genes to affect pattern formation. Interestingly, although these candidate genes are the ectodine blotch related genes, little blotch specific patterns were found. Instead, sequences in *C. macrops* appeared much more similar to species without innovative anal fin pigmentation. These results suggested that - compared to the ectodine blotch- species with eggspots might have gained much more independence in their gene network compared to the ancestral fin gene network. Besides, eggspots could also have different roles as ectodine blotch, and selection might be stronger on it. In addition, the ectodine blotch and haplochromine blotch might be originated independently.

New genes with new functions——gene duplication in cluster for apolipoprotein D (ApoD) gene family

Classical molecular evo-devo originated from the discovery of highly conserved developmental Homeobox genes (Hox genes) in distantly related species (Carroll 1995). However, except for the mechanisms of “old genes playing new tricks” by co-option of pre-existing genes, it is largely unknown how new genes broke up conservative expression patterns thus contributing to phenotypic novelties. In **Chapter 4**, I focus on the consequences of gene duplication and their contribution to novelties by studying the expansion of a multi-ligand transporter related gene family, ApoD gene family, via duplication in clusters in teleosts. I found that, across teleosts, orthologous genes in the same physical order along their respective cluster exhibited similar expression pattern. Combined with the results of positive selection detection in coding regions, several models (eg. neofunctionalization, subfunctionalization, dosage effects) were proposed to explain the evolution of these duplicates, especially related to two putative innovations (lower pharyngeal jaw and anal fin pigmentation) in cichlid fish. The mechanisms behind the maintenance of the duplicates in

clusters were discussed from a functional and structural point of view, with a special focus on the role of repeat elements. Taken together, the ApoD gene family provides an ideal model to study the evolution of new gene functions, gene duplication, gene cluster maintenance and their relationship with teleost radiation.

Further perspectives——evo-devo research enters postgenomic era

Next generation sequencing technology provides an unprecedented opportunity to study evo-devo and make it possible to focus on the gene network level instead of individual genes (Wagner 2012; Roux et al. 2015), even without further functional experiments (Lynch et al. 2011). For example, a previous study has shown that a major candidate identified by QTL mapping is indeed causing cavefish albinism (Protas et al. 2006). In addition, with CHIPseq techonlogy (Chromatin Immunoprecipitation combining with high through output), researchers can identify genome-wide TF binding events or epigenetic marks indicative of enhancers or other types of gene regulatory elements easily (Lynch et al. 2011), which will be the future plan of our study. Recent rapid technology development of functional assays, such as regulatory interspaced short palindromic repeats (CRISPR/CAS9) nuclease system (Cong et al. 2013; Mali et al. 2013), or Tol2-mediated transgenic make functional experiment available in non-model species, including cichlid fish (Juntti et al. 2013; Li et al. 2014). This will be definitely useful to test the anal fin pigmentation related candidate genes which were identified in this thesis. The role of the expansion of ApoD gene family via duplication in cluster in the development of teleosts is our ongoing project. All these technologies facilitate that evo-devo research enters the postgenomic era (Lesoway 2016).

References

Carroll SB. 1995. Homeotic genes and the evolution of arthropods and chordates. Nature. 376:479–485.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. Science. 339:819–823.

Juntti SA, Hu CK, Fernald RD. 2013. Tol2-mediated generation of a transgenic haplochromine cichlid, Astatotilapia burtoni. PLoS One. 8:e77647.

Lesoway MP. 2016. The future of Evo-Devo: the inaugural meeting of the Pan American Society for evolutionary developmental biology. Evol. Dev. 18:71–77.

Li M, Yang H, Zhao J, Fang L, Shi H, Li M, Sun Y, Zhang X, Jiang D, Zhou L, et al. 2014. Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. Genetics. 197:591–599.

Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene

regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 43:1154–1159.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013. RNA-guided human genome engineering via Cas9. Science. 339:823–826.

Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. Nat. Genet. 38:107–111.

Roux J, Rosikiewicz M, Robinson-Rechavi M. 2015. What to compare and how: Comparative transcriptomics for Evo-Devo. J. Exp. Zool. B. Mol. Dev. Evol. 324:372–382.

Santos ME, Baldo L, Gu L, Boileau N, Musilova Z, Salzburger W. The molecular basis of independently evolved anal fin pigment patterns in cichlid fish (submitted).

Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, Belting H-G, Affolter M, Salzburger W. 2014. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:5149.

Wagner GP. 2012. Next Gen Devo-Evo. J. Exp. Zool. B. Mol. Dev. Evol. 318:519–520.

Acknowledgements

First of all, I would like to express my deepest gratitude to my cool and nice supervisor, Prof. Walter Salzburger, who gave me the precious opportunity to come to Switzerland to pursue my PhD and to experience the diversified cultures in Europe, which could be viewed as my second life. Many thanks to him for giving me the freedom to choose the topic I like. His free but professional style makes me feel relax and happy in the scientific area. So many thanks for his full support, guidance, patience and respect throughout my whole PhD study in Switzerland.

And many thanks to Dr. M. Santos Emilia who gave me the first guidance to the topic of anal fin pigmentation. Her enthusiasm and passion on science made a deep impression on me, which inspired me througout my PhD study.

And many thanks to Philippe Demougin, whose assistance with illumina sequencing library construction makes a beautiful start of my PhD project.

I would also like to thank Nicolas Boileau for his professional support in the lab, which is invaluable throughout my PhD project.

And special thank to Attila Rüegg for both his professional support in the aquarium room and friendship.

Many thanks to Lukas Zimmerman for the IT support, who always help me communicate with computer with their own “strange“ language.

And I also would like to express my thanks to Prof. Dieter Ebert, whose questions always make me think deeply about my project and get new ideas.

Many thanks to the whole Salzburger group for the support throughout my PhD. Especially to Marco Colombo and my dear office mates for the fun time.

And I also would like to thank my master supervisor Prof. Zhengwang Zhang, who gave me the opportunities to broaden my eyes during my master study.

And many thanks to Prof. Yang Liu who inspired me to pursue my PhD here.

So many thanks to my dear Chinese friends Tingting Zhou, Jie Zhang, Yang Liu, Shengkai pan, Liang Li, De Chen, Li Jiang, Xiangjiang Zhan and Bo Xu for their professional support and long-distance friendships.

Many thanks to Yanyan Xu, Laura Liegaspi and the members of dancing club *Chinesischer Tanzverein Baselland* for the invaluable friendships in Switzerland.

And a special thank to Ms. Stoll Zuzanna who let me share a large, warm and nice room with her and let me spend Christmas and many festivals with her family in Switzerland, which make me feel like at home.

Finally, I would like to express my special thanks to my dear Mother, who always gives me the encouragement and unconditional support to my choice.

Personal information

Langyu Gu (谷浪屿)

Born in 1988, Jilin province, China
langyugu@gmail.com

Education

August 2012 to June 2016 PhD student in Evolutionary Biology, Zoological Institute, Basel University, Basel, Switzerland.
Supervised by Prof. Walter Salzburger.

September 2009 to July 2012 Master degree in Zoology, College of Life Sciences, Beijing Normal University, Beijing, China.
Supervised by Prof. Zhengwang Zhang.
Ranking 1st in major with Major GPA: 3.85. Awarded with distinction.

September 2005 to July 2009 Bachelor of Science in Biology, College of Biological Sciences, China Agricultural University, Beijing, China.
Curricular Overall GPA: 3.7.
Postgraduate candidates exempt from Admission exam, which is only granted to excellent student.

Publications

Langyu Gu and Walter Salzburger**.** Expansion via duplication of a multiple-ligand transporter related gene family in cluster in teleost fish. (in preperation). Molecular Biology and Evolution.

Langyu Gu, Natacha Espirito Santo, Zuzana Musilova, Nicolas Boileau, Walter Salzburger. Sequence evolution of the genes involved in the convergent evolution of innovative anal fin pigmentation patterns in cichlid fish. (in preperation) Molecular Biology and Evolution.

Langyu Gu, M. Emilia Santos, Walter Salzburger. The genetic basis of convergent evolution of two innovative anal fin pigmentation patterns in cihclid fish---- haplochromine eggspots and ectodine blotches. (in preperation) BMC Genomics.

Santos M. Emilia, Baldo Laura, **Gu Langyu**, Boileau Nicolas, Musilova Zuzana and Salzburger Walter. 2016. Comparative transcriptomics of anal fin pigmentation patterns in cichlid fishes. BMC Genomics, 17: 712.

Langyu Gu*, Yang Liu*, Pinjia Que, Zhengwang Zhang. 2013. Quaternary climate and environmental changes have shaped genetic differentiation in a Chinese pheasant endemic to the eastern margin of the Qinghai-Tibetan Plateau. Molecular

Phylogenetics and Evolution, 67:129-139. (* These authors contributed equally to this work).

Langyu Gu, Yang Liu, Ning Wang, Zhengwang Zhang. 2012. A panel of polymorphic microsatellites in the Blue Eared Pheasant (*Crossoptilon auritum*) developed by cross-species amplification. Chinese Birds, 3: 103-107.

Ning Wang, Jiang Chang, **Langyu Gu**, Zhengwang Zhang. 2009. Polymorphic microsatellites in the Reeve's pheasant developed by cross-species amplification. European Journal of Wildlife Research, 55: 627-629.

Courses

2016 Eco-evo-devo Summer School in Oxford Brookes University, UK.

2015 Writing to be published for the Natural Sciences, University of Basel, Switzerland.

2014 Workshop Successful Fund acquisition for PhD, University of Basel, Switzerland.

2014 EMBO Practical Course on Computational Molecular Evolution. Greece.

2013 Genetics of Adaptation Workshop at the Evolutionary Biology Center, Uppsala University, Sweden.

2013 RNA-seq: from experiment planning to statistical data analysis. Swiss Institute of Bioinformatics, Switzerland.

2009 Training course on Molecular Genetics and Genomics at CAAS (Chinese Academy of Agriculture Sciences)-ILR (International Livestock Research Institute)-JLLFGR (Joint Laboratory on Livestock and Forage Genetic Resources), Beijing, China.

Meetings

2015 Oral communication: "Gene network re-wiring in the convergent evolution of a key innovation: egg dummy in cichlid fish". Congress of the European Society for Evolutionary Biology (ESEB 2015), Lausanne, Switzerland.

2015 Oral Communication: "Genetic basis oft he origination and evolution of a key innovation----egg dummy in cichlid fish". Research seminar, Nüsslein-Volhard Lab in Tübingen, German.

2015 Poster communication: "The role of internal repetition within protein on convergent evolution and adaptive radiation in cichlid fish". Annual Meeting of the Society for Molecular Biology and Evolution (SMBE 2015), Wien, Austria.

2013 Poster communication: "Comparative transcriptomics analysis to identify candidate genes for a pigment trait in cichlid fish". Congress of the European Society

for Evolutionary Biology (ESEB 2013), Lisbon, Portugal.

2013 Poster communication: Quaternary climate and environmental changes have shaped genetic differentiation in a Chinese pheasant endemic to the eastern margin of the Qinghai-Tibetan Plateau. Biology 13, Basel, Switzerland.

2011 Oral Communication: "Population genetic structure of the Blue Eared Pheasant, Crossoptilon auritum; implications for conservtaion". The 7th Chinese Ornithologists Conference. Lanzhou University, China.

Teaching experience and project supervision

2015 Supervisor of a scientific project: The role of gene duplication clusters in adaptive radiation in cichlid fishes. Students: Elia Heule, Jeanne-Elea Baumann. Level: 3rd year Bachelor students, University of Basel, Switzerland.

2015 Supervisor of a scientific project: A comparative analysis of egg-dummie related genes in East African cichlid fish. Student: Natacha Espirito Santo. Level: Master student, University of Basel, Switzerland.

2010 Teaching assistant in Ornithology in Xiaolongmen National Forest Park, Beijing, China.

2009 to 2010 Teaching assistant in Conservation Biology, Beijing Normal University, Beijing, China.

Grants

2012 to 2016 PhD Scholarship fund by Basel University, Basel, Switzerland.

2009 to 2012 Academic Scholarship fund by Beijing Normal University, China.

2007 to 2008 Xizhi scholarship, which is awarded only to outstanding students in academic, China Agricultural University, Beijing, China.

2007 to 2008 Academic Excellence Award, China Agricultural University, Beijing, China.

2006 to 2007 Xizhi Scholarship, which is awarded to outstanding students in academic, China Agricultural University, Beijing, China.

2006 to 2007 Academic Excellence Award, China Agricultural University, Beijing, China.

2005 to 2006 Academic Excellence Award, China Agricultural University, Beijing, China.

Languages

Chinese (Madrin):

Mother language

English:

Spoken: Fluent; Reading: Fluent; Writing: Fluent. Listenning: Fluent.

German:

Spoken: Basics. Reading: Basics. Writting: Basics. Listenning: Basics.

Miscellaneous

2012-2016 Member of dancing club *Chinesischer Tanzverein Baselland* for traditional Chinese dancing.

2012-2016 Experience different cultures in Europe.

2012 Part-time job in China Science Daily newspaper, working on topics about genome research, gene patent and food safety.

2011 Volunteer in Wildlife Conservation Society (WCS) to protect wildlife in China.

2011 Silver Medal in the 7th Chinese Ornithologists' Conference for the academic oral communication on protection of a Chinese endemic bird, Blue Eared Pheasant.

2010 Third Prize of College Students Social Practice on Biogas Service System Survey in China, sponsored by Royal Dutch Shell Company, Chinese Agriculture Ministry.

2009 Volunteer as a teacher in WEST SUNSHINE project to teach students in poor regions in western China.

2009 Radio-tracking in Dongzhai National Park in Henan province, China.

2005-2009 Student Union member in China Agricultural University.