The evolution of cichlid fish egg-spots is linked with a cis-regulatory change

M. Emilia Santos, Ingo Braasch, Nicolas Boileau, Britta S. Meyer, Loïc Sauteur, Astrid Böhne, Heinz-Georg Belting, Markus Affolter & Walter Salzburger

The origin of novel phenotypic characters is a key component in organismal diversification; yet, the mechanisms underlying the emergence of such evolutionary novelties are largely unknown. Here we examine the origin of egg-spots, an evolutionary innovation of the most species-rich group of cichlids, the haplochromines, where these conspicuous male fin colour markings are involved in mating. Applying a combination of RNAseq, comparative genomics and functional experiments, we identify two novel pigmentation genes, fhl2a and fhl2b, and show that especially the more rapidly evolving b-paralog is associated with egg-spot formation. We further find that egg-spot bearing haplochromines, but not other cichlids, feature a transposable element in the cis-regulatory region of fhl2b. Using transgenic zebrafish, we finally demonstrate that this region shows specific enhancer activities in iridophores, a type of pigment cells found in egg-spots, suggesting that a cis-regulatory change is causally linked to the gain of expression in egg-spot bearing haplochromines.
The de novo evolution of complex phenotypic traits poses a challenge to evolutionary biology. While selection explains adaptation and speciation in an adequate manner, it is more difficult to conceive how selection would trigger the origin of evolutionary novelties such as insect wings, feathers, tetrapod limbs, flowers, the mammalian placenta, beetle horns or butterfly eye-spots. The function of egg-spots has been implicated with the mating behaviour of the female-mouthbrooding haplochromines. Immediately upon spawning, a haplochromine female gathers up her eggs into the mouth; the male then presents his egg-spots to which the female responds by snatching and bringing her mouth close to the male’s genital opening; upon discharging sperm, the eggs become fertilized inside the female’s mouth. The mother subsequently broods and carries her progeny in the oral cavities for several weeks after fertilization.

Here we are interested in the molecular basis of the anal fin egg-spots of haplochromine cichlids. The main advantages of the cichlid egg-spot system are that (i) the evolutionary innovation of interest emerged just a few million years ago and hence is recent compared with most other evolutionary novelties studied so far; (ii) the phylogenetic context in which the novel trait evolved is known and living sister clades to the lineage featuring the novelty still exist, and (iii) the genomes of two outgroup species lacking the trait and of three derived species featuring the trait are available. This allows us to study early events involved in the origin of an evolutionary innovation in an assemblage of phenotypically diverse, yet closely related and genetically similar species. Using RNAseq, we identify two novel candidate pigmentation genes, the a- and b-paralogs of the four and a half LIM domain protein 2 (fhl2) gene, and show that both genes, but especially the more rapidly evolving b-copy, are associated with the formation of egg-spots. We then find that egg-spot bearing haplochromines—but not an egg-spot-less ancestral haplochromine and not the representatives from more basal

---

**Figure 1 | The egg-spots of haplochromine cichlids.** (a) Phylogeny of the East African cichlid fishes based on a new multimarker data set. The haplochromines are the most species-rich and derived group of cichlids in East Africa. One of the common features of haplochromines is the presence of egg-spots on the anal fin of males. Note that one of the ancestral lineages, represented here by *P. philander*, does not show this characteristic trait. *Substr-br*, substrate brooders; *mouthbr*, mouthbrooders; *spp.*: species. (b) Examples of male anal fin patterns in East African cichlids. Haplochromine egg-spots (upper panel) vary in size, shape, number and colouration. Non-haplochromines and basal haplochromine *P. philander* (lower panel) do not show this trait. (c) A typical mating cycle of haplochromine cichlids.
cichlid lineages—exhibit a transposable element insertion in close proximity to the transcription initiation site of fhl2b. A functional assay with transgenic zebrafish reveals that only a haplochromine-derived genetic construct featuring the SINE (short interspersed repetitive element) insertion drove expression in a special type of pigment cells, iridophores. Together, our data suggest that a cis-regulatory change (probably in the form of a SINE insertion) is responsible for the gain of expression of fhl2b in iridophores, contributing to the evolution of egg-spots in haplochromine cichlids.

**Results**

**fhl2 paralogs: novel candidates for egg-spot morphogenesis.** As a first step, we performed an illumina-based comparative transcriptomic experiment (RNAseq) between male (with egg-spots) and female (without egg-spots) anal fins in the haplochromine cichlid *Astatotilapia burtoni*. Two of the most differentially expressed genes according to RNAseq were the a- and b-paralogs of fhl2 (~4 log₂-fold and ~5 log₂-fold differences, respectively; see Supplementary Table 2). These paralogs result from the teleost genome duplication\(^{17}\) (Supplementary Fig. 2). The four and a half LIM domain protein 2 (Fhl2) is known as a transcriptional co-activator of the androgen receptor and the Wnt-signalling pathway\(^{18,19}\); Fhl2 plays a role in cell-fate determination and pattern formation, in the organization of the cytoskeleton, in cell adhesion, cell motility and signal transduction; furthermore, it regulates the development of heart, bone and musculature in vertebrates\(^{20,21}\).

Expression of fhl2a and fhl2b is egg-spot specific. To confirm the results obtained by RNAseq, we performed quantitative real-time PCR (qPCR) experiments (Fig. 2a), this time also comparing egg-spot versus non-egg-spot tissue within male anal fins. In addition, we tested another haplochromine species, *Cynotilapia pulpican*, with a different egg-spot arrangement to exclude positional effects of gene expression on the anal fin. In both species, the two duplicates of fhl2 were overexpressed in egg-spots (A. burtoni: fhl2a: \(t_s = 10.77, P = 0.0001; t_{fhl2b}: t_s = 4.362, P = 0.0073\). C. pulpican: fhl2a: \(t_s = 5.031, P = 0.0073; t_{fhl2b}: t_s = 9.154, P = 0.0008\)). We then tested the expression of both fhl2 paralogs in the four main developmental stages of egg-spot formation in *A. burtoni*\(^{22}\) and compared it with other candidate pigmentation genes (including the previously identified xanthophore marker *csf1ra*, the melanophore marker *mitfa* and the iridophore marker *pnp4a*). We found that the expression of both fhl2 paralogs increases substantially throughout anal fin and egg-spot development, and both genes showed higher expression levels compared with the other pigmentation genes (Fig. 2b); fhl2b shows the highest increase in expression exactly when egg-spots begin to form. Furthermore, we corroborate that the expression...
domain of both fhl2a and fhl2b matches the conspicuously coloured inner circle of egg-spots with RNA in situ hybridization (see Fig. 2c for results on fhl2b).

fhl2a and fhl2b evolved under purifying selection. In general, phenotypic differences can arise via mutations affecting the function of proteins or via changes in gene regulation. Therefore, we examined coding sequence evolution in the two fhl2 paralogs to test for positive selection and potential change of function in a phylogenetically representative set of 26 East African cichlids. We found that the two fhl2 genes are highly conserved in cichlids, with few amino-acid differences between species and an average genetic divergence (0.4% in fhl2a and 0.7% in fhl2b) that lies below the transcriptome-wide average of 0.95% (ref. 23). None of the observed amino-acid changes was correlated with the egg-spot phenotype (Supplementary Table 7).

Greater functional specialization of fhl2b in haplochromines. Usually, after a gene duplication event, the duplicates go through a period of relaxed selection, during which one of the two copies can diversify and acquire new functions. We found that the b-copy of fhl2 shows an elevated rate of molecular evolution compared with its paralog (fhl2a), which more closely resembles the ancestral sequence (Fig. 3a). An additional series of qPCR experiments in 12 tissues revealed that, in cichlids, fhl2a is primarily expressed in heart, bony structures and muscles, whereas fhl2b is highly expressed in the eye, and further in skin and the egg-spots of haplochromines (Fig. 3b,c). This is different to the gene expression profiles in medaka, where both duplicates are highly expressed in heart, skin and eye tissues; and in zebrafish, where the two paralogs are primarily expressed in heart, eye and (pharyngeal) jaw tissues, with fhl2a showing rather low levels of gene expression (Supplementary Figs 3 and 4). When compared with the other teleost fishes examined here, our results suggest that the haplochromine fhl2a retained most of the previously described functions, whereas the more rapidly evolving fhl2b obtained new expression patterns. Together, the gene expression profile and the pattern of sequence evolution make fhl2b a prime candidate gene for the morphogenesis of haplochromine egg-spots.

fhl2b shows an AFC-SINE insertion in species with egg-spot. Since there were no changes in the coding regions of fhl2a and fhl2b that are specific to the egg-spot bearing haplochromines, we shifted our focus towards the analysis of putative regulatory elements, exploring the recently available genomes of five East African cichlids (including the egg-spot bearing haplochromines A. burtoni, P. nyererei, Metriaclima zebra and the egg-spot-less non-haplochromines Neolamprologus brichardi and Oreochromis niloticus). The non-coding region of fhl2a shows homology with other teleosts (Oryzias latipes, Takifugu rubripes, Tetraodon nigroviridis and Gasterosteus aculeatus) and we identified four conserved non-coding elements (CNEs) in all species examined (Supplementary Fig. 5a). These CNEs might thus represent conserved regulatory regions responsible for ancestral conserved functions of fhl2a in teleosts. We might be missing cichlid-specific regulatory regions in important upstream regions although, as our capacity to detect lineage-specific enhancers is limited owing to the small sample size for each lineage and the high background conservation level present in cichlids.

Concerning fhl2b, we did not find any CNE that is shared by cichlids and other teleosts (Supplementary Fig. 5b). Strikingly, however, we found a major difference that is shared by the three egg-spot bearing haplochromines: the presence of a transposable element upstream of fhl2b. Specifically, we identified a SINE belonging to the cichlid-specific AFC-SINEs (African cichlid family of SINEs25), which inserted ~800 bp upstream of the transcriptional start site of fhl2b (Supplementary Fig. 6). To confirm that this insertion is associated with the egg-spot phenotype, we sequenced the upstream region of fhl2b in 19 cichlid species. The insertion was indeed present in nine additional, egg-spot bearing haplochromine species, yet absent in all 10 non-haplochromines examined (Supplementary Table 8). Importantly, we found that one haplochromine species lacks the AFC-SINE element, namely P. philander. This species belongs to one of the basal lineage of haplochromines (Fig. 1a), which is characterized by the absence of egg-spots (Fig. 1b). This suggests that the AFC-SINE upstream of fhl2b is not characteristic to the entire haplochromine clade, but to those that feature egg-spots, thus linking the SINE insertion to the origin of this evolutionary innovation.

Haplochromine fhl2b regulatory region drives iridophore expression. A long-standing hypothesis proposes that ubiquitous genomic repeat elements are potential regulators of transcription, and could thereby generate evolutionary variations and novelties26,27. SINEs are known for their capability of ‘transcriptional rewiring’, that is, to change the expression patterns of genes by bringing along new regulatory sequences when inserted in close proximity to a gene’s transcriptional initiation site27. In order to test whether the insertion of an AFC-SINE close to fhl2b functions as an enhancer of gene expression, we aimed for a functional experiment. We were particularly interested to find out whether there were changes in enhancer activity between AFC-SINE-positive haplochromines and other cichlids lacking both the insertion and the egg-spot phenotype. To this end, we designed reporter constructs containing the upstream region of fhl2b (~2 kb upstream to intron 1) of three cichlid species linked to the coding region of green fluorescent protein (GFP), and injected these constructs into zebrafish (Danio rerio) embryos to generate transgenic lines. We switched to the zebrafish system here, as no functioning transgenesis was available for haplochromine cichlids at the time the study was performed (owing to the small number of eggs per clutch associated with the characteristic female-mouthbrooding behaviour). The three constructs were derived from A. burtoni (haplochromine with egg-spots, AFC-SINE⁺), P. philander (haplochromine without egg-spots, AFC-SINE⁻) and N. sexfasciatus (lamprologine, AFC-SINE⁻), respectively (Fig. 4a).

We were able to produce stable transgenic zebrafish lines for each of the three constructs to examine the expression of GFP. Importantly, we found striking differences in expression between the A. burtoni construct and the two constructs lacking the AFC-SINE. Of the three reporter lines, only the AFC-SINE⁺ showed GFP expression in iridophores, a silvery-reflective type of pigment cells (Fig. 4b,c and Supplementary Fig. 7). This experiment demonstrates the presence of novel enhancer activities in the regulatory region of fhl2b in derived haplochromines and strongly suggests that these came along with the SINE insertion.

Iridophores and egg-spot development. The egg-spot phenotype has previously been associated with pigment cells containing pteridines (xanthophores)16,22, whereas our new results indicate an auxiliary role of iridophores in egg-spot formation. We thus re-evaluated the adult egg-spot phenotype by removing the pteridine pigments of the xanthophores (Fig. 4e). We indeed found that A. burtoni egg-spots show a high density of iridophores, which is further corroborated by the increase in gene expression of the iridophore marker pnp4a during egg-spot
Figure 3 | Gene tree of the two fhl2 paralogs and expression profiling in East African cichlid fishes. (a) Bayesian inference phylogeny of the orthology and paralogy relationships between cichlids, other teleosts (O. latipes, D. rerio, Ta. rubripes and G. aculeatus) and tetrapods (Anolis carolinensis and Mus musculus) fhl2 sequences. This gene tree is important for generating functional hypotheses about both duplicates, and to infer the ancestral state of the fhl2 gene before duplication. Our phylogeny indicates that fhl2a is more similar to the ancestral state, while fhl2b is apparently evolving faster in teleosts. Values at the tree nodes represent posterior probabilities. In Supplementary Fig. 2, we present a synteny analysis supporting the origin of teleost fhl2 duplicates in the teleost genome duplication. (b) Relative quantity (RQ) of fhl2a and fhl2b gene expression in 12 tissues (three replicates per tissue) in C. pulpican, an egg-spot bearing haplochromine from Lake Malawi. The error bars represent the s.e.m. (c) RQ of fhl2a and fhl2b gene expression in 12 tissues in N. crassus, a substrate spawning lamprologine that has no egg-spots. In both species, gill tissue was used as reference; in N. crassus, ‘egg-spots’ corresponds to the fin region where haplochromines would show the egg-spot trait. In C. pulpican (b), fhl2a is highly expressed in heart, in pigmented tissues (eye, skin and egg-spot) and in craniofacial traits (oral jaw and lower pharyngeal jaw); fhl2b is mainly expressed in the pigmented tissues. N. crassus (c) shows a similar expression patterns for fhl2a and fhl2b, with the difference that fhl2a does not show high expression levels in jaw tissues, and fhl2b is not highly expressed in skin and fin tissue. These results suggest that fhl2b shows a higher functional specialization, and that it might be involved in the morphogenesis of sexually dimorphic traits such as pigmented traits including egg-spots. LPJ, lower pharyngeal jaw bone.
formation (Fig. 2b). With the exception of the proximal region of the anal fin, the number of iridophores is greatly reduced in the fin tissue surrounding egg-spots (Supplementary Fig. 8a). Interestingly, this proximal region is the only area of the anal fin besides the egg-spots where we observed fhl2 expression with RNA in situ hybridization (see Fig. 2c for fhl2b), once more linking fhl2 expression with iridophores (and less so with xanthophores, which are very rare in this region). In the non-haplochromine N. crassus, which features a yellow anal fin pattern containing xanthophores, we did not find iridophores in

---

**Figure 4 | The molecular basis of egg-spot formation.**

(a) The egg-spot bearing haplochromines feature an AFC-SINE insertion in close proximity to the transcriptional start site of fhl2b, which is absent in the ancestral and egg-spot-less genus Pseudocrenilabrus and in all non-haplochromines. The sequences from the three species shown here were the ones used to engineer the reporter constructs, where the fhl2b coding sequence was substituted by GFP. (b) In transgenic zebrafish, only the AFC-SINE \(^+\) construct showed GFP expression in the iridophores, a type of pigment cells (one of them is indicated by a yellow arrow). The upper panel depicts bright-field images of 3-day-old zebrafish embryo trunks; the lower panel shows the respective embryos under ultraviolet light. The green signal in the AFC-SINE negative N. sexfasciatus line (marked with an asterisk) is auto-fluorescence from the yolk extension. (c) Higher magnification image from A. burtoni AFC-SINE \(^+\) reporter construct driving GFP expression in the iridophores. Orientation in b,c: bottom: anterior, top: posterior. (d) Top-down view of a trunk of a 3-day-old AFC-SINE-positive zebrafish embryo. The left panel depicts a bright-field image where the iridophores of the dorsal stripe are illuminated by the incident light (yellow arrows). The right panel depicts GFP expression of the same embryo. The GFP signal co-localizes with iridophores. (e) Cellular basis of egg-spots: this series of images shows that egg-spots are made up of xanthophores, iridophores and scattered melanophores. Image 1 shows an A. burtoni fin with two egg-spots. Image 2 shows the same fin without pteridine pigments (xanthophores are not visible anymore). Images 3 and 4 are higher magnification images of the egg-spots without pteridine under slightly different light conditions confirming that egg-spots have a high density of iridophores (examples of this cell type are highlighted with arrows). UTR, untranslated region.
the xantophore-rich region (Supplementary Fig. 9), suggesting that the xantophore/iridophore pattern is unique to haplochromine egg-spots. Importantly, we also observed that iridophores appear early in the newly forming egg-spot of haplochromines, that is, before the first xantophores start to aggregate (Supplementary Fig. 8b).

In zebrafish, stripe development is initiated by iridophores, which serve as morphological landmarks for stripe orientation in that they attract further pigment cells such as xantophores by expressing the csf1 ligand gene\(^\text{29,30}\). Interestingly, it has previously been shown that a gene encoding a Csfr receptor known for its role in xanthophore development in zebrafish, csf1ra, is expressed in haplochromine egg-spots\(^\text{16}\). We thus examined the expression of the ligand csf1b and showed that its relative level of gene expression doubles during egg-spot development, and that this increase coincides with the emergence of the phenotype (Supplementary Fig. 10). This leads us to suggest that a similar pigment cell type interaction mechanism might be involved in egg-spot patterning as the one described for zebrafish\(^\text{29,30}\). The specific mode of action of fin patterning in haplochromine cichlids, and how Fhl2b interacts with the Csfr/Csfr1 system, remains to be studied in the future.

**Contribution of fhl2a in egg-spot formation.** The role of the more conserved and functionally constrained a-paralog of fhl2 in egg-spot development cannot be dismissed. Its temporally shifted increase in gene expression compared with fhl2b (Fig. 2b) suggests that fhl2a most likely acts as a more downstream factor involved in pigment pattern formation. We were nevertheless interested in uncovering the regulatory region responsive for this expression pattern. The first intron of fhl2a shows two CNEs that are common across percomorph fish (Supplementary Fig. 5). Using the same strategy as described above, we generated a transgenic zebrafish line containing exon 1 and intron 1 of \textit{A. burtoni} linked to GFP. This construct drove expression in heart in zebrafish embryos, which is consistent with the reported function of fhl2a in tetrapods\(^\text{20}\), whereas there was no indication of a pigment cell related function for this reporter construct (Supplementary Fig. 7e). An alignment between the genomic regions of the two fhl2 paralogs shows that there were no CNEs in common and generally very little homology between them, suggesting that the regulation of the expression of fhl2a in egg-spots might proceed in a different way (Supplementary Fig. 11).

**Discussion**

In this study, we were interested in the genetic and developmental basis of egg-spots, an evolutionary innovation of the most species-rich group of cichlids, the haplochromines, where these conspicuous colour markings on the anal fins of males play an important role in mating\(^\text{11-13}\) (Fig. 1).

We first performed a comparative RNAseq experiment that led to the identification of two novel candidate pigmentation genes, the a- and b-paralogs of the four and a half LIM domain protein 2 (fhl2) gene. We then confirmed, with qPCR and RNA \textit{in situ} hybridization, that the expression domain of both duplicates indeed matches the conspicuously coloured inner circle of egg-spots (Fig. 2). Especially the more rapidly evolving b-copy of fhl2 emerged as strong candidate gene for egg-spot development, as its expression profile mimics the formation of egg-spots (Figs 2b and 3). Interestingly, we found that the egg-spot bearing haplochromines, but not other cichlids, feature a transposable element in the \textit{cis}-regulatory region of fhl2b. Finally, making use of transgenic zebrafish, we could show that a \textit{cis}-regulatory change in fhl2b in the ancestor of the egg-spot bearing haplochromine cichlids (most likely in the form of the AFC-SINE insertion) resulted in a gain of expression in iridophores, a special type of pigment cells found in egg-spots (Fig. 4). This in turn might have led to changes in iridophore cell behaviour and to novel interactions with pigmentation genes (csf1b, csf1ra and pnp4a), thereby contributing to the formation of egg-spots on male anal fins. The specific mode of action of the SINE insertion, and how the fhl2b locus interacts with these other pigmentation genes remains elusive at present. Addressing these questions would require functional studies in haplochromines, which are, however, hampered by the specific mechanisms involved in the trait complex of interest (mouthbrooding makes it notoriously difficult to obtain enough eggs—in a controlled manner—to make such experiments feasible).

Our results are also suggestive of an important role of the a-copy of fhl2 in cichlid evolution. With our qPCR experiments, we provide strong evidence that fhl2a is involved in jaw tissue in zebrafish (Supplementary Fig. 3) and, importantly, in the pharyngeal jaw apparatus of cichlids (Fig. 3b,c), another putative evolutionary innovation of this group. The pharyngeal jaw apparatus is a second set of jaws in the pharynx of cichlids that is functionally decoupled from the oral jaws and primarily used to process food\(^\text{11,12,15}\). Interestingly, fhl2a has previously been implicated in the evolution of fleshy lips in cichlids\(^\text{31}\), which is yet another ecologically relevant trait. From a developmental perspective, the main tissues underlying these traits—the craniofacial cartilage (the jaw apparatus) and pigment cells (egg-spots)—have the same origin, the neural crest, which itself is considered an evolutionary key innovation of vertebrates\(^\text{32}\). It thus seems that the function of fhl2 in cichlids may have been split into (a) an ecologically important, that is, naturally selected, scope of duties, and (b) a role in colouration and pigmentation more likely to be targeted by sexual selection.

Taken together, our study permits us to propose the following hypothesis for the origin of cichlid egg-spots: In one of the early, already female-mouthbrooding, haplochromines the insertion of a transposable element of the AFC-SINE family in the \textit{cis}-regulatory region of fhl2b, and its associated recruitment to the iridophore pigment cell pathway, mediated the evolution of egg-spots on the anal fins—possibly from the so-called \textit{perleckmuster} common to many cichlids\(^\text{16}\). The conspicuous anal fin spots were fancied by haplochromine females, which—just like many other cichlids and also the ancestral and egg-spot-less haplochromine genus \textit{Pseudocrenilabrus}—have an innate bias for yellow/orange/red spots that resemble carotenoid-rich prey items\(^\text{33}\), leading to the fixation of the novel trait. In today’s haplochromines, egg-spots seem to have a much broader range of functions related to sexual selection\(^\text{34}\).

Most of the currently studied evolutionary innovations comprise relatively ancient traits (for example, flowers, feathers, tetrapod limb, insect wings and mammalian placenta), thereby making it difficult to scrutinize their genetic and developmental basis. Here we explored a recently evolved novelty, the anal fin egg-spots of male haplochromine cichlids. We uncovered a regulatory change in close proximity to the transcriptional start site of a novel iridophore gene that likely contributes to the molecular basis of the origin of egg-spots in the most rapidly diversifying clade of vertebrates. This, once more, illustrates the importance of changes in \textit{cis}-regulatory regions in morphological evolution\(^\text{2}\).

**Methods**

**Samples.** Laboratory strains of \textit{A. burtoni}, \textit{C. pulpicia}, \textit{Astatoreocromis aliausdi}, \textit{P. nyererei}, \textit{Labidochromis caeruleus}, \textit{Pseudotropheus elegans} and \textit{N. caussid} were kept at the University of Basel (Switzerland) under standard conditions (12 h light/12 h dark; 26 °C pH 7). Before dissection, all specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following an approved procedure (permit no. 2317).
issued by the cantonal veterinary office Basel). Individuals of all other specimens were collected in the southern region of Lake Tanganyika (Zambia) under the permit of Fisheries, Republic of Zambia, and processed in the field following our standard operating procedure9. Tissues for RNA extraction were stored in RNA later (Ambion, USA), and tissues for genomic DNA extraction were stored in ethanol and shipped to the University of Basel.

**RNA and DNA extractions.** Isolation of RNA was performed according to the TRIzol protocol (Invitrogen, USA) after incubating the dissected tissues in 750 μl of TRIzol at 4 °C overnight or, alternatively, for 8–16 h (in order to increase the RNA yield after long-term storage). The tissues were then homogenized with a Bead-Buster (FastPrep-24; MP Biomedicals, France). Subsequent DNA treatment was performed with DNA-Free kit (Ambion). 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). Genomic DNA was extracted using a high salt extraction method (modified from ref. 35).

**Differential gene expression analysis using RNAseq.** We used a transcriptomic approach (RNAseq) to identify genes differentially expressed between male and female anal fins of A. burtoni. Library construction and sequencing of RNA extracted from three male and three female anal fins (at the developmental stage of 30 mm standard length) were performed with the TruSeq RNA-to-cDNA Kit (Illumina, USA) following the manufacturer’s guidelines. These fragments were then sequenced on an Illumina Genome Analyzer IIX. Each sample was sequenced in one lane and with a read length of 76 bp.

The reads were then aligned to an embryonic A. burtoni reference transcriptome assembled by Broad Institute (http://www.broadinstitute.org/models/tilapia). This transcriptome is not annotated and each transcript has a nomenclature where the first term codes for the parent contig and the third term codes for alternatively spliced transcripts (CompX_cX_seqX). The reference transcriptome was indexed using NOVOINDEX (www.novocraft.com) with default parameters. Then, RNASeq variants (rRNAseq) were mapped against the reference transcriptome with a maximum alignment (t) score of 30, a minimum of good-quality base pair per read (i) of 25 and a successive trimming factor (s) of 5. Reads that did not match these criteria were discarded. Since the reference transcriptome has multiple transcripts/isosforms belonging to the same gene, all read alignment locations were reported (FALL). The mapping results were reported (o) in SAM format. The output SAM file was then transformed into BAM format, sorted, indexed and converted to count files (number of reads per transcript) using SAMTOOLS version 0.1.18 (ref. 47). The count files were subsequently concatenated into a single data set—count table—and analyzed using the software EDGE (Sung et al., 2010). Deseq2 (Love et al., 2014) was used to test significant differences in gene expression between male and female anal fins. The 10 most differentially expressed transcripts were identified by BLASTx48 against GenBank’s non-redundant database (Supplementary Table 2).

**Cloning of fhl2a and fhl2b and RNA in situ hybridization.** A. burtoni fhl2a and fhl2b coding fragments were amplified by PCR (for primer information, see Supplementary Table 3) using Phusion Master Mix with High Fidelity buffer (New England BioLabs, USA) following the manufacturer’s guidelines. These fragments were then sequenced into the human 18sRNA (Invitrogen) Plasmid extractions were done with GenElute Plasmid Miniprep Kit (Sigma-Aldrich). RNA probes were synthesized with the DIG RNA labelling kit (SP6/T7) (Roche). The insertion and direction of the fragments were confirmed by Sanger sequencing using M13 primers (available with the cloning kit) and BigDye Terminator reaction chemistry (Applied Biosystem). The in situ Hybridization Analyser (Applied Biosystems). In situ hybridization was performed in 12 fins from A. burtoni males, six for fhl2a and six for fhl2b. The protocol was executed as described in ref. 16, except for an intermediate protease K treatment (20 min at a final concentration of 15 μg ml−1) and for the hybridization temperature (65 °C).

**Synteny analysis of teleost fhl2 paralogs.** The Synteny Database (http://syntenydb.uoregon.edu27) was used to generate dotplots of the human FHL2 gene (ENSG00000115641) region on chromosome Hsa2 and the genomes of medaka (Supplementary Fig. 2a) and zebrafish (Supplementary Fig. 2b). Double-conserved synteny between the human FHL2 gene and the fhl2a and fhl2b paralogs in teleost genomes provide evidence that the fhl2a and fhl2b paralogs were generated during the teleost genome duplication.

**fhl2a/fhl2b coding region sequencing and analysis.** We then used cDNA pools extracted from anal fin tissue to amplify and sequence the coding region of fhl2a and fhl2b in a phylogenetically representative set of 26 cichlid species (21 Tangan- yikan species, three species from Lake Malawi and two species from the Lake Victoria basin). This taxon sampling included 14 species belonging to the haploclines and 12 species belonging to other East African cichlid tribes not featuring the egg-spot trait (Supplementary Table 4). fhl2a and fhl2b coding regions were fully sequenced (from start to stop codon) in five individuals per species in order to evaluate the rate of molecular evolution among cichlids. For PCR amplification, we used Phusion Master Mix and cichlid-specific primers (for primer information, see Supplementary Table 3) designed with Primer3 (ref. 55). PCR products were visualized with 1% agarose gel in 1× TAE buffer using GelRed (Biotium, USA). In cases where multiple bands were present, we purified the correct size fragment from the gel using the GenElute Gel Extraction Kit (Sigma-
Aldrich). 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 AB3130xL Genetic Analyzer. Sequences were corrected, trimmed and aligned manually in CODONCODE ALIGNER (CodonCode Corporation).

**fh2 phylogenetic analysis.** *fh2* and *fh2b* sequences from non-cichlid teleosts and *fh2* sequences from tetrapods were retrieved from ENSEMBL66 (species names, gene names and accession numbers are available in Supplementary Table 5). We then constructed gene trees based on these sequences and on a subset of the cichlid sequences obtained in the previous step (information available in Supplementary Table 4) in order to confirm the orthologous and paralogous relationships of both duplicates. Sequences were aligned with CLUSTALW2 (ref. 57) using default parameters. The most appropriate model of sequence evolution was determined using jmodeltest as described above. Phylogenetic analyses were performed with RAxML (1 million generations; 25% burn-in).

**Tests for positive selection in *fh2a* and *fh2b***. Using PAUP* 4.0b10 (ref. 58), we first compiled a maximum likelihood tree based on the mitochondrial ND2 gene, including all species used for the positive selection analyses (see Supplementary Table 6 for species and GenBank accession numbers). We used the GTR + Γ model with base frequencies and substitution rate matrix estimated from the data (as suggested by model selection). We then ran a maximum likelihood search implemented in RAxML v. 7.2.4 to test for branch-specific adaptive evolution in *fh2a* and *fh2b* applying the branch-site model (free-ratios model with Ω allowed to vary)59,60. The branch comparisons and results are shown in Supplementary Table 7.

**Identification of CNEs.** We then made use of the five available cichlid genomes61 to identify CNEs that could explain the difference in expression of *fh2a* and *fh2b* between haplochromines and non-haplochromines (note that there are three haplochromine genomes available: A. burtoni, Pu. nyererei, M. zebra; and two genomes belonging to more ancestral cichlid lineages: N. brichardi and Or. niloticus). For this analysis, we also included the respective genomic regions of four other teleost species (O. latipes, Ta. rubripes, Te. nigroviridis and G. aculeatus). More specifically, we extracted the genomic scaffolds containing *fh2a* and *fh2b* from the available cichlid genomes using BLAST v. 2.2.25 and testing individual scaffolds for the CNE characteristics. The resulting hits were refined in order to extract fragments of ≈5–6 kb in length containing the CNE motif as described above. These fragments were then aligned using the RepeatMasker program. We then used the SINE insertion sequence 5`-CCGCATGGGACAGTGCTGCTGAGCACC-3` to identify fragments containing CNEs, applying the RepeatMasker program (http://www.girinst.org/RepeatMasker). In order to confirm the orthologous and paralogous relationships of both duplicates, sequences were aligned with CLUSTALW2 using default parameters. The most appropriate model of sequence evolution was determined using jmodeltest as described above. Phylogenetic analyses were performed with RAxML (1 million generations; 25% burn-in).

**Fixation and dehdydration of cichlid fins.** In order to determine the pigment cell composition of egg-spots (and especially whether they contain iridophores in addition to xanthophores), we dissected *A. burtoni* anals fins. To better understand the morphological differences between non-haplchromine and haplochromine fins, we further dissected three *N. crassus* anal fins. To visualize iridophores, we removed the pteridine pigments of the overlying xanthophores by fixating the fin in 4% paraformaldehyde–PBS for 1 h at room temperature and washing it in a series of methanol:PBS dilutions (25%, 50%, 75% and 100%). Pictures were taken after 6 days in 100% methanol at −20 °C.

**References**
1. Pigliucci, M. What, if anything, is an evolutionary novelty? Philos. Sci. 75, 887–898 (2008).
2. Carroll, S. B. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. Cell 134, 25–36 (2008).
3. Wagner, G. P. & Lynch, V. J. Evolutionary novelties. Curr. Biol. 20, R48–R52 (2010).
4. Wagner, A. The molecular origins of evolutionary innovations. Trends Genet. 27, 397–410 (2011).
5. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design (Blackwell Science, 2001).
6. Schierwater, B. Evidence for ecological speciation and its alternative. Science 323, 737–741 (2009).
7. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. Nat. Genet. 43, 1154–1159 (2011).
8. Beldade, P. & Brakefield, P. M. The genetics and evo-devo of butterfly wing patterns. Nat. Rev. Genet. 3, 442–452 (2002).
9. Salzburger, W., Mack, T., Verheyen, E. & Meyer, A. Out of Tanganika: genesis, explosive speciation, key-innovations and phylogeny of the haplochromine cichlid fishes. BMC Evol. Biol. 5, 17 (2005).
10. Kocher, T. D. Adaptive evolution and explosive speciation: the cichlid fish model. Nat. Rev. Genet. 5, 288–298 (2004).
11. Salzburger, W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 18, 169–185 (2009).
12. Fryer, G. & Iles, T. The Cichlid Fishes of the Great Lakes of Africa: their Biology and Evolution (Oliver & Boyd, 1972).
13. Wickler, W. 'Egg-dummies' as natural releasers in mouth-breeding cichlids. Nature 194, 1092–1093 (1962).
14. Queiroz, M. E. & Salzburger, W. How cichlids diversify. Science 338, 619–621 (2012).
15. Mushchik, M., Indermahr, A. & Salzburger, W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22, 2362–2368 (2012).
16. Salzburger, W., Brasch, 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. 5, 51 (2007).
17. Brasch, I. & Posluthwait, J. in Polyploidy and Genome Evolution (eds Solits, D. E. & Solits, P. S.) 341–383 (Springer, 2012).
18. Müller, J. M. et al. H2L2, a novel tissue-specific coactivator of the androgen receptor. *EMBO J.* 19, 359–369 (2000).
19. Bruhn, J. et al. The LIM-only protein H2L2 controls osteoblast mesenchymal cell differentiation through non-canonical Wnt signalling. *Bone* 50, S76 (2012).
20. Johannessen, M., Müller, S., Hansen, T., Moens, U. & Van Ghelue, M. The multifunctional roles of the four-and-a-half-LIM only protein H2L2. *Cell. Mol. Life Sci.* 63, 268–284 (2006).
21. Kadras, J. L. & Beckerle, M. C. The LIM domain: from the cytoskeleton to the nucleus. *Nat. Rev. Mol. Cell Biol.* 5, 920–931 (2004).
22. Heule, C. & Salzburger, W. The ontogenetic development of egg-spots in the haplochromine cichlid fish *Astatotilapia burtoni*. *J. Fish Biol.* 78, 1588–1593 (2011).
23. Baldo, L., Santos, M. E. & 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.* 3, 443–455 (2011).
24. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* 290, 1151–1155 (2000).
25. Takahashi, K., Tera, Y., Nishida, M. & Okada, N. A novel family of short interspersed repetitive elements (SINEs) from cichlids: the patterns of insertion of SINEs at orthogonal loci support the proposed monophyly of four major groups of cichlid fishes in Lake Tanganyika. *Mol. Biol. Evol.* 15, 391–407 (1998).
26. Britten, R. J. & Davidson, E. H. Gene regulation for higher cells: a theory. *Science* 165, 349–357 (1969).
27. Britten, R. J. & Davidson, E. Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Q Rev. Biol.* 46, 111–138 (1971).
28. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405 (2008).
29. Frohnhöfer, H. G., Krauss, J., Maischein, H.-M. & Nüsslein-Volclhard, C. Iridophores and their interactions with other chromatophores are required for stripe formation in zebrafish. *Development* 140, 2997–3007 (2013).
30. Patterson, L. B. & Parichy, D. M. Interactions with iridophores and the tissue environment required for patterning melanophores and xanthophores in the zebrafish adult pigment stripe formation. *PLoS Genet.* 9, e1003561 (2013).
31. Manousaki, T. et al. Parsing parallel evolution: ecological divergence and differential gene expression in the adaptive radiations of thick-lipped Midas cichlid fishes from Nicaragua. *Mol. Ecol.* 22, 650–669 (2013).
32. Shimeld, S. M. & Holland, P. W. Vertebrate innovations. *Proc. Natl Acad. Sci. USA* 97, 4449–4450 (2000).
33. Eggert, B., Klaefiger, Y., Theis, A. & Salzburger, W. A sensory bias has triggered the evolution of egg-spots in cichlid fishes. *PLoS ONE* 6, e25601 (2011).
34. Theis, A., Salzburger, W. & Parichy, D. The function of anal fin egg-spots in the cichlid fish *Astatotilapia burtoni*. *Bioinformatics* 27, e197 (2012).
35. Miller, S., Dykes, D. & Polesky, H. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res.* 16, 1215 (1988).
36. Meyer, B. S. & Salzburger, W. A novel primer set for multilocus phylogenetic inference in East African cichlid fishes. *Mol. Ecol. Resour.* 12, 1097–1104 (2012).
37. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376–380 (2005).
38. Binladen, J. et al. The use of coded PCR primers enables high-throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. *PLoS ONE* 2, e917 (2007).
39. Schmieder, R. & Edwards, R. Quality control and preprocessing of metagenomic data sets. *Bioinformatics* 27, 863–864 (2011).
40. Karsee, M. et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* 28, 1647–1649 (2012).
41. Katoh, K., Kuma, K., Toh, H. & Miyata, T. MAFFT version 5: improvement in alignment accuracy and speed. *Nucleic Acids Res.* 33, 518–524 (2005).
42. Darrida, D., Taboada, G. L., Doallo, R. & Posada, D. *JModelTest* 2: more models, new heuristics and parallel computing. *Nat. Methods* 9, 772 (2012).
43. Schwarz, G. Estimating the dimension of a model. *Ann. Stat.* 6, 461–464 (1978).
44. Ronquist, F. et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542 (2012).
45. Zwickl, D. J. Genetic Algorithm Approaches for the Phylogenetic Analysis of Large Biological Sequence Data Sets Under the Maximum Likelihood Criterion (University of Texas, 2006).
46. Sukumaran, J. & Holder, M. T. DendroPy: a Python library for phylogenetic computing. *Bioinformatics* 26, 1569–1571 (2010).
Additional information
Accession codes: All nucleotide sequences reported in this study have been deposited in GenBank/EMBL/DDBJ under the accession codes KM263618 to KM264016. All the short reads have been deposited in GenBank/EMBL/DDBJ Sequence Read Archive (SRA) under the BioProject ID PRJNA25755.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing financial interests: The authors declare no competing financial interests.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Santos, M. E. et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 5:5149 doi: 10.1038/ncomms6149 (2014).

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/4.0/