

RESEARCH ARTICLE

Open Access

# Sequence analyses of the *distal-less homeobox* gene family in East African cichlid fishes reveal signatures of positive selection

Eveline T Diepeveen\*, Fabienne D Kim and Walter Salzburger

## Abstract

**Background:** Gen(om)e duplication events are hypothesized as key mechanisms underlying the origin of phenotypic diversity and evolutionary innovation. The diverse and species-rich lineage of teleost fishes is a renowned example of this scenario, because of the fish-specific genome duplication. Gene families, generated by this and other gene duplication events, have been previously found to play a role in the evolution and development of innovations in cichlid fishes - a prime model system to study the genetic basis of rapid speciation, adaptation and evolutionary innovation. The *distal-less homeobox* genes are particularly interesting candidate genes for evolutionary novelties, such as the pharyngeal jaw apparatus and the anal fin egg-spots. Here we study the *dlx* repertoire in 23 East African cichlid fishes to determine the rate of evolution and the signatures of selection pressure.

**Results:** Four intact *dlx* clusters were retrieved from cichlid draft genomes. Phylogenetic analyses of these eight *dlx* loci in ten teleost species, followed by an in-depth analysis of 23 East African cichlid species, show that there is disparity in the rates of evolution of the *dlx* paralogs. *Dlx3a* and *dlx4b* are the fastest evolving *dlx* genes, while *dlx1a* and *dlx6a* evolved more slowly. Subsequent analyses of the nonsynonymous-synonymous substitution rate ratios indicate that *dlx3b*, *dlx4a* and *dlx5a* evolved under purifying selection, while signs of positive selection were found for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b*.

**Conclusions:** Our results indicate that the *dlx* repertoire of teleost fishes and cichlid fishes in particular, is shaped by differential selection pressures and rates of evolution after gene duplication. Although the divergence of the *dlx* paralogs are putative signs of new or altered functions, comparisons with available expression patterns indicate that the three *dlx* loci under strong purifying selection, *dlx3b*, *dlx4a* and *dlx5a*, are transcribed at high levels in the cichlids' pharyngeal jaw and anal fin. The *dlx* paralogs emerge as excellent candidate genes for the development of evolutionary innovations in cichlids, although further functional analyses are necessary to elucidate their respective contribution.

**Keywords:** *Distal-less homeobox* gene, Molecular evolution, Cichlid fishes, Teleost fishes, Positive selection, Differential selection, Gene duplication, dN/dS

## Background

Teleost fishes (Teleostei) are among the most diverse lineages on Earth and with nearly 30,000 species the most species-rich vertebrate group. This is in stark contrast to the more basal non-teleost ray-finned fishes that are characterized by small numbers of species. A causal explanation for this discrepancy in speciation rates between the derived Teleostei and the non-teleost ray-finned fishes might be the fish-specific genome duplication (FSGD) that

occurred in the ancestor of modern teleosts ([1-4] and references therein). It has been hypothesized that the FSGD has laid down the genetic conditions necessary for the evolution of phenotypic diversity [5], although the exact causes of diversification of such a large clade are likely to be more complex and most probably also include other factors [6].

The Hox gene clusters, which evolved through both tandem and whole genome duplications, represent illustrative examples for the contribution of duplicated genes to morphological evolution across the animal kingdom (see e.g., [7-9]). Together with other homeotic genes,

\* Correspondence: evelinetdiepeveen@gmail.com  
Zoological Institute, University of Basel, Vesalgasse 1, 4051 Basel, Switzerland

Hox genes play a crucial role in the development of the multicellular body plan (e.g., anterior-posterior patterning; [10]). Furthermore, Hox genes are known to be involved in the development of evolutionary novelties, such as walking limbs and the wings of insects [11-15]. It has been shown that different mechanisms such as *cis*-regulatory evolution, changes in protein function and post-transcriptional regulation of the Hox genes contribute to morphological diversification (reviewed in e.g., [8,15,16]).

East African cichlid fishes show a remarkable level of phenotypic diversity between closely related species and constitute the most diverse adaptive radiations known [17-21]. Although several smaller radiations of cichlid fishes exist outside of Africa (e.g., in Central and South America), an astonishingly high number of cichlid species (close to 1900 species [22]) evolved in and around lakes Malawi, Victoria and Tanganyika in the last few million to several thousand years [23,24]. The various cichlid species differ in body shape, coloration, reproductive biology and mouth morphology [25-27] - traits which are thought to, at least partly, underlie the evolutionary success of cichlid fishes [18,27,28]. Furthermore, several morphological innovations are unique to cichlids or specific lineages thereof. The highly modified and morphologically diverse pharyngeal jaw apparatus, for example, correlates with the diversity in foraging strategies exploited by the different cichlid species [27-29]. The occurrence of several color morphs within species, sexual color dimorphism and anal fin egg-spots are three characteristic features of the extremely species-rich and mouthbrooding haplochromine lineage [30].

As a result of their great phenotypic diversity and high number of species, cichlid fishes provide an ideal set up to examine the genetic basis of rapid speciation, evolutionary innovations and adaptation [21,31-37]. An important strategy is the study of so-called candidate genes, i.e., genes with known functions in development in other organisms such as zebrafish. For example, it has been shown that *csf1ra*, which was identified as xanthophore marker in zebrafish [38,39] is involved in the morphogenesis of the egg-spots of haplochromine cichlids [31]. Furthermore, species-specific jaw shapes of different cichlid species correlate with differences in early *bmp4* expression patterns, a gene which has also the potential to change the mandibular morphology in zebrafish [40]. Many of these candidate genes belong to larger gene families such as the endothelin family of ligands and receptors that are putatively involved in the morphogenesis of the pharyngeal jaw apparatus and pigmentation [36], and the above mentioned Hox gene clusters [41].

Recently, Renz et al. [35] characterized seven *distal-less* homeobox (*dlx*) genes and examined their expression patterns in the developing pharyngeal arches and/or pharyngeal teeth of the haplochromine cichlid *Astatotilapia burtoni*. The vertebrate *dlx* genes are widely known for their

crucial roles in the development (of components) of the nervous system, craniofacial skeleton and connective tissue and in the formation of appendages [reviewed in 42]. These functions seem to be conserved across a wide range of animal taxa. For example, the vertebrate *dlx* genes are homologs of, and share several functions with, the single *Distal-less* (*dll*) gene of *Drosophila* [42]. Within vertebrates, the expression patterns of *dlx* homologs is similar in early development [35,42-45]. At the same time, *dlx* genes have been implicated with evolutionary novelties such as the eyespots in various butterfly species [46-48], the insect antenna [49,50] and the vertebrate craniofacial bones [51].

Phylogenetic analyses and the chromosomal arrangements of the vertebrate *dlx* genes suggest that the extant *dlx* repertoire has evolved by an initial tandem duplication, followed by two rounds of whole genome duplication in the lineage towards vertebrates and a third one in the lineage towards teleost fishes, the FSGD [35,44,45,52]. These duplication events resulted in multiple so-called *dlx* clusters, in which two *dlx* genes are located in a tail-to-tail arrangement on the respective chromosome. Linked *dlx* genes are transcribed coincidentally due to shared *cis*-regulating elements in the intergenic regions [35,42,43]. Four of these *dlx* clusters have been identified in teleost fish; *dlx1a-dlx2a*, *dlx3a-dlx4a*, *dlx3b-dlx4b* and *dlx5a-dlx6a* [43,44]. Seven of these *dlx* genes have been identified in the cichlid *A. burtoni*, where they are expressed in tissues that make up putative evolutionary innovations [35].

Here, we analyzed the *dlx* repertoire and diversity in detail in a phylogenetically representative set of 23 East African cichlid species in order to study the molecular evolution of this prominent developmental gene family. To this end, we first performed phylogenetic comparisons of the *dlx* proteins including the sixty amino acids long homeobox domain in a range of teleost fishes in combination with blast searches of these sequences against the draft genomes of four cichlid species. Teleost and cichlid-specific phylogenies were examined to compare the rate of evolution of both between and within *dlx* gene trees. Several studies have shown that loci putatively involved in evolutionary innovations are characterized by adaptive protein evolution in cichlids [31,36,53]. Therefore, all loci were screened for elevated rates of protein evolution by means of  $d_N/d_S$  analyses. Our analyses indicate the presence of *dlx3a* in cichlids and that the *dlx* repertoire of cichlid fishes is shaped by differential selection pressures and rates of evolution, with signs of positive selection on specific sites in *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b*.

## Methods

### Dlx protein sequence comparison in teleost fishes

The sequences of nine *dlx* proteins (i.e., *dlx1a*, *dlx2a*, *dlx2b*, *dlx3a*, *dlx3b*, *dlx4a*, *dlx4b*, *dlx5a* and *dlx6a*) of seven teleost

species (i.e., zebrafish (*Danio rerio*), Atlantic cod (*Gradus morhua*), three-spined stickleback (*Gasterosteus aculeatus*), spotted green pufferfish (*Tetraodon nigroviridis*), Japanese pufferfish (*Takifugu rubripes*), Japanese medaka (*Oryzias latipes*) and Nile Tilapia (*Oreochromis niloticus*)) were obtained from Ensemble (release 68, July 2012; see Additional file 1 for accession numbers). *Dlx2b* was excluded from all further analyses, due to its lineage-specific loss in percomorphs, to which all studied species belong except *D. rerio* and *G. morhua* (see [35]). Sequences were aligned with Tcoffee [54,55], ambiguous sites were removed and tblastx searches were performed to determine *dlx* protein sequences in the draft cichlid genomes of *Astatotilapia burtoni*, *Neolamprologus brichardi* and *Pundamilia nyererei* (BROAD Institute, unpublished data; see Additional file 1 for scaffold numbers). To determine the rate of evolution for each of the *dlx* proteins, phylogenetic analyses were performed in PAUP\* 4.0 [56] under parsimony settings and the number of amino acid changes was obtained. *D. rerio* or *G. morhua* was used as outgroup species and bootstrap analyses with 100 replicates were conducted to test the robustness of the obtained topologies. Next, the sixty amino acids long homeobox domain was extracted from the sequences and aligned to the homeobox domain of the single *Distal-less* (*Dll*) gene of *Drosophila melanogaster* [Ensemble: FBgn0000157] in Geneious 5.6 [57] for closer inspection of the conservation of the domain and to identify gene-specific substitutions.

#### Cichlid samples and genomic DNA sequencing

White muscle and/or fin clip samples were collected during fieldwork in Zambia in 2007 and 2008 using a standard operating procedure described in [29]. In total 23 Lake Tanganyikan cichlid species were included in this study (Additional file 2). Genomic DNA was extracted following a standard Proteinase K protocol [58]. Cichlid-specific PCR primers were designed based on available and/or draft genomic and transcriptomic cichlid sequences, which were identified by tblastx searches of publicly available *dlx* sequences from other teleost species (see Additional file 1 for species and accession numbers). This was done for eight *dlx* loci: *dlx1a*, *dlx2a*, *dlx3a*, *dlx3b*, *dlx4a*, *dlx4b*, *dlx5a* and *dlx6a* (see Additional file 3 for primer sequences). Standard PCR reactions, purification steps and sequencing reactions were set up and performed as described elsewhere [36]. PCR products of the partially sequenced loci were visualized with GelRed (Biotium) on a 1.5% agarose gel and sequenced on a 3130xl capillary sequencer (Applied Biosystems). Partial sequences were aligned and visually inspected using Codon Code Aligner 3.7.1 (CodonCode Corporation, Dedham, MA). Exon/intron boundaries were determined by homology comparisons with the sequences from the other

teleost species. All generated cichlid *dlx* sequences have been deposited into GenBank [GenBank: KC285366-KC285546] (Additional file 2).

#### Phylogenetic analyses of cichlid samples

Individual gene trees were constructed using maximum likelihood in PAUP\* 4.0 [56] and Bayesian Inference in MrBayes 3.2 [59,60]. The best-fitting model of nucleotide substitution was determined with the corrected Akaike information criteria and likelihood ratio tests conducted in jModeltest 0.1.1 [61,62]. Bootstrap analyses with 100 replicates were performed in PAUP\* and MrBayes was run for 10,500,000 generations. *Oreochromis tanganyicae* was used as outgroup (see e.g., [63]). Phylogenetic analysis of a concatenated dataset of 9.2 kb was performed as described above in PAUP\* to generate a common input tree file (treeBASE submission 14433) for the subsequent analyses.

Coding sequence data of the 23 cichlid species (treeBASE submission 14433) was assessed with both site- and branch-site models as implemented in the program Codeml of the software package PAML (Phylogenetic Analysis by Maximum Likelihood) 4.3 [64,65]. The following parameters were estimated for all eight *dlx* datasets under different models: the nonsynonymous/synonymous substitution rate ratio,  $\omega$ , the proportion of sites assigned to an  $\omega$  category,  $p_{0,1,2}$ , and the  $p$  and  $q$  parameters of the  $\beta$  distribution. Tests of positively selected sites were conducted by performing Likelihood Ratio Tests (LRT) of the following model comparisons: M1a (Nearly Neutral) with M2a (Positive Selection), M7 ( $\beta$ ) with M8 ( $\beta$  &  $\omega_s \geq 1$ ), and M8a ( $\beta$  &  $\omega_s = 1$ ) with M8. The comparison between M0 (one-ratio) and M3 (discrete) was used as a test of variable  $\omega$  among sites. The naive empirical Bayes (NEB; [66,67]) and the Bayes empirical Bayes (BEB; [68]) criteria were used to calculate the posterior probabilities for site classes and the BEB was used to identify sites under positive selection when the LRT was significant. To test whether the *dlx* genes evolved under non-neutral evolution in specific lineages a LRT between the null model ( $\omega_s = 1$ ) and the alternative model ( $\omega_s \geq 1$ ) was performed in the branch-site analyses. Branches of interest, or so-called foreground branches, were chosen based on the results of the phylogenetic analyses and branch tests performed in Hyphy ([69], following [36]).

Additional tests of positive selection on the partial *dlx* sequences were performed with the Sitewise Likelihood Ratio estimation of selection program (SLR; [70]) v1.3. The common input tree file was used (see above) and the significance level was set to 95%.

Amino acid substitutions were screened for possible effect on protein function with the program SIFT (Sorting Intolerant from Tolerant; [71]).

## Results

### Dlx protein sequence comparison in teleost fishes

The tblastx searches of the teleost *dlx* proteins resulted in the retrieval of eight *dlx* genes in all four cichlid species. Furthermore, the genomic locations of these *dlx* loci (Additional file 1) indicate that four *dlx* clusters are present in the cichlid lineage: *dlx1a-dlx2a*; *dlx3a-dlx4a*; *dlx3b-dlx4b* and *dlx5a-dlx6a*. All other teleost species examined contain this full set of genes, except zebrafish, in which *dlx3a* could not be located, and medaka, in which *dlx4b* is missing, as previously noted [35,44,45]. Interestingly, in contrast to Renz *et al.* [35] we do find evidence for the existence of *dlx3a* in cichlids, including *A. burtoni* (Figure 1, Additional file 4).

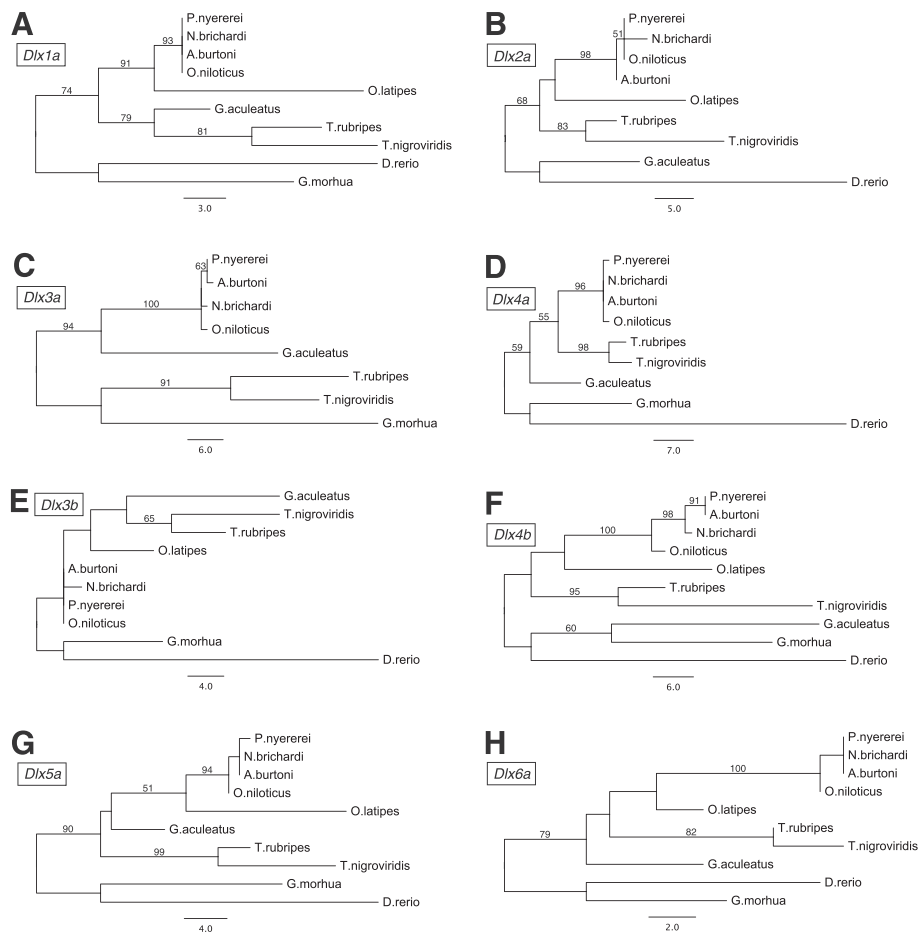
The sixty amino acid long homeobox domain of the eight teleost *dlx* proteins are highly conserved among teleost fish and even between teleosts and the single Dll protein of *D. melanogaster* (Additional file 4). Despite the high level of conservation, several locus-specific amino acid substitutions are present in the paralogs, making it

possible to distinguish between individual *dlx* homeobox domains.

Phylogenetic analyses of the *dlx* protein sequences were performed to examine the rate of evolution of the *dlx* paralogs in teleost fishes. The overall and relative longest trees were found for *dlx4b* and *dlx3a*, while for *dlx1a* and *dlx6a* the shortest tree lengths were observed (Figure 1 and Table 1). Typically the longest branches were observed in the two basal species *D. rerio* and *G. morhua*. Interestingly, relatively long branch lengths for the branch towards the four cichlid species were observed for *dlx3a* and *dlx6a*, indicating elevated rates of molecular evolution. The opposite scenario was observed in the overall more conserved *dlx1a* and *dlx5a* proteins. To study these effects in more detail cichlid specific gene trees were constructed.

### The rate of *dlx* gene evolution in East African cichlid fishes

To reconstruct the molecular evolutionary history of the *dlx* homologs in East African cichlid species, we determined the rate of evolution and the signatures of selection



**Figure 1** Maximum Likelihood phylogenetic hypotheses for the eight *dlx* paralogs in teleost fishes. (A) *Dlx1a* (254 amino acids (aa)). (B) *Dlx2a* (276 aa). (C) *Dlx3a* (307 aa). (D) *Dlx4a* (259 aa). (E) *Dlx3b* (283 aa). (F) *Dlx4b* (257 aa). (G) *Dlx5a* (285 aa). (H) *Dlx6a* (247 aa). Bootstrap probabilities (PAUP\*) above 50% are shown.

**Table 1 Overall and relative tree lengths of teleost protein phylogenies**

dlx paralog	Length (# amino acids)	Tree length (absolute)	Tree length (relative)
dlx1a	254	93	0.366
dlx2a	276	122	0.442
dlx3a	307	179	0.583
dlx3b	283	108	0.382
dlx4a	259	123	0.475
dlx4b	257	222	0.864
dlx5a	285	125	0.439
dlx6a	247	56	0.227

pressure in a phylogenetically representative set of 23 species. The gene trees of the obtained partial cichlid *dlx* sequences resulted in various polytomies (Additional file 5), probably due to the limited size of some of the datasets (minimum of 0.7 kb). Although for each gene tree specific branches were observed with relative long branches, there is not a particular species or clade that has evolved under faster rates of evolution in all of the *dlx* loci examined. Interestingly, three branches have relative long branch lengths in multiple topologies: the branch towards the Lamprologini (*dlx2a*, *dlx4a* and *dlx5a*), *C. leptosoma* (*dlx3b*, *dlx4a* and *dlx5a*) and *C. furcifer* (*dlx1a* and *dlx6a*). The relative tree lengths (Additional file 5 and Table 2) of these gene trees reveal similar results as the teleost protein trees, with *dlx4b* and *dlx3a* evolving fastest and *dlx1a* and *dlx6a* evolving more slowly.

**Observed signatures of selection pressure in cichlid *dlx* loci**

To investigate signatures of selection pressure in the *dlx* loci, we performed detailed analyses of the  $d_N/d_S$  ratios. Maximum likelihood parameter estimations for  $\omega$ ,  $p_{0,1,2}$  and  $p$  and  $q$  under different evolutionary models can be found in Table 3 for all eight *dlx* loci. Estimations of  $\omega$  under the M0 model suggest that the *dlx* genes evolved under purifying selection with  $\omega$  ranging from 0.0001 (*dlx5a*) to 0.457 (*dlx2a*). A small proportion of sites,

**Table 2 Overall and relative tree lengths of cichlid *dlx* gene trees**

dlx paralog	Length (# base pairs)	Tree length (absolute)	Tree length (relative) 10 <sup>-4</sup>
dlx1a	737	0.036	0.483
dlx2a	1371	0.094	0.684
dlx3a	666	0.061	0.910
dlx3b	1972	0.120	0.609
dlx4a	1166	0.104	0.892
dlx4b	722	0.068	0.937
dlx5a	1538	0.093	0.607
dlx6a	1710	0.093	0.542

0.00001-24.2%, was estimated to have evolved neutrally ( $\omega = 1$ ) under the M1a model. By using models that allow  $\omega$  to vary among sites, 0.7-12.3% of sites was detected with  $\omega > 1$  in *dlx1a*, *dlx2a*, *dlx3a*, *dlx4b* and *dlx6a*. Overall, most sites are estimated to have evolved under purifying selection, with highest proportions found in *dlx3b*, *dlx4a* and *dlx5a*.

Likelihood ratio tests of the subsequent model comparisons (Table 4) resulted in the rejection of the null models in only the following comparisons per loci: *dlx1a* (M8a-M8), *dlx2a* (all four comparisons), *dlx3a* (M0-M3; M8a-M8) and *dlx4b* (all four comparisons). Positively selected sites were detected with the BEB in *dlx2a* (5 sites), *dlx3a* (1 site) and *dlx4b* (3 sites; see Table 4, Figure 2). The less constraining analyses with the NEB resulted in two more putative positively selected sites in *dlx1a* (1) and *dlx2a* (1; Figure 2). Fewer positively selected sites were identified by the SLR analyses for *dlx2a* (position: 36; significance: 99%), *dlx3a* (37, 157; 99%, 95%) and *dlx4b* (145; 99%).

None of the performed LTRs of the branch-site analyses were significant ( $1 \geq p \geq 0.20$ ) indicating that although the  $\omega$  ratios do vary among sites (see above), the  $\omega$  ratios do not vary significantly among lineages.

**Amino acid substitutions and their predicted effect on function**

Next, the individual amino acid substitutions were examined in more detail. The total protein length and the number of amino acid substitutions per locus are shown in Table 5 (see also Figure 2 and Additional file 6). A relative large number of substitutions was observed in *dlx2a* (13), *dlx3a* (16) and *dlx4b* (10), while in *dlx5a* no substitution was found. Most of the amino acid substitutions are species-specific (i.e., observed in a single species), although lineage-specific substitutions were observed for the lamprologines (*dlx2a*, *dlx3a*, *dlx4b*), ectodines (*dlx2a*) and haplochromines (*dlx2a*, *dlx3a*). None of the observed amino acid substitutions have a predicted effect on the protein functions (see Table 5), although two substitutions were observed in the homeobox domain of *dlx2a* (Figure 2).

**Selection regimes on the *dlx* clusters**

It is known that the paired members of each of the four *dlx* clusters (Additional file 4) are transcribed concurrently [35,42-45]. To characterize if the members of the same *dlx* cluster evolved at similar rates and under similar selection regimes, we had a closer inspection of these paired genes. First, the teleost *dlx* protein and cichlid gene trees show that overall and relative tree lengths (or the rate of evolution) differ between the two genes within a cluster. Loci with the highest (*dlx3a*: 0.583/0.910 and *dlx4b*: 0.864/0.937) or the smallest (*dlx1a*: 0.366/0.483 and *dlx6a*: 0.227/0.542) tree lengths are never observed

**Table 3 Site model parameter estimates generated by the CodeML analyses for the eight *dlx* paralogs**

Locus	Parameter estimates under different models						
	M0 (one ratio)	M1a (neutral)	M2a (selection)	M3 (discrete)	M7 (β)	M8 (β & ω)	M8a (β & ω <sub>s</sub> = 1)
<i>dlx1a</i>	$\omega = 0.111$	$p_0 = 0.940, \omega_0 = 0$ $p_1 = 0.060, \omega_1 = 1$	$p_0 = 0.980, \omega_0 = 0$ $p_1 = 0, \omega_1 = 1$ <b><math>p_2 = 0.020, \omega_2 = 12.580</math></b>	$p_0 = 0.640, \omega_0 = 0$ $p_1 = 0.340, \omega_1 = 0$ <b><math>p_2 = 0.020, \omega_2 = 12.580</math></b>	$p = 0.005, q = 0.049$	$p = 0.005, q = 22.651$ $p_0 = 0.980$ <b><math>p_1 = 0.020, \omega = 12.580</math></b>	$p = 1.333, q = 24.671$ $p_0 = 0.936$ $p_1 = 0.065, \omega = 1$
<i>dlx2a</i>	$\omega = 0.457$	$p_0 = 0.833, \omega_0 = 0$ $p_1 = 0.167, \omega_1 = 1$	$p_0 = 0.879, \omega_0 = 0$ $p_1 = 0.110, \omega_1 = 1$ <b><math>p_2 = 0.010, \omega_2 = 18.903</math></b>	$p_0 = 0, \omega_0 = 0$ $p_1 = 0.986, \omega_1 = 0.081$ <b><math>p_2 = 0.014, \omega_2 = 16.629</math></b>	$p = 0.005, q = 0.011$	$p = 0.005, q = 0.046$ $p_0 = 0.989$ <b><math>p_1 = 0.011, \omega = 18.053</math></b>	$p = 0.005, q = 12.618$ $p_0 = 0.833$ $p_1 = 0.167, \omega = 1$
<i>dlx3a</i>	$\omega = 0.320$	$p_0 = 0.758, \omega_0 = 0$ $p_1 = 0.242, \omega_1 = 1$	$p_0 = 0.921, \omega_0 = 0$ $p_1 = 0.034, \omega_1 = 1$ <b><math>p_2 = 0.045, \omega_2 = 4.451</math></b>	$p_0 = 0.877, \omega_0 = 0$ <b><math>p_1 = 0.120, \omega_1 = 2.099</math></b> <b><math>p_2 = 0.003, \omega_2 = 11.657</math></b>	$p = 0.005, q = 0.017$	$p = 0.015, q = 0.088$ $p_0 = 0.032$ <b><math>p_1 = 0.032, \omega = 5.096</math></b>	$p = 0.005, q = 13.826$ $p_0 = 0.758$ $p_1 = 0.242, \omega = 1$
<i>dlx3b</i>	$\omega = 0.047$	$p_0 = 1, \omega_0 = 0.047$ $p_1 = 1E-5, \omega_1 = 1$	$p_0 = 1, \omega_0 = 0.047$ $p_1 = 0, \omega_1 = 1$ $p_2 = 0, \omega_2 = 1$	$p_0 = 0.243, \omega_0 = 0.047$ $p_1 = 0.262, \omega_1 = 0.047$ $p_2 = 0.495, \omega_2 = 0.047$	$p = 4.896, q = 99.00$	$p = 4.895, q = 99.00$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$	$p = 4.895, q = 99.00$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$
<i>dlx4a</i>	$\omega = 0.050$	$p_0 = 1, \omega_0 = 0.050$ $p_1 = 1E-5, \omega_1 = 1$	$p_0 = 1, \omega_0 = 0.050$ $p_1 = 0, \omega_1 = 1$ $p_2 = 0, \omega_2 = 1$	$p_0 = 0.216, \omega_0 = 0.050$ $p_1 = 0.384, \omega_1 = 0.050$ $p_2 = 0.400, \omega_2 = 0.050$	$p = 5.267, q = 99.00$	$p = 5.266, q = 99.00$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$	$p = 0.137, q = 2.305$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$
<i>dlx4b</i>	$\omega = 0.259$	$p_0 = 0.872, \omega_0 = 0$ $p_1 = 0.128, \omega_1 = 1$	$p_0 = 0.993, \omega_0 = 0.090$ $p_1 = 0, \omega_1 = 1$ <b><math>p_2 = 0.007, \omega_2 = 12.858</math></b>	$p_0 = 0, \omega_0 = 0$ $p_1 = 0.993, \omega_1 = 0.090$ <b><math>p_2 = 0.007, \omega_2 = 12.858</math></b>	$p = 0.005, q = 0.042$	$p = 9.871, q = 99.00$ $p_0 = 0.993$ <b><math>p_1 = 0.007, \omega = 12.859</math></b>	$p = 0.005, q = 30.817$ $p_0 = 0.872$ $p_1 = 0.128, \omega = 1$
<i>dlx5a</i>	$\omega = 1E-4$	$p_0 = 1, \omega_0 = 0$ $p_1 = 1E-5, \omega_1 = 1$	$p_0 = 1, \omega_0 = 0$ $p_1 = 0, \omega_1 = 1$ $p_2 = 0, \omega_2 = 1$	$p_0 = 0.424, \omega_0 = 0$ $p_1 = 0.420, \omega_1 = 0$ $p_2 = 0.156, \omega_2 = 0$	$p = 0.005, q = 1.294$	$p = 0.005, q = 99.00$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$	$p = 0.005, q = 44.274$ $p_0 = 1$ $p_1 = 1E-5, \omega = 1$
<i>dlx6a</i>	$\omega = 0.029$	$p_0 = 0.967, \omega_0 = 0$ $p_1 = 0.033, \omega_1 = 1$	$p_0 = 0.992, \omega_0 = 0$ $p_1 = 0, \omega_1 = 1$ <b><math>p_2 = 0.008, \omega_2 = 6.336</math></b>	$p_0 = 0.899, \omega_0 = 0$ $p_1 = 0.093, \omega_1 = 0$ <b><math>p_2 = 0.008, \omega_2 = 6.336</math></b>	$p = 0.066, q = 1.696$	$p = 0.005, q = 2.390$ $p_0 = 0.992$ <b><math>p_1 = 0.008, \omega = 6.336</math></b>	$p = 0.005, q = 1.672$ $p_0 = 0.967$ $p_1 = 0.033, \omega = 1$

Note:  $p_{0-2}$  are the proportions of sites assigned to an  $\omega$  category or to a beta distribution with  $p$  and  $q$  as parameters.  $\omega$  ratios greater than one and their corresponding proportions are depicted in bold.

**Table 4 Likelihood ratio test (LRT) statistics of site model comparisons for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b***

Locus	Test	LRT (2ΔI)	p	Selected sites (BEB)
<i>dlx1a</i>	M0 vs M3	8.416	0.077	-
	M1a vs M2a	3.396	ns	-
	M7 vs M8	3.680	ns	-
	M8a vs M8	5.084	0.012	-
<i>dlx2a</i>	M0 vs M3	70.438	<0.001	-
	M1a vs M2a	39.198	<0.001	<b>36A, 48 T, 255I</b>
	M7 vs M8	43.093	<0.001	<b>36A, 48 T, 195A, 254A, 255I</b>
	M8a vs M8	39.168	<0.001	See M7 vs M8 comparison
<i>dlx3a</i>	M0 vs M3	12.605	0.013	-
	M1a vs M2a	3.858	ns	-
	M7 vs M8	4.258	ns	-
	M8a vs M8	3.872	0.025	37S
<i>dlx4b</i>	M0 vs M3	39.110	<0.001	-
	M1a vs M2a	16.940	<0.001	48Q, 135A, <b>145 T</b>
	M7 vs M8	17.367	<0.001	48Q, 135A, <b>145 T</b>
	M8a vs M8	16.931	<0.001	See M7 vs M8 comparison

LRT values, p-values and positively selected sites identified by the BEB ( $p < 0.01$  (in bold) and  $p < 0.05$  (in italic); CodeML) are shown.

within the same cluster. Furthermore, the mode of selection seems to differ between members of the same *dlx* clusters as well. While strong purifying selection was observed for *dlx3b*, *dlx4a* and *dlx5a*, their paired cluster members *dlx4b*, *dlx3a* and *dlx6a* show sign of elevated  $\omega$ -values. A notable exception to this observation is the *dlx1a-dlx2a* cluster. For both genes a proportion of sites was found with elevated  $\omega$ -values (note that the proportion is considerably bigger for *dlx2a*). These observations indicate that although clusters are transcribed concurrently, selection seems to act on the individual gene level rather than on the level of the *dlx* gene clusters. Also the observed patterns are not in concordance with the two groups of homeobox domains that emerged from the initial tandem duplication (see [52] and Additional file 4).

## Discussion

In this work, we present a detailed evolutionary characterization of the *dlx* gene repertoire in East African cichlid fishes. Previously, Renz *et al.* [35] studied the embryonic expression patterns of *dlx* genes in cichlids and showed that they are expressed in e.g., the developing jaw apparatus and anal fin, tissues that contribute to two putative evolutionary innovations: the pharyngeal jaw and the egg-spots on the anal fin of the cichlid *A. burtoni*. Here, we study the molecular evolution of *dlx* genes in a representative set of 23 East African cichlid species. We performed comparative phylogenetic analyses and detailed screens of nonsynonymous-synonymous substitution rate ratios to determine the selective pressure acting

upon these candidate genes for evolutionary novelties in cichlid fishes.

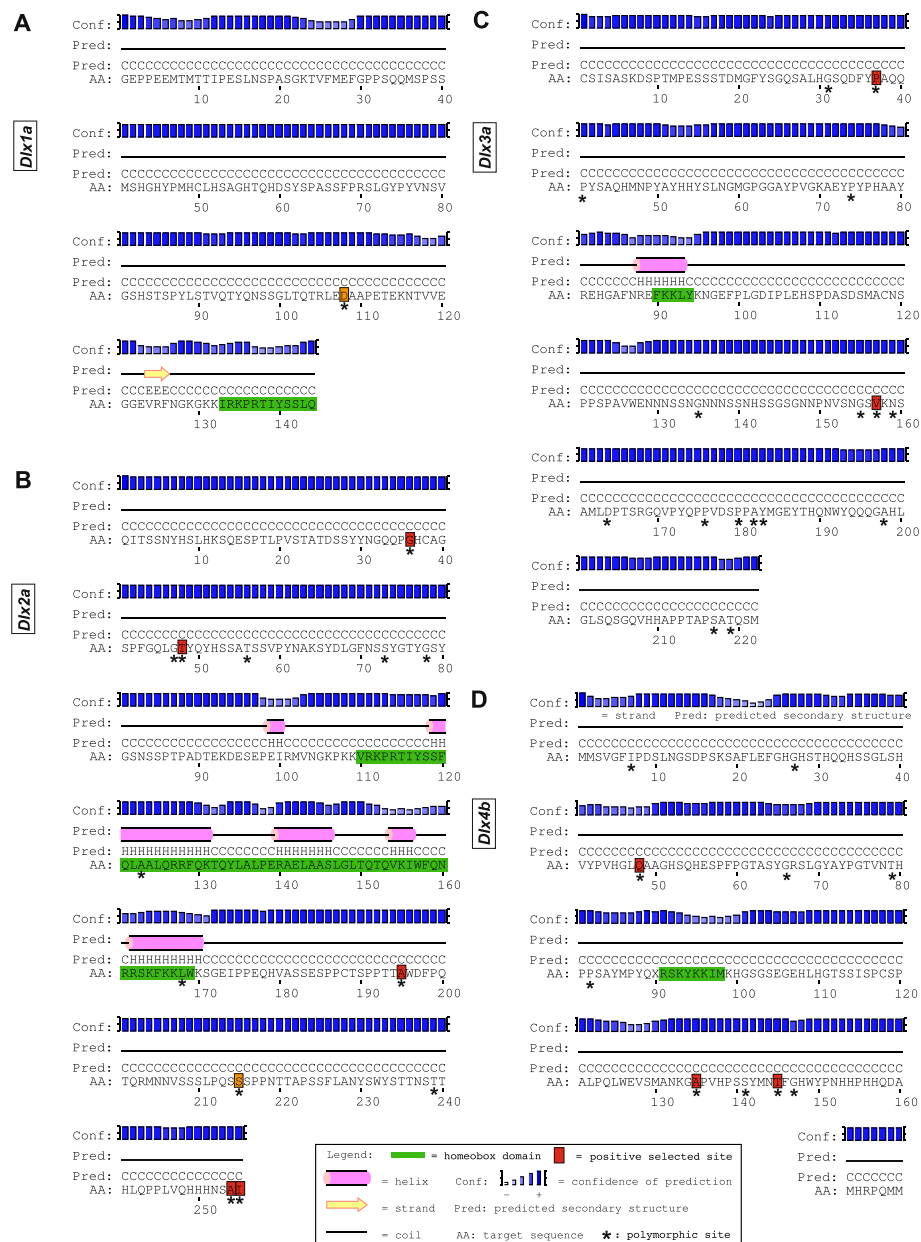
### *Dlx3a* did not get lost in the cichlid lineage

Our phylogenetic analyses of *dlx* proteins extends previous analyses (e.g., [35]) by the inclusion of cod [72] and four different cichlid species (i.e., *O. niloticus*, *N. brichardi*, *A. burtoni* and *P. nyererei*; BROAD Institute). Although our results agree with most of the available hypotheses on the evolutionary loss of *dlx* genes in specific teleost lineages (i.e., *dlx3a* in zebrafish and *dlx4a* in medaka), we did detect *dlx3a* in cichlids and thus refute the cichlid-specific gene loss hypothesis of *dlx3a* put forward by Renz *et al.* [35]. Not only were we able to locate this gene in all four cichlid genomes examined (Additional file 1), we also gathered partial gene sequences for this locus in all 23 cichlid species included (Additional files 4 and 6). Furthermore, in-house tblastx searches of this newly identified paralog against preliminary cichlid EST libraries (BROAD Institute, unpublished data) resulted in multiple hits, providing proof of its expression in – at least – *Astatotilapia burtoni*, *Oreochromis niloticus* and *Metriaclima zebra*.

### Selection on *dlx* paralogs in relation to gene duplication events

Gene-wide estimates of the  $d_N/d_S$  ratios indicate that all loci evolved under purifying selection ( $\omega < 1$ ), indicating strong selection against deleterious mutations, commonly observed in functional proteins. Additional analyses of individual codons indicate that the sequenced regions of *dlx3b*, *dlx4a* and *dlx5a* evolved under purifying selection, while positive selection acting on specific codons was detected for a small proportion of sites (i.e., up to 12%) for *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b* (i.e., a smaller number of positively selected sites was found with more stringent SLR analyses for *dlx2a*, *dlx3a* and *dlx4b*). Plausible reasons for the excess of nonsynonymous mutations in these loci are either lowered functional constraints or directional selection, as Sumiyama and colleagues suggested for *Dlx7* in mouse [73]. Different modes of selection are thus found to have acted on the *dlx* paralogs in cichlids after the genome duplication events.

Differential selection after gen(om)e duplication is a commonly observed phenomenon and is associated with the fate of the gene duplicates i.e., non-, sub- or neofunctionalization. Sub- and neofunctionalization are adaptive processes by which either spatial or temporal partitioning of the ancestral function or the evolution of complete new functions take place [5,74-76]. While ancestral functions can be maintained by retaining the protein sequences and preventing deleterious mutations through purifying selection, relaxed selection on the other duplicate can lead to the introduction of mutations and subsequent divergence



**Figure 2 Secondary structure and positively selected sites for four partially sequenced *Astatotilapia burtoni* Dlx proteins.** Secondary structure predictions were obtained from the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>). Positively selected sites identified by the site model analyses (CodeML) and the SLR analyses are highlighted in red (BEB and/or SLR) or orange (NEB) boxes. (A) Dlx1a. (B) Dlx2a. (C) Dlx3a. (D) Dlx4b.

[5,75,76]. Most of these changes are deleterious and are followed by the loss of the gene over time (i.e., nonfunctionalization). On rare occasions the mutations can lead to an altered function of the protein (i.e., neofunctionalization; change within the protein) or altered expression pattern (subfunctionalization; change in regulatory regions), which can be characterized by elevated  $\omega$  values and the maintenance of the mutations results in divergence of the two duplicates.

Many studies have focused on duplicated genes in relation to divergence of duplicates (see e.g., [77-80] and references therein). An interesting case of subfunctionalization was described in leaf-eating Colobine monkeys, in which the pancreatic ribonuclease gene (*RNASE1*), necessary to digest its specialized diet, was duplicated [81,82]. Although the two gene-products are used in the same process (i.e., digestion of bacterial RNA), the duplicate gene shows many substitutions, while the ancestral locus



**Table 5 Amino acid substitution and their predicted effect on function for the eight cichlid *dlx* loci**

Locus	L	S	E
<i>dlx1a</i>	144	1	Tolerated
<i>dlx2a</i>	255	13	Tolerated
<i>dlx3a</i>	222	16	Tolerated
<i>dlx3b</i>	160	4	Tolerated
<i>dlx4a</i>	96	1	Tolerated
<i>dlx4b</i>	167	10	Tolerated
<i>dlx5a</i>	271	0	-
<i>dlx6a</i>	236	2	Tolerated

Length of obtained protein sequence (L), the number of amino acid substitutions (S) and the predicted effect of the substitutions on the function generated by the SIFT analyses (E) are shown.

did not change [81]. Similar patterns of heterogeneity in amino acid substitutions or differential selection were also observed by Dermitzakis and Clark [83] between duplicates of several developmental gene families (e.g., *Notch*, *Bmp* and *Hox9*) in mouse and human. Interestingly, differential selection regimes acting on paralogs were also found in the murine *Dlx3-Dlx7* cluster, with *Dlx7* evolving more rapidly than *Dlx3* [73]. Our results of differential selection acting on the cichlid *dlx* paralogs are thus comparable to previously studied cases of duplicated genes. We even detect a similar pattern as Sumiyama *et al.* [73], with *dlx4b* evolving more rapidly than *dlx3b* (i.e., relative tree length 0.937 vs 0.609).

The adaptive protein evolution as observed in *dlx1a*, *dlx2a*, *dlx3a* and *dlx4b* together with the evolutionary history of the gene family, could thus be a sign of possible new or altered functions of these *dlx* paralogs in cichlids. Although we did not observe amino acid substitutions with predicted apparent effect on the protein function in our partial sequences, other mechanisms, such as *cis*-regulatory evolution might have altered the expression patterns after gene duplication. Gene expression analyses in cichlid and zebrafish indicate that clusters are often transcribed concurrently and that the *dlx* duplicates exhibit overlapping expression patterns in particular during the development of brain and pharyngeal arches [35,44,45]. This co-expression of the *dlx* clusters is controlled through intergenic *cis*-regulatory regions [35,42,43]. While mutations in these regions are expected to affect the expression of both paralogs, changes in the coding regions of the *dlx* loci are likely to affect the individual *dlx* locus' function, which could lead to neofunctionalization.

#### Selection pressure on *dlx* paralogs in relation to evolutionary innovations

We found an interesting pattern comparing our  $d_N/d_S$  results with the expression patterns found by Renz *et al.*

[35] in relation to evolutionary novelties in cichlids. In the developing pharyngeal teeth and the anal fin *dlx3b*, *dlx4a* (not in anal fin) and *dlx5a*, the exact loci for which we found strong patterns of purifying selection, are expressed at high levels. Although this observation seems to contradict other cases in which candidate genes showed accelerated rates of protein evolution (see [31,53,84]), they do not stand alone (see e.g., [36]). It has been shown that minor changes in the complex genetic pathways underlying the development of morphological structures can lead to the evolution of novelties (see e.g., [85]). Furthermore, many cases of morphological adaptation are driven by *cis*-regulatory evolution (reviewed in [86]). Several intergenic *cis*-regulatory elements have been identified in the *dlx* clusters in *A. burtoni* by Renz *et al.* [35], but the functional characterization in cichlids is yet to be performed. It is thus possible that only a small fraction of genes involved in the evolutionary novelties in cichlids show signs of adaptive evolution and that the three *dlx* loci were co-opted for their ancestral functions.

According to Renz *et al.* [35], the five *dlx* genes for which we found signatures of positive selection, are either not expressed at all or at low levels during pharyngeal teeth and anal fin development in the cichlid *A. burtoni*. Low levels of gene expression were observed for *dlx2a* in the developing pharyngeal teeth in cichlids [35], while higher *dlx2a* expression levels were observed in other teleost species [33,44,45]. *Dlx4b* and *dlx6a* expression has previously been shown in the developing pharyngeal teeth of zebrafish and/or medaka [44,45], but has not been observed in cichlids (yet). Furthermore, multiple *dlx* genes, including loci with signatures of positive selection, appear to be expressed in the developing anal fin tissue at time points coinciding with egg-spot development in *A. burtoni* (E. Santos, personal communications). Therefore, it is likely that several *dlx* paralogs, for which we found signs of positive selection, are involved in the development of evolutionary innovations in cichlids, in contrast to the initial findings of Renz *et al.* [35]. Future detailed and extended functional analyses should be conducted to elucidate their role in the development of these evolutionary important traits in cichlid fishes.

#### Conclusions

In this study, we provide an in depth molecular evolutionary analysis of the *dlx* gene repertoire in teleost fishes. We located and generated partial sequences for *dlx3a* in 23 East African cichlid species, refuting the hypothesis of Renz *et al.* [35] that *dlx3a* got lost in the cichlid lineage. Phylogenetic analyses of the teleost *dlx* gene repertoire show that substantial differences exist in the rate of evolution among teleost *dlx* paralogs. In addition, analyses of the nonsynonymous-synonymous

substitution rates of the cichlid *dlx* paralogs revealed strong differences in the selection pressure acting upon *dlx* paralogs and cluster members. Although differential selection pressure after gene duplication is a putative sign of new or altered functions, we observed a link between the *dlx* loci under strong purifying selection, in particular, and high expression levels in two cichlids' novelties; the pharyngeal jaw and anal fin. This indicates that other mechanisms than adaptive protein evolution are likely to be involved in the co-option of these genes. Furthermore, several (preliminary) studies found that at least three other *dlx* paralogs, for which we found signs of positive selection, are actually expressed in the developing pharyngeal teeth and/or haplochromine anal fin. Hence, the *dlx* paralogs appear as candidate genes for the development of evolutionary innovations in cichlids, although further functional analyses should elucidate the role of positive selection therein.

### Availability of supporting data

The datasets supporting the results of this article are publicly available in the GenBank repository under accession numbers: KC285366-KC285546 and in the treeBASE repository under submission number 14433, <http://purl.org/phylo/treebase/phyloids/study/TB2:S14433>.

### Additional files

**Additional file 1: Accession numbers and/or genomic location of the teleost *dlx* sequences.**

**Additional file 2: Specimen information and GenBank Accession numbers.**

**Additional file 3: Primer information and primer sequences.**

**Additional file 4: Protein comparison of the teleost *dlx* homeobox domains.** Depicted are the amino acid sequences of the homeobox domains for each of the four teleost clusters: *dlx1a-dlx2a*, *dlx4a-dlx3a*, *dlx4b-dlx3b* and *dlx6a-dlx5a* in comparison with the single Dll homeobox sequence (here depicted in duplo) of *Drosophila melanogaster*. Sequences can be divided in two groups; *dlx1a*, *dlx4a*, *dlx4b* and *dlx6a* versus *dlx2a*, *dlx3a*, *dlx3b* and *dlx5a*. The two sixty amino acid long homeobox domains of each cluster are depicted in separate boxes. The top graph displays the mean pairwise identity of all sequences (i.e., green = 100% identity and brown  $\geq$  30% identity). Numbers represent the amino acid position within the homeobox.

**Additional file 5: Maximum likelihood gene trees based on 23 cichlid species for the eight *dlx* loci.** Bootstrap values (PAUP\*) and Bayesian posterior probabilities (MrBayes) above 50% are shown respectively above and below the branches. A color key for the ten studied cichlid lineages is given in the box below the figure. (a) *Dlx1a* (737 base pairs (bp); TPM3uf model). Two major polytomies were recovered. The lamprologines cluster together with the Boulengerochromini, Bathybatini and the Cyphotilapiini. *A. burtoni* is found at the base with *O. tanganyicae*. (b) *Dlx2a* (1371 bp; HKY + I model). Polytomous tree with all members of the lineages Lamprologines, Ectodines, Haplochromines and Limnchromines recovered as monophyletic clades. (c) *Dlx3a* (666 bp; HKY model). Polytomous tree, with only the Lamprologines recovered as monophyletic clade. (d) *Dlx4a* (1166 bp; TPM3uf + I + G). Polytomous relationships were observed between multiple lineages, although most lineages are monophyletic except the Haplochromines (e) *Dlx3b* (1972 bp; GTR + I + G). Moderately

resolved tree. (f) *Dlx4b* (722 bp; TPM3uf). Mostly polytomous relationships between species, except the Limnchromini and most members of the Lamprologines. (g) *Dlx5a* (1538 bp; TIM2 + G). Basal polytomy divides ingroup species except *G. permaxillaris*, in two big clades. (h) *Dlx6a* (1710 bp; TIM3 + G). Limnchromines, Lamprologines and Haplochromines recovered as monophyletic clades, although the relationships between lineages are largely polytomous.

#### Additional file 6: Four partially sequenced cichlid *Dlx* proteins.

Depicted are the amino acid sequences of *Astatotilapia burtoni* (a, c, d) and *Ctenochromis horei* (b). Secondary structure predictions were obtained from the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>). (a) *Dlx3b*. (b) *Dlx4a*. (c) *Dlx5a*. (d) *Dlx6a*.

### Competing interest

The authors declare that they have no competing interests.

### Authors' contributions

ETD, FDK and WS conceived the study. FDK generated the data. ETD and FDK analyzed the data. ETD and WS wrote the paper. All authors read and approved the final manuscript.

### Authors' information

ETD is a PhD student and FDK a master student in the group of WS. WS is a Professor of Zoology and Evolutionary Biology at the University of Basel. The research of his team focuses on the genetic basis of adaptation, evolutionary innovation and animal diversification of mainly the exceptionally diverse cichlid fishes.

### Acknowledgements

We would like to express our gratitude to past and current members of the Salzburger lab for their contribution to sampling during fieldwork; to Brigitte Aeschbach and Nicolas Boileau for assistance during lab work; to Britta Meyer for advice on the phylogenetic analyses; to Emília Santos for help on designing the study and comments on earlier drafts of this manuscript; and to Richard Kluin for grammatical advice. We would also like to thank the BROAD institute for sharing unpublished cichlid genome sequence data with the community. The valuable suggestions of two anonymous reviewers greatly helped improving this manuscript. This study was supported by the Freiwillige Akademische Gesellschaft Basel (Dissertation support grant to ETD), European Research Council (Starting Grant "INTERGENADAPT" to WS) and the Swiss National Science Foundation (Grant 3100A0\_122458 to WS).

Received: 9 May 2013 Accepted: 5 July 2013

Published: 17 July 2013

### References

1. Meyer A, Van de Peer Y: **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27**:937–945.
2. Volff JN: **Genome evolution and biodiversity in teleost fish.** *Heredity* 2005, **94**:280–294.
3. Taylor JS, Braasch I, Frickey T, Meyer A, Van de Peer Y: **Genome duplication, a trait shared by 22,000 species of ray-finned fish.** *Genome Res* 2003, **13**:382–390.
4. Taylor JS, Van de Peer Y, Braasch I, Meyer A: **Comparative genomics provides evidence for an ancient genome duplication event in fish.** *Phil Trans R Soc Lond B* 2001, **356**:1661–1679.
5. Ohno S: *Evolution by gene duplication.* New York: Springer Verlag; 1970.
6. Santini F, Harmon LJ, Carnevale G, Alfaro ME: **Did genome duplication drive the origin of teleosts? A comparative study of diversification in ray-finned fishes.** *BMC Evol Biol* 2009, **9**:194.
7. De Rosa R, Grenier JK, Andreevas T, Cook CE, Adoutte A, Akam M, Carroll SB, Balavoine G: **Hox genes in brachiopods and priapulids and protostome evolution.** *Nature* 1999, **399**:772–776.
8. Lemons D, McGinnis W: **Genomic evolution of Hox gene clusters.** *Science* 2006, **313**:1918–1922.
9. Ruddle FH, Bartels JL, Bentley KL, Kappen C, Murtha MT, Pendleton JW: **Evolution of Hox genes.** *Annu Rev Genet* 1994, **28**:423–442.
10. Gehring WJ, Hiromi Y: **Homeotic genes and the homeobox.** *Annu Rev Genet* 1986, **20**:147–173.

11. Cohn MJ, Tickle C: **Limbs: a model for pattern formation within the vertebrate body plan.** *Trends Genet* 1996, **12**:253–257.
12. Zakany J, Duboule D: **The role of Hox genes during vertebrate limb development.** *Curr Opin Genet Dev* 2007, **17**:359–366.
13. Weatherbee SD, Nijhout HF, Grunert LW, Halder G, Galant R, Selegue J, Carroll SB: **Ultrabithorax function in butterfly wings and the evolution of insect wing patterns.** *Curr Biol* 1999, **9**:109–115.
14. Warren RW, Nagy L, Selegue J, Gates J, Carroll SB: **Evolution of homeotic gene regulation and function in flies and butterflies.** *Nature* 1994, **372**:458–461.
15. Pick L, Heffer A: **Hox gene evolution: multiple mechanisms contributing to evolutionary novelties.** *Ann N Y Acad Sci* 2012, **1256**:15–32.
16. Pearson JC, Lemons D, McGinnis W: **Modulating Hox gene functions during animal body patterning.** *Nat Rev Genet* 2005, **6**:893–904.
17. Seehausen O: **African cichlid fish: a model system in adaptive radiation research.** *Proc R Soc B* 2006, **273**:1987–1998.
18. Salzburger W: **The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes.** *Mol Ecol* 2009, **18**:169–185.
19. Kornfield I, Smith PF: **African cichlid fishes: model systems for evolutionary biology.** *Annu Rev Ecol Syst* 2000, **31**:163–196.
20. Kocher TD: **Adaptive evolution and explosive speciation: the cichlid fish model.** *Nat Rev Genet* 2004, **5**:288–298.
21. Santos ME, Salzburger W: **How cichlids diversify.** *Science* 2012, **338**:619–621.
22. Turner GF, Seehausen O, Knight ME, Allender CJ, Robinson RL: **How many species of cichlid fishes are there in African lakes?** *Mol Ecol* 2001, **10**:793–806.
23. Genner MJ, Seehausen O, Lunt DH, Joyce DA, Shaw PW, Carvalho GR, Turner GF: **Age of cichlids: new dates for ancient lake fish radiations.** *Mol Biol Evol* 2007, **24**:1269–1282.
24. Verheyen E, Salzburger W, Snoeks J, Meyer A: **Origin of the superclade of cichlid fishes from Lake Victoria, East Africa.** *Science* 2003, **300**:325–329.
25. Barlow GW: *The cichlid fishes: nature's grand experiment in evolution.* Cambridge: Perseus publishing; 2000.
26. Coulter GW: *Lake Tanganyika and its life.* Oxford: British Museum (Natural History) and Oxford University Press; 1991.
27. Fryer G, Iles TD: *The cichlid fishes of the Great Lakes of Africa: their biology and evolution.* Edinburgh: Oliver & Boyd; 1972:1–334.
28. Liem KF: **Evolutionary strategies and morphological innovations: Cichlid pharyngeal jaws.** *Systematic Zoology* 1973, **22**:425–441.
29. Muschick M, Indermaur A, Salzburger W: **Convergent evolution within an adaptive radiation of cichlid fishes.** *Curr Biol* 2012, **22**:2362–2368.
30. 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.
31. 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.
32. 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.
33. Fraser GJ, Hulsey CD, Bloomquist RF, Uyesugi K, Manley NR, Strelman JT: **An ancient gene network is co-opted for teeth on old and new jaws.** *Plos Biol* 2009, **7**:e31.
34. Fraser GJ, Bloomquist RF, Strelman JT: **A periodic pattern generator for dental diversity.** *BMC Biol* 2008, **6**:32.
35. Renz AJ, Gunter HM, Fischer JM, Qiu H, Meyer A, Kuraku S: **Ancestral and derived attributes of the dlx gene repertoire, cluster structure and expression patterns in an African cichlid fish.** *EvoDevo* 2011, **2**:1.
36. Diepeveen ET, Salzburger W: **Molecular characterization of two endothelin pathways in East African cichlid fishes.** *J Mol Evol* 2011, **73**:355–368.
37. Kuraku S, Meyer A: **Genomic analysis of cichlid fish 'natural mutants'.** *Curr Opin Genet Dev* 2008, **18**:551–558.
38. Parichy DM, Ransom DG, Paw B, Zon LI, Johnson SL: **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**:3031–3044.
39. Parichy DM, Turner JM: **Temporal and cellular requirements for Fms signaling during zebrafish adult pigment pattern development.** *Development* 2003, **130**:817–833.
40. Albertson RC, Strelman JT, Kocher TD, Yelick PC: **Integration and evolution of the cichlid mandible: the molecular basis of alternate feeding strategies.** *P Natl Acad Sci Usa* 2005, **102**:16287–16292.
41. Hoegg S, Boore JL, Kuehl JV, Meyer A: **Comparative phylogenomic analyses of teleost fish Hox gene clusters: lessons from the cichlid fish *Astatotilapia burtoni*.** *BMC Genomics* 2007, **8**:317.
42. Panganiban G, Rubenstein J: **Developmental functions of the *Distal-less/Dlx* homeobox genes.** *Development* 2002, **129**:4371–4386.
43. Ellies DL, Stock DW, Hatch G, Giroux G, Weiss KM, Ekker M: **Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes.** *Genomics* 1997, **45**:580–590.
44. Borday-Birraux V, Van der Heyden C, Debais-Thibaud M, Verreijdt L, Stock DW, Huysseune A, Sire J-Y: **Expression of *Dlx* genes during the development of the zebrafish pharyngeal dentition: evolutionary implications.** *Evol Dev* 2006, **8**:130–141.
45. Debais-Thibaud M, Germon I, Laurenti P, Casane D, Borday-Birraux V: **Low divergence in *Dlx* gene expression between dentitions of the medaka (*Oryzias latipes*) versus high level of expression shuffling in osteichthyan.** *Evol Dev* 2008, **10**:464–476.
46. Beldade P, Brakefield PM, Long AD: **Contribution of *Distal-less* to quantitative variation in butterfly eyespots.** *Nature* 2002, **415**:315–318.
47. Reed RD, Serfas MS: **Butterfly wing pattern evolution is associated with changes in a Notch/*Distal-less* temporal pattern formation process.** *Curr Biol* 2004, **14**:1159–1166.
48. Brakefield PM, Gates J, Keys D, Kesbeke F, Wijngaarden PJ, Monteiro A, French V, Carroll SB: **Development, plasticity and evolution of butterfly eyespot patterns.** *Nature* 1996, **384**:236–242.
49. Sunkel CE, Whittle J: ***Brista*: A gene involved in the specification and differentiation of distal cephalic and thoracic structures in *Drosophila melanogaster*.** *Roux's Arch Dev Biol* 1987, **196**:124–132.
50. Dong P, Chu J, Panganiban G: **Coexpression of the homeobox genes *Distal-less* and *homothorax* determines *Drosophila* antennal identity.** *Development* 2000, **127**:209–216.
51. Gordon CT, Brinas I, Rodda FA, Bendall AJ, Farlie PG: **Role of *Dlx* genes in craniofacial morphogenesis: *Dlx2* influences skeletal patterning by inducing ectomesenchymal aggregation in ovo.** *Evol Dev* 2010, **12**:459–473.
52. Stock DW, Ellies DL, Zhao ZY, Ekker M, Ruddle FH, Weiss KM: **The evolution of the vertebrate *Dlx* gene family.** *P Natl Acad Sci Usa* 1996, **93**:10858–10863.
53. Terai Y, Morikawa N, Okada N: **The evolution of the pro-domain of bone morphogenetic protein 4 (*Bmp4*) in an explosively speciated lineage of East African cichlid fishes.** *Mol Biol Evol* 2002, **19**:1628–1632.
54. Notredame C, Higgins DG, Heringa J: **T-Coffee: a novel method for fast and accurate multiple sequence alignment.** *J Mol Biol* 2000, **302**:205–217.
55. Poirot O, O'Toole E: **Tcoffee@igs: a web server for computing, evaluating and combining multiple sequence alignments.** *Nucleic Acids Res* 2003, **31**:3503–3506.
56. Swofford DL: *Phylogenetic Analysis Using Parsimony (\*and Other Methods).* Sunderland: Sinauer Associates; 2003.
57. *Geneious version 5.6 created by Biomatters.* Available from <http://www.geneious.com/>.
58. Bruford MW, Hanotte O, Brookfield J, Burke T: **Multilocus and single-locus DNA fingerprinting.** In *Molecular genetic analysis of populations: a practical approach.* Edited by Hoelzel AR. Oxford: Oxford University Press; 1998:287–336.
59. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2001, **17**:754–755.
60. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19**:1572–1574.
61. Guindon S, Gascuel O: **A simple, fast, and accurate algorithm to estimate large phylogenies by maximum-likelihood.** *Syst Biol* 2003, **52**:696–704.
62. Posada D: **jModelTest: phylogenetic model averaging.** *Mol Biol Evol* 2008, **25**:1253–1256.
63. Salzburger W, Meyer A, Baric S, Verheyen E, Sturmbauer C: **Phylogeny of the Lake Tanganyika cichlid species flock and its relationship to the Central and East African haplochromine cichlid fish faunas.** *Systematic Biology* 2002, **51**:113–135.
64. Yang Z: **PAML, A program package for phylogenetic analysis by maximum likelihood.** *Comput Appl Biosci* 1997, **13**:555–556.
65. Yang Z: **PAML 4: phylogenetic analysis by maximum likelihood.** *Mol Biol Evol* 2007, **24**:1586–1591.
66. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**:929–936.

67. Yang Z, Nielsen R, Goldman N: **Codon- substitution models for heterogeneous selection pressure at amino acid sites.** *Genetics* 2000, **155**:431–449.
68. Yang Z, Wong W, Nielsen R: **Bayes empirical Bayes inference of amino acid sites under positive selection.** *Mol Biol Evol* 2005, **22**:1107–1118.
69. Kosakovsky Pond SL, Frost S, Muse SV: **HyPhy: hypothesis testing using phylogenies.** *Bioinformatics* 2005, **22**:676–679.
70. Massingham T, Goldman N: **Detecting amino acid sites under positive selection and purifying selection.** *Genetics* 2005, **169**:1753–1762.
71. Ng P, Henikoff ST: **SIFT: predicting amino acid changes that affect protein functions.** *Nucleic Acids Research* 2003, **31**:3812–3814.
72. Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrøm M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A, Wetten OF, Lanzén A, Winer R, Knight J, Vogel J-H, Aken B, Andersen Ø, Lagesen K, Tooming-Klunderud A, Edvardsen RB, Tina KG, Espelund M, Nepal C, Previti C, Karlsen BO, Mourm T, Skage M, Berg PR, Gjøen T, Kuhl H, *et al*: **The genome sequence of Atlantic cod reveals a unique immune system.** *Nature* 2011, **477**:207–210.
73. Sumiyama K, Irvine SQ, Stock DW, Weiss KM, Kawasaki K, Shimizu N, Shashikant CS, Miller W, Ruddle FH: **Genomic structure and functional control of the *Dlx3-7* bigene cluster.** *P Natl Acad Sci Usa* 2002, **99**:780–785.
74. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J: **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531–1545.
75. Sidow A: **Gen(om)e duplications in the evolution of early vertebrates.** *Curr Opin Genet Dev* 1996, **6**:715–722.
76. Ohta T: **Simulating evolution by gene duplication.** *Genetics* 1987, **115**:207–213.
77. Marotta M, Piontkivska H, Tanaka H: **Molecular trajectories leading to the alternative fates of duplicate genes.** *PLoS ONE* 2012, **7**:e38958.
78. Conrad B, Antonarakis SE: **Gene duplication: a drive for phenotypic diversity and cause of human disease.** *Annu Rev Genom Human Genet* 2007, **8**:17–35.
79. Prince VE, Pickett FB: **Splitting pairs: the diverging fates of duplicated genes.** *Nature Reviews Genetics* 2002, **3**:827–837.
80. Braasch I, Volf JN, Scharl M: **The endothelin system: evolution of vertebrate-specific ligand-receptor interactions by three rounds of genome duplication.** *Mol Biol Evol* 2009, **26**:783–799.
81. Zhang J, Zhang Y-P, Rosenberg HF: **Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey.** *Nat Genet* 2002, **30**:411–415.
82. Zhang J: **Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys.** *Nat Genet* 2006, **38**:819–823.
83. Dermitzakis ET, Clark AG: **Differential selection after duplication in mammalian developmental genes.** *Mol Biol Evol* 2001, **18**:557–562.
84. 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 fishes.** *Mol Biol Evol* 2002, **19**:574–578.
85. Wagner GP, Lynch VJ: **Evolutionary novelties.** *Current Biology* 2010, **20**:R48–R52.
86. Prud'homme B, Gompel N, Carroll SB: **Emerging principles of regulatory evolution.** *P Natl Acad Sci Usa* 2007, **104**:8605–8612.

doi:10.1186/1471-2148-13-153

**Cite this article as:** Diepeveen *et al.*: Sequence analyses of the *distal-less homeobox* gene family in East African cichlid fishes reveal signatures of positive selection. *BMC Evolutionary Biology* 2013 **13**:153.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

