Heterogeneous Conservation of Dlx Paralog Co-Expression in Jawed Vertebrates
Mélanie Debiais-Thibaud, Cushla J. Metcalfe, Jacob Pollack, Isabelle Germon, Marc Ekker, Michael Depew, Patrick Laurenti, Véronique Borday-Birraux, Didier Casane

To cite this version:
Mélanie Debiais-Thibaud, Cushla J. Metcalfe, Jacob Pollack, Isabelle Germon, Marc Ekker, et al.. Heterogeneous Conservation of Dlx Paralog Co-Expression in Jawed Vertebrates. PLoS ONE, 2013, 8 (6), pp.e68182. 10.1371/journal.pone.0068182. hal-01622687

HAL Id: hal-01622687
https://hal.umontpellier.fr/hal-01622687
Submitted on 26 Oct 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
Heterogeneous Conservation of Dlx Paralog Co-Expression in Jawed Vertebrates

Mélanie Debiais-Thibaud1, Cushla J. Metcalfe2*, Jacob Pollack3, Isabelle Germon2, Marc Ekker3, Michael Depew4, Patrick Laurenti2,5, Véronique Borday-Birraux2,5*, Didier Casane2,5*

1 Institut des Sciences de l’Evolution, Université de Montpellier II, UMR5554, Montpellier, France, 2 Laboratoire Évolution Génome et Spéciation UPR9034 CNRS, Gif-sur-Yvette, France, 3 Center for Advanced Research in Environmental Genomics, University of Ottawa, Ottawa, Canada, 4 Department of Orthopaedic Surgery, University of California San Francisco, San Francisco, California, United States of America, 5 Université Paris Diderot, Paris, France

Abstract

Background: The Dlx gene family encodes transcription factors involved in the development of a wide variety of morphological innovations that first evolved at the origins of vertebrates or of the jawed vertebrates. This gene family expanded with the two rounds of genome duplications that occurred before jawed vertebrates diversified. It includes at least three bigene pairs sharing conserved regulatory sequences in tetrapods and teleost fish, but has been only partially characterized in chondrichthyans, the third major group of jawed vertebrates. Here we take advantage of developmental and molecular tools applied to the shark Scyliorhinus canicula to fill in the gap and provide an overview of the evolution of the Dlx family in the jawed vertebrates. These results are analyzed in the theoretical framework of the DDC (Duplication-Degeneration-Complementation) model.

Results: The genomic organisation of the catshark Dlx genes is similar to that previously described for tetrapods. Conserved non-coding elements identified in bony fish were also identified in catshark Dlx clusters and showed regulatory activity in transgenic zebrafish. Gene expression patterns in the catshark showed that there are some expression sites with high conservation of the expressed paralog(s) and other expression sites with events of paralog sub-functionalization during jawed vertebrate diversification, resulting in a wide variety of evolutionary scenarios within this gene family.

Conclusion: Dlx gene expression patterns in the catshark show that there has been little neo-functionalization in Dlx genes over gnathostome evolution. In most cases, one tandem duplication and two rounds of vertebrate genome duplication have led to at least six Dlx coding sequences with redundant expression patterns followed by some instances of paralog sub-functionalization. Regulatory constraints such as shared enhancers, and functional constraints including gene pleiotropy, may have contributed to the evolutionary inertia leading to high redundancy between gene expression patterns.

Introduction

The Osteichthyan Dlx Gene Family

Dlx genes encode a family of homeodomain transcription factors with various roles in embryogenesis, notably in many shared derived characters (synapomorphies) that evolved with the diversification of vertebrates [1]. This gene family displays a conserved genomic organization in jawed vertebrates with the clustering of Dlx1 with Dlx2, Dlx3 with Dlx4 and Dlx5 with Dlx6. The mouse is the best studied model organism in terms of the functional analysis of this family of transcription factors, with a series of mutants showing that they are often redundant. Further studies in the zebrafish have shown that most of their roles are conserved in bony vertebrates [reviewed in [2]]. The earliest expression of Dlx genes is found in the non-neural ectoderm, including the preplacodal region, in early neurula [3–6]. Dlx genes are expressed in pre-migrating neural crest cells [5,7] and paired sensory placodes [5,6,8,9] (neural crest and sensory placodes are vertebrate synapomorphies), as well as in some migrating neural crest cells streams giving rise to neural crest cell-derived mesenchyme of the pharyngeal arches [5,6,10]. Associated with this expression, Dlx genes have been shown to have a function in the specification of neural crest cells in Xenopus [11], and later in the regionalization of the pharyngeal arches and their derivatives in mouse [12] and zebrafish [13] (regionalized arches are a gnathostome (jawed vertebrates) synapomorphy). Dlx transcription factors are expressed in the anterior brain (telencephalon and diencephalon [5,6,10], which are vertebrate synapomorphies).
where they have a role in specifying GABA-ergic interneurons [14,15]. In addition they are involved in the development of the sensory circuitry associated to the eyes and olfactory and otic organs [16–18]. Some Dlx genes are expressed in the later differentiating surface epidermis (arising from the gastrula non-neural ectoderm) and have been shown to play a role during papilla-derived appendage development: hair, tooth, and feather [9,19–22]. Most Dlx genes are transcribed in the epithelial and mesenchymal compartments of the developing paired limb buds (gnathostome synapomorphies) and median fold (a vertebrate synapomorphy) [5,6,23]. In humans, DLX3 and DLX6 have been shown to be involved in a fore- and hind-limb developmental pathway activated by p63 [24] through an enhancer located more than 250 kbp away from the bigene cluster [25]. Finally, Dlx genes are expressed in the developing cartilage and bones (both dermal and cartilage bones [10,26,27], which also are vertebrate synapomorphies) where they are involved in the differentiation of chondrocytes and osteocytes [28,29].

**Gnathostome Outgroups and Origins of the Dlx Genomic Organisation**

Within the gnathostomes, Dlx genes are found as three tandem bigene clusters in the genome (six coding sequences, Dlx1 to Dlx6), reviewed in [30] with additional genes (single or tandem clusters) in teleost genomes due to an ancestral whole genome duplication (dlx1a and dlx1b to dlx6a and dlx6b) [31,32]. Within the gnathostome sister group (cyclostomes), Dlx genes from the lamprey and hagfish have been identified but could not be strictly designated as members of the gnathostome Dlx1 to Dlx6 orthology groups [Petromyzon marinus 4 genes [1]; Lampetra japonica, 6 genes [33], Eptatretus burgeri, 6 genes [34]]. Outside the vertebrates, three Dlx genes have been identified in the urochordate Ciona, two organized as a bigene tandem cluster [35], while a single gene is found in amphioxus (AmphiDll [36]) and most protostomes (Drosophila melanogaster [37], but see [38] for exceptions). The current hypothesis is therefore that there was an ancestral tandem duplication after the divergence of the cephalochordates and the common ancestor of the urochordates and vertebrates (Figure 1), making this gene family a chordate gene family which expanded from two ancestral genes in vertebrates. This expansion is due to two series of whole genome duplications before the divergence of vertebrates [33], resulting in four bigene clusters, one being lost early, leaving three clusters in the gnathostomes: Dlx1/Dlx2, Dlx3/Dlx4, Dlx5/Dlx6. The phylogenetic relationships between those orthology groups link Dlx1, Dlx4 and Dlx6 to one of the Dlx gene from the ancestral single bigene cluster, while Dlx2, 3 and 5 are related to the other Dlx gene of this same ancestral cluster [30] (and see Figure 1). Because of their role in the development of key structures in vertebrate and gnathostome evolution, Dlx gene expression patterns have been examined extensively in lampreys and amphioxus. In the lamprey, Dlx genes are transcribed in the anterior brain, in the pharyngeal arches in a gnathostome-like fashion, in the olfactory and otic organs, pre-migrating and migrating neural crest cells, and the median fin fold [1,39]. In amphioxus, Amphi-Dll expression was detected in a zone putatively homologous to the anterior brain within the neural vesicle, in the non-neural ectoderm during late gastrulation, and in cells associated with the photosensitive organ [36].

The lamprey Dlx complement has not been characterized at the genomic level, but the conservation of the bi-gene tandem clusters in bony fish has allowed comparison of the non-coding regions surrounding the Dlx coding sequences. Comparing mammals to zebrafish, highly conserved non-coding elements (CNEs) could be identified around the Dlx1/Dlx2 pair: I12a, I12b within the intergenic sequence and URE2 upstream of Dlx1, and within the Dlx5/Dlx6 intergenic sequence: I56i and I56ii ([39–41], also see Figure 2). These CNEs are considered enhancers for the Dlx genes as they are able to drive expression of a transgene in specific Dlx expressing zones of the embryo: I12a in the pharyngeal arches (mouse [41]), I12b and URE2 in the anterior brain and pharyngeal arches (mouse [42,43]) and zebrafish [44,45], I56i in the anterior brain and pharyngeal arches (mouse [40,41]) or only in the anterior brain (zebrafish [46]) and I56ii in the anterior brain (mouse only [47], no enhancer activity in zebrafish [46]). A comparison of the intergenic sequence between Dlx3 and Dlx4 did not identify CNEs shared between the mouse and zebrafish [32] but one highly conserved element between mouse and human, I37-2, could be tested and was shown to have positive regulatory activity in the pharyngeal arches [48]. Outside the gnathostomes, no match for known CNEs could be identified in the P. marinus draft genome [32]. Outside the vertebrates, other regulatory sequences could be identified in Ciona, which were shown to be involved in driving expression in the non-neural ectoderm [49]. However the vertebrate CNEs could not be identified in Ciona and the Ciona element could not be isolated in vertebrates, suggesting...
that the conserved regulatory sequences obtained in bony fish were modified at some time between the “urochordates/vertebrates” divergence node and the “chondrichthyans/bony fish” divergence node.

**Vertebrate Evolution and the Chondrichthyan Lineage**

Within gnathostomes, little is known about the Dlx genes outside the bony vertebrate group (chiefly represented by mouse and zebrafish as model organisms). The sister clade, chondrichthyans, which groups together the sharks, rays, and chimaeras, has been much less well studied in terms of embryonic development, and therefore represents an attractive branch of the gnathostome tree. Six Dlx genes have been identified in the leopard shark *Triaakis semifasciata*, which have a usual bony fish genomic organization (at least two pairs, Dlx1–2, Dlx5–6, intergenic regions of 7–10 kb, and conserved intron-exon boundaries) [30]. We have isolated and identified 6 Dlx coding sequences in the small-spotted catshark *Scyliorhinus canicula* and described their expression pattern during tooth development [50], and recent thorough comparative studies of branchial arches development used Dlx gene expression in chondrichthyan species [51,52]. These two studies focused on the putative role of Dlx gene products (and others) in regionalizing branchial arches in the wide scheme of a Dlx code. A full description of gene expression pattern is still lacking both in terms of a wide window of organogenesis stages and in terms of global expression in the embryos. Using sequence data from the elephant shark, *Callorhinus milii*, one Dlx gene pair Dlx1–Dlx2 was identified on a single BAC associated with putative orthologous 112a and URE2 sequences [45]. A full view of the Dlx gene complement in a chondrichthyan species is therefore still lacking, in particular one which includes expression data, to enable us to propose ancestral expression domains in gnathostomes and a comprehensive view of the evolution of Dlx gene regulation within vertebrates. Here, we describe embryonic expression patterns over the neurulation and early organogenesis stages (from stage 16 to stage 26) and the genomic organization of the Dlx gene family in the catshark, and test the putative activity of some conserved non-coding elements identified around the coding sequences. These findings are analyzed in the context of gnathostome evolution, using *Ciona* and the amphioxus as outgroups. This comparative approach allows the identification of evolutionary patterns in jawed vertebrates after the series of gene duplications that occurred before jawed vertebrate evolution. In this analysis, we considered the theoretical framework of the Duplication-Degeneration-Complementation (DDC) model [53] and described occurrences of (1) neo-functionalization of some paralogs by acquisition of new expression sites; (2) sub-functionalization between paralogs, through differential degeneration of the ancestral expression sites. We could characterize this sub-functionalization as total (only one member of the family has kept the ancestral expression site) or partial (more than one) but also as early (before the diversification of jawed vertebrates) or late (during their diversification). This analysis highlights very different rates of expression pattern degeneration between paralogs of one gene family, leading to heterogeneous conservation of co-expression between paralogs from one expression site to another.

**Results**

**Catshark Dlx Genes and their Genomic Organization**

Based on exons 1 and 3, six Dlx genes have been previously identified and assigned to Dlx1 to Dlx6 orthology groups [50]. We obtained four full-length cDNA sequences from cDNA libraries made from several developmental stages in the catshark [54]: Dlx2, Dlx3, Dlx4 and Dlx5. Mapping the Dlx5 cDNA sequence from the catshark and the Dlx6 sequence from *Triaakis semifasciata* [30] onto the BAC containing the Dlx5–Dlx6 bigene cluster showed that the gene organization observed in all other gnathostomes was conserved in the catshark. The Dlx5 and Dlx6 coding regions both consist of three exons with the homeodomain coding sequence within exons 2 and 3. The STOP codons of these two genes were separated by a roughly 10 kb sequence in the BAC. Using long-range PCR we amplified the approximately 8 kb intergenic sequence between Dlx1 and Dlx2 (herein named SCinter12), and the 10 kb intergenic sequence between Dlx3 and Dlx4 (SCinter34). The intergenic regions were cloned and completely (SCinter12) or partially (SCinter34:5 kb at the Dlx3 end, 4.6 kb at the Dlx4 end, poly-G in central position) sequenced using primer walking. Alignment of these intergenic sequences with mouse and zebrafish orthologs using the VISTA genome browser tool identified regions within the SCinter12 and SCinter36 sequences that are conserved in the gnathostomes (greater than 75% conservation over 100 bp). No similarly conserved region could be identified within SCinter34 (see Figure S1). Conserved sequences from SCinter12 and SCinter36 are homologous to the putative enhancers previously characterized in mouse and zebrafish: 112a, 112b, 156i and 156ii (Table 1, Figure 2). Note that URE2 could not be identified because of its putative location outside of the Dlx gene cluster.

**Catshark Dlx Gene Expression Patterns**

Dlx probes were used to examine expression patterns of all six Dlx genes during the early neurulation to mid-organogenesis stages (stage 15 to 25). Hybridization using all probes designed against
the six Dlx coding sequences resulted in a strong signal in various structures (see supplementary material in [50]) with positive expression from stage 15 to 25 for Dlx3, Dlx4, Dlx5 and Dlx6; from stage 17 to 25 for Dlx2; and from stage 19 to 25 for Dlx1. The expression patterns are presented here as (i) non-neural ectoderm, sensory placodes and sensory vesicles; (ii) paired fins, median fin fold and analia-genitalia; (iii) anterior brain; (iv) neural crest cells and branchial arch mesenchyme. For each zone of expression, we mapped our results and those previously described from mouse, zebrafish, Ciona and amphioxus onto a chordate phylogenetic tree. We then proposed hypotheses to depict the evolutionary steps of Dlx expression patterns, both before and during the divergence of the extant gnathostomes (see Figure 1). In the case of the zebrafish data, we summed the expression patterns of both teleost-specific tandems expressed. Note that Dlx6 mRNAs were detected in both the head and trunk ectoderm at stage 17 (Figure 5 C1–3). Positive staining was obtained at stage 21 for Dlx3, 4 and 5 probes in both thickened, well-defined, olfactory placodes and folding otic placodes (see Figure 5 A1–2, B1–2, C1–2). Expression of Dlx6 could also be detected in the most peripheral zone of the closing otic placode (Figure 5 D’2) but not in the olfactory placode although it was expressed in cephalic ectoderm anterior to the placode (Figure 5 D’1). Later expression of Dlx3, Dlx4, Dlx5 and Dlx6 could be observed at stage 25 in the folding olfactory placode and otic vesicles (Figure 6 C–F 1–2). Late expression of Dlx1 and Dlx2 could be detected in a lateral region of the otic vesicle starting around stage 24 (Figure 6 A1–B1).

To analyze these results within a maximum parsimony framework, we separated them into two groups: early expression in the olfactory or otic placodes (before they fold) and late expression (after folding of the placodes into vesicles). Homologous expression patterns were included from Ciona as it has been proposed that the paired atrial siphon primordia are homologous to vertebrate otic placodes [56]. Our data were compared with that from Ciona, mouse [3,9,57] and zebrafish [5,6]. Sensory placodes are not found in amphioxus even though gene expression pattern data suggest that there may be homologous cell types in this organism [58]. The evolutionary scenario obtained shows ancestral expression of Dlx genes during the early development of paired sensory placodes, followed by early loss of expression by the Dlx1–Dlx2 cluster before gnathostome diversification, and late loss of Dlx6 expression in the catshark lineage and of Dlx3–Dlx4 in the mouse lineage (Figure 7A). Full sub-functionalization is therefore observed in the mouse lineage while only partial sub-functionalization has happened in the zebrafish and catshark lineages. Late expression of all Dlx genes in paired sensory vesicles has been retained from the last gnathostome ancestor (Figure 7B) but note that expression patterns are not strictly comparable for all genes in the catshark, Dlx5–Dlx6 genes are expressed in the dorsal aspects of the otic vesicle while Dlx1–Dlx2 genes are localized in a more lateral aspect. The observed conserved redundancy may therefore hide more subtle sub-functionalization between paralogs in subsets of their expression patterns.

(i) Non-neural ectoderm, sensory placodes and sensory vesicles. Dlx3, Dlx4, Dlx5 and Dlx6 transcripts were detected as early as stage 15 (early neurulation stage) in the non-neural ectoderm, on the rostral and lateral border of the neural plate (Figure 3). Dlx6 expression was restricted to the ectodermal margin of the neural plate border (Figure 3 H–J) while Dlx3, Dlx4 and Dlx5 expression domains extended fully to the embryonic and extra-embryonic ectoderm (Figure 3 A–C and histological sections for Dlx5 in Figure 3 D–G, histological details for Dlx3 and Dlx4 are not shown but were identical to Dlx5). During stage 16, when the neural tube starts folding, expression of Dlx3 to 6 was still observed in the non-neural ectoderm. In particular, Dlx6 transcripts were restricted to the folding and fusing part of the ectoderm at the neural tube border (Figure 3 K–M).

This first set of results of early non-neural ectodermal expression was compared to homologous expression patterns in amphioxus [36], Ciona [55], zebrafish [5,6] and mouse [3,4] and mapped onto the Dlx cluster tree (Figure 4). A maximum parsimony reconstruction suggests the following evolutionary scenario: ancestral Dlx expression in chordates was followed by several losses, one before the diversification of gnathostomes for the Dlx1–Dlx2 cluster and two independent losses, one in the zebrafish lineage (loss of dlx5a-6a expression) and the other in the mouse lineage (Dlx3–Dlx4). This case therefore illustrates late (during lineage divergence) but full sub-functionalization of one ancestral expression site (neural plate border) in the mouse and zebrafish lineage (only one Dlx tandem expressed) and only partial sub-functionalization in the catshark lineage (two Dlx tandems expressed). Note that Dlx6 expression pattern in the catshark was much more restricted but this difference was not taken into account in the analysis.

### Table 1. Catshark CNE length and similarity defined through comparison to the mouse (Mus musculus) and zebrafish (Danio rerio) intergenic sequences.

<table>
<thead>
<tr>
<th>CNE similarity</th>
<th>length</th>
<th>similarity</th>
<th>Length</th>
<th>inserts</th>
</tr>
</thead>
<tbody>
<tr>
<td>I11a</td>
<td>94%</td>
<td>429 bp</td>
<td>87%</td>
<td>461 bp</td>
</tr>
<tr>
<td>I11b</td>
<td>72%</td>
<td>318 bp</td>
<td>72%</td>
<td>215 bp</td>
</tr>
<tr>
<td>I56i</td>
<td>76%</td>
<td>338 bp</td>
<td>75%</td>
<td>285 bp</td>
</tr>
<tr>
<td>I56ii</td>
<td>74%</td>
<td>283 bp</td>
<td>74%</td>
<td>227 bp</td>
</tr>
</tbody>
</table>

Expression from stage 15 to 25 for Dlx3, 4, 5 and 6 were expressed in the cephalic ectoderm (Figure 5 and Figure 6). As early as stage 17, messenger RNAs of Dlx3, 4, 5 and 6 were detected in the developing sensory organs, in particular in the prospective olfactory and otic placodes (Figure 5 A1–2, B1–2, C1–2). At the same stage, Dlx6 mRNAs could be detected in the presumptive olfactory region, but not yet in the presumptive otic placode (Figure 5 D1–2). Expression of Dlx3 and Dlx4 was also seen in the cephalic ectoderm but not in the lateral trunk ectoderm (Figure 5 A1–3 and B1–3), with expression in the lens placode at stage 21 (Figure 5 A1’ and B1’). In contrast, Dlx5 mRNAs were detected in both the head and trunk ectoderm at stage 17 (Figure 5 C1–3). Between stage 16 and stage 25, Dlx3, 4, 5 and 6 were expressed in the cephalic ectoderm (Figure 5 and Figure 6). As early as stage 17, messenger RNAs of Dlx3, 4, 5 and 6 were detected in the developing sensory organs, in particular in the prospective olfactory and otic placodes (Figure 5 A1–2, B1–2, C1–2). At the same stage, Dlx6 mRNAs could be detected in the presumptive olfactory region, but not yet in the presumptive otic placode (Figure 5 D1–2). Expression of Dlx3 and Dlx4 was also seen in the cephalic ectoderm but not in the lateral trunk ectoderm (Figure 5 A1–3 and B1–3), with expression in the lens placode at stage 21 (Figure 5 A1’ and B1’). In contrast, Dlx5 mRNAs were detected in both the head and trunk ectoderm at stage 17 (Figure 5 C1–3). Between stage 16 and stage 25, Dlx3, 4, 5 and 6 were expressed in the cephalic ectoderm (Figure 5 and Figure 6). As early as stage 17, messenger RNAs of Dlx3, 4, 5 and 6 were detected in the developing sensory organs, in particular in the prospective olfactory and otic placodes (Figure 5 A1–2, B1–2, C1–2). At the same stage, Dlx6 mRNAs could be detected in the presumptive olfactory region, but not yet in the presumptive otic placode (Figure 5 D1–2). Expression of Dlx3 and Dlx4 was also seen in the cephalic ectoderm but not in the lateral trunk ectoderm (Figure 5 A1–3 and B1–3), with expression in the lens placode at stage 21 (Figure 5 A1’ and B1’). In contrast, Dlx5 mRNAs were detected in both the head and trunk ectoderm at stage 17 (Figure 5 C1–3). Positive staining was obtained at stage 21 for Dlx3, 4 and 5 probes in both thickened, well-defined, olfactory placodes and folding otic placodes (see Figure 5 A1–2, B1–2, C1–2). Expression of Dlx6 could also be detected in the most peripheral zone of the closing otic placode (Figure 5 D’2) but not in the olfactory placode although it was expressed in cephalic ectoderm anterior to the placode (Figure 5 D’1). Later expression of Dlx3, Dlx4, Dlx5 and Dlx6 could be observed at stage 25 in the folding olfactory placode and otic vesicles (Figure 6 C–F 1–2). Late expression of Dlx1 and Dlx2 could be detected in a lateral region of the otic vesicle starting around stage 24 (Figure 6 A1–B1).
Dlx genes before their duplication during the two rounds of genome duplication with subsequent loss in the catshark lineage (Dlx1–Dlx2 cluster). These results highlight very low (catshark) to no sub-functionalization (mouse, zebrafish) of Dlx genes in these structures. In the median fin fold, the scenario is identical except that all mouse Dlx genes have lost this zone of expression as the structure itself has disappeared from mammalian embryos (Figure 8B, expression in the zebrafish are from [6]). Finally, data for the analia genitalia were more difficult to compile: for the mouse, we complemented data from [60] with that from the GUDMAP Project [61] and in situ hybridization results at E14.5 from [62], available at MGI [63]. Together these data show that all Dlx genes are expressed during genital tubercle development and/or the genito-urinary system; for the zebrafish, we used data submitted by [64] in ZFIN [65] at 36hpf, showing that all Dlx genes except for dlx1a is transcribed in the developing analia genitalia at this stage. The scenario obtained was almost identical to that proposed for the AER zone, except for the loss of dlx1a expression in the zebrafish lineage and the loss of Dlx6 expression in the catshark lineage (Figure 8C). Again, a very low level of sub-functionalization is observed for Dlx expression patterns in these structures.

(iii) Anterior brain. Dlx5 mRNAs were detected in the most anterior part of the neural tube at stage 17 (Figure 5 C1). During stage 21, faint expression of Dlx5 (Figure 5C′1) and Dlx6 (not shown) was still detected in the anterior-most part of the telencephalon at the level of the olfactory placodes and still seen until at least stage 25 (not shown). Starting around stage 25, some cells of the ventral-most part of the telencephalon (subpallium) started showing expression of Dlx1, Dlx2 and Dlx5 (Figure 9 A–C) while scattered positive cells were also observed in the diencephalon (Figure 9 A, D–F).

To analyze these results, we excluded the earlier and anterior-most expression of Dlx5 and Dlx6 in the catshark telencephalon, because this zone of expression was not strictly homologous to those previously described in mouse and zebrafish. We only kept the later expression of Dlx genes in the development of the ventral telencephalon and diencephalon, and compared them to data from mouse and zebrafish (reviewed in [40] and [6]). The expression of Amphi-Dll has been described in an anterior subset of
expression in the subsequent pharyngeal arch development has been published elsewhere [52].

Of the Dlx genes in gnathostomes, only Dlx2 cognates are expressed in pre-migrating and migrating neural crest cells, as described in the catshark (this work), zebrafish [3] and mouse [7]. Neural crest cells are a vertebrate synapomorphy: no homologous structure can be found in amphioxus. In *Ciona*, some cells have been proposed as the putative homolog to the vertebrate neural crest cells but they have been shown not to express Dlx homologs [55]. The most parsimonious evolutionary hypothesis is therefore a single unique cooption of the Dlx2 coding sequence in the development of this structure after gnathostome paralog duplication but before gnathostome divergence (Figure 12) followed by strong conservation of this expression pattern during gnathostome diversification (but see [68]). This would be the only example of neo-functionalization observed in gnathostome Dlx paralogs.

**Putative Catshark Dlx Regulatory Elements**

Transient transgenic zebrafish were obtained by injection of a GFP-reporter plasmid. This plasmid is negative for expression by itself but becomes active when an enhancer sequence is inserted. We built eight constructs with each of the four identified CNEs (I12a, I12b, I56i, I56ii) in both possible orientations. We obtained transient positive GFP expression in the nervous system of zebrafish embryos with the I56i and I12b sequences from the catshark genome (mean 40% positive embryos, n > 200 for each CNE, n > 100 for each CNE in each orientation), GFP expression could be detected with these sequences in both orientations suggesting they act as true regulatory sequences and not as promoters. Specific GFP-positive cells were observed in the anterior part of the developing brain only, in the telencephalon of embryos from 24 hours post fertilization (hpf) to 72 hpf (Figure 13 and not shown) consistent with the telencephalic expression observed at stage 25 with Dlx1 and Dlx2 (Figure 9). Other sites of expression observed in our description of catshark Dlx gene expression patterns were not sites of GFP expression in our transgenic zebrafish embryos. GFP expression was never observed with I56ii and I12a in embryos from 1 dpf until 3 dpf (n > 100 for each CNE in each orientation), consistent with the absence of expression obtained with zebrafish I12a and I56ii in zebrafish transient transgenesis at early embryonic stages (not shown, n > 100 embryos for each construct).

**Discussion**

**Conservation of the Genomic Structure and Associated CNEs**

The Dlx genomic organization in the catshark showed that such organization is highly conserved within the gnathostomes: that is, three bigene clusters in a tail-to-tail arrangement, including an approximately 10 kb intergenic sequence with slight temporal and space differences suggesting at least partial sub-functionalization between paralogs in osteichthyans.

(iv) **Neural crest cells and pharyngeal arch mesenchyme.** Dlx2 transcripts were detected in the cranial neural crest cells before and during (stage 17–19) their migration from the neural tube to the pharyngeal arches (Figure 11A–E). Dlx2 expression was localized all along the anterior-posterior axis, in a dorsal aspect of the neural tube from stage 17 to stage 19 equivalent to trunkal neural crest cells (Figure 11A–E, K) but was then observed only in cranial neural crest cells during their migration toward the head and pharyngeal arches mesenchyme (Figure 11G–J). Dlx2-positive cranial neural crest cell migration was complete at stage 21. A complete description of Dlx gene
Figure 5. *Dlx* gene expression patterns at stage 17 and 21 in the catshark. Target gene name is indicated on top of each column, stage of development is indicated on top of each box: A–D, A’–D’: lateral views of whole-mount embryos; A–D dorsal to the top, anterior to the left; A’–D’ are anterior to the top, dorsal to the right. A1–3, B1–3, C1–3, D1–3, A’1–3, B’1–3, C’1–3, D’1–3 are transverse sections with dorsal to the top; A1–D1 go across the presumptive olfactory placode (black arrow) as located by section plan #1 on the A panel; A2–D2 go across the presumptive otic placode (double arrowhead), plan #2 on the A panel; A3–D3 go across the trunk with developing median fin fold (asterisk), plan #3 on the A panel; A’1–D’1 go across the olfactory placode (olp) and lens placode (arrowhead), plan #1 on the A’ panel; A’2–D’2 go across the folding otic vesicle (otv), plan #2 on the A’ panel; A’3–D’3 go across the trunk with developing median fin fold (mff), plan #3 on the A’ panel. ce: cephalic epithelium; cv: cephalic vesicle; fb: forebrain; lm: lateral mesoderm; n: notochord; nt: neural tube; oc: optic cup; rb: rhombencephalon. Scale bars: A–D, A’–D’: 200 μm; A1–3, B1–3, C1–3, D1–3: 50 μm; A’1–3, B’1–3, C’1–3, D’1–3: 100 μm.

doi:10.1371/journal.pone.0068182.g005
 structural arrangement of the Dlx genes and on the intergenic CNE sequences at least over the past 430 Ma. Previously, similar results in other vertebrates were considered good examples of highly conserved gene regulatory blocks linking developmental genes to their shared enhancers [40,70].

Our transient transgenesis assays demonstrated that enhancer activity in the brain is conserved for catshark I12b and I56i CNEs.
when tested in zebrafish (Figure 13), as has been previously shown for homologous sequences from zebrafish and mice [40,41,43,44,46]. However, major technical advances are still necessary to test the enhancer activity of these sequences within catshark embryos to support these results and describe their activity in situ. The results we obtained in zebrafish suggest that enhancer activity is known to be positive are written black while genes for which no expression has been recorded are indicated in grey.
doi:10.1371/journal.pone.0068182.g007

Single Gene vs Cluster Regulation

Enhancer sharing by two genes of the same cluster has been proposed to explain both the conservation of the genomic organization and overlapping Dlx gene expression [40]. However, the detailed and complete expression patterns we describe here show clear differences between the Dlx clusters (Table 2). Expression of Dlx3 and Dlx4 were identical (over the observed stages of development and at the level of precision of an in situ hybridization) while there were clear differences between Dlx5 and Dlx6 expression patterns: Dlx6 was repeatedly found expressed within a subset of the Dlx5 spatial expression pattern, notably in the folding zones of organs such as the neural plate border, olfactory and otic placodes (see Figure 3, 5, 6). Dlx1 and Dlx2 expression patterns were more restricted and showed both zones of co-expression (late expression in the otic vesicle, telencephalon) and in one case, Dlx2 specific expression in the cranial and trunkal neural crest cells before or during their migration. The hypothesis of shared enhancers for both genes of a bigene cluster in the catshark therefore seems applicable for most Dlx genes. The Dlx5–Dlx6 situation suggests an additional hypothesis: as Dlx6 was expressed within a subset of the Dlx5 expression pattern, and notably often in the folding/fusing zones, there may be a shared enhancer for this cluster with an additional restricting regulatory sequence specifically linked to Dlx6. This putative regulatory situation in the catshark is concordant with proposed evolutionary scenarios in which there is a lineage-specific loss of Dlx6 expression in the early otic placode, paired and median fins, anterior brain and restriction of expression in the non-neural ectoderm, late otic and olfactory placode development. This scenario also supports the hypothesis that there originally was co-expression of Dlx5 and Dlx6 driven by a shared enhancer, followed by lineage-specific addition of a Dlx6 negative regulator modulating the original expression pattern.

Heterogeneous Rates of Sub-functionalization among Expression Sites

The catshark gene expression patterns described here show both similarities and differences with those of other gnathostomes. Reconstruction of ancestral states with maximum parsimony resulted in various evolutionary scenarios depending on the expression site involved. Only one case of neo-functionalization
is highlighted with our results: Dlx2 expression in premigratory and migratory neural crest cells. This novelty originated before gnathostomes diversified and was highly conserved in all three lineages analyzed (Figure 12). Other expression sites were all inherited from the ancestral Dlx bigene cluster. We describe one scenario with no detectable sub-functionalization (late expression of all Dlx genes in the paired sensory placodes, Figure 7B) and two scenarios with weak sub-functionalization (loss of Dlx1–Dlx2 expression only in paired fins/limbs (Figure 8A) and the same scenario with additional loss of catshark Dlx6 and zebrafish dlx1a expression in analia genitalia (Figure 8C)). On the other hand, other sites of expression show more extensive paralog sub-functionalization: expression in the non-neural ectoderm (Figure 4), early paired sensory placode development (Figure 7A), expression of all but Dlx3–Dlx4 in the developing forebrain (Figure 10).

Lineage-specific evolutionary events may explain other peculiar losses: for example, the case of the Dlx3–Dlx4 cluster is well-known in the mouse (and most probably represent the situation in all

Figure 9. Dlx gene expression during brain development in the catshark. Target gene name and stage of development are indicated on each panel. Transverse sections through whole-mount hybridized embryos at stage 25, dorsal to the top, at the level of the telencephalon (A–C) or more posterior through the diencephalon (D–F). Dlx1, Dlx2 and Dlx5 are expressed in scattered cells of the telencephalon (arrowhead in A–C) and of the diencephalon (arrow in A and D–F). Note expression of Dlx1 in scattered cells of the head mesenchyme at the mandibular level, interpreted as cranial neural crest cells (cnc). Scale bars: 50 μm.

doi:10.1371/journal.pone.0068182.g009
mammalian species) where there has been both recruitment of these genes in placenta development [71,72], and a specific case of protein function modification for Dlx4, from a nuclear transcription factor to cytoplasmic protein [73]. Modification of Dlx expression patterns is therefore probably not independent in early developing paired sensory placodes and neural plate border, while placental expression may be considered as a modified “non-neural ectoderm expression” at the neurula stage. Another case of non-independency in our evolutionary scenarios is the loss of mouse Dlx expression during median fin fold development, because of the simple loss of this structure, even at embryonic stages, in mammals (Figure 8B). Finally we observe a very similar set of evolutionary scenarios in the case of limb/median fin fold/analia genitalia, with

Figure 10. Evolutionary scenario for the expression of Dlx genes in anterior brain. See Figure 4 for legends. doi:10.1371/journal.pone.0068182.g010

Figure 11. Dlx2 expression in neural crest cells in the catshark. A–E: lateral views of whole-mount embryos, anterior is to the top, dorsal to the right, between stage 17 and 21. F: schematic of the head for embryo in panel B with location (blue colored) of the various streams of migrating neural crest cells: trigeminal (tg) stream separated in the premandibular (pm) and mandibular (m) streams; hyoid (h) stream; branchial (b) stream. G–K are transverse sections made on the embryo shown in panel B. The section level is located on panel F: G = section 1(S1) to K = section 5 (S5). A: stage 17, expression is restricted to the neural crest cells both at the cranial and trunk level. B: streams of cranial neural crest cells migrate ventrally: Dlx2 expression detected in the various cranial neural crest streams and most posteriorly in the trunk neural crest cells (tnc). D, E: after migration, Dlx2-positive cells are found in the mesenchyme of branchial arches (ba), ba1: first branchial arch, ba2: second branchial arch, oc: optic vesicle, nt: neural tube. Scale bars: A–E: 200 μm; G–K: 100 μm. doi:10.1371/journal.pone.0068182.g011

Figure 12. Evolutionary scenario for the expression of Dlx genes in premigrating and migrating neural crest cells. See Figure 4 for legends. doi:10.1371/journal.pone.0068182.g012

Dlx expression during median fin fold development, because of the simple loss of this structure, even at embryonic stages, in mammals (Figure 8B). Finally we observe a very similar set of evolutionary scenarios in the case of limb/median fin fold/analia genitalia, with
redundant expression patterns for all Dlx genes except in the case of the catshark Dlx6 and Dlx1–Dlx2 cluster (Figure 8). This observation is consistent with the hypothesis of a common gene regulatory network which may have been coopted in early vertebrates from the median fin fold towards the lateral mesoderm to make the paired appendages [74]. A common gene regulatory network has also been proposed for the development of both distal limb and genitalia [75] which would explain overall similarities in the evolution of Dlx expression patterns in these three structures.

In all cases, the evolutionary scenarios proposed here involve much fewer evolutionary steps than the one we previously proposed for Dlx expression in gnathostome teeth [50] where the discrete expression pattern in tooth buds are very different when the two genes of a cluster are compared, and also very different from one lineage to another. This difference may be because of the detail in which we looked at expression patterns (for example, here we coded dorsal expression of Dlx3–6 and the lateral expression of Dlx1–2 in the otic vesicle in a similar way). However, the low conservation of Dlx expression pattern in teeth and scales is most probably the result of complete independency of gene expression pattern between genes of a single cluster, therefore also showing extensive independent changes in expression patterns during vertebrate evolution [50]. These variations in the pace of change in gene expression patterns may therefore be dependent on: (i) the putative functional redundancy within the family (as in Figures 4 and 7A) as opposed to strict specificity of the paralog(s) involved in one zone (as in Figures 10 and 12), and (ii) the possibility of each gene of a cluster evolving independently or not, which may be more likely for some structures than for others (previously discussed).

**Evolution of Regulatory Sequences vs Evolution of Expression Patterns**

As major players in chordate development and in vertebrate and gnathostome morphological novelties, Dlx genes have been extensively examined in model species and gnathostome outgroups. With the description of the Dlx gene complement in the
catshark, we obtained several evolutionary scenarios for Dlx gene expression modification before and after the vertebrate whole-genome duplications. These results show that there are some morphological regions with highly conserved gene expression within the gnathostomes, such as migrating neural crest cells, late sensory placodes and brain either because of early sub-functionalization (before lineage divergence) or because of non-sub-functionalization. The two former zones of expression have not been linked yet to any putative regulatory sequence while the latter has been linked to gnathostome conserved non-coding elements which have highly conserved regulatory activity in mouse, zebrafish and possibly in the catshark. In contrast, expression in the gnathostome neural plate border showed that there was sub-functionalization after the second round of whole-genome duplications and during gnathostome diversification. One enhancer element was identified just upstream of the Ciona DllB promoter and shown to drive expression in the non-neural ectoderm. Both results are congruent with the presence of a non-neural ectoderm enhancer linked to the ancestral Dlx cluster which may have then been lost differentially in various lineages, and whose sequence has evolved significantly during urochordate and gnathostome evolution. Finally, two sets of results suggest ancestral involvement of Dlx genes in gnathostome early placode and fin/limb/analia genitalia development, with secondary losses in the different lineages. No enhancer has been described for the regulation of early placode expression, but the regulation of Dlx5 and Dlx6 in the mammalian limb has been proposed to act through a promoter as well as a long-range enhancer (located more than 250 kb upstream of Dlx6) in a response to p63 binding [24, 25]. Sequence conservation for this long-range enhancer appears to be restricted to amniotes (see Supplementary Figure 8 in [25]), though, as suggested by our evolutionary scenario, the enhancer activity may be conserved within gnathostomes.

**Conclusion**

The Duplication-Degeneration-Complementation model [53] suggests that there is sub-functionalization of each paralog after duplication and deep evolutionary time. This has been reported for several gnathostome homeobox gene families with expression pattern complementation between duplicates or loss of one of the duplicates [76]. However, our description of gnathostome Dlx expression patterns depicts various evolutionary scenarios including frequent conservation of the ancestral expression pattern for all, or most, duplicates (e.g. in fins, sensory placodes/vesicles) leading to high redundancy in paralog expression patterns. These observations may be interpreted as a sign of functional complementation between paralogs though functional studies of Dlx genes tend to show redundancy between them instead [8, 12]. This situation with multiple redundant paralogs has been observed in other homeobox families within the gnathostomes, such as the Hox genes [77]. Hox genes are also organized as clusters with shared regulatory sequences, and also are highly pleiotropic genes. We interpret these data as a consequence of an “evolutionary inertia” where the genomic structure, presence of shared regulatory sequences, and gene pleiotropy, act as a brake which strongly slows down the evolutionary process of degeneration and complementation among paralogs.

**Materials and Methods**

**Ethic Statements**

For the zebrafish experiments, we obtained a permit for protocol #BL-256 from the Ottawa University Animal Care Committee. The manipulations on zebrafish embryos were all performed after euthanasia with MS222 so that potential suffering was minimized. No permit was needed for catshark manipulations because they were limited to early embryonic stages, all less than stage 26.

**cDNAs and BAC Sequences**

Dlx gene sequences from *Triakis semifasciata* were used to design degenerate primers to amplify partial catshark cDNA sequences (for details see [50]) and were compared using BLAST to the 5 cDNA libraries obtained from catshark embryos (for details see [54]) to obtain full-length sequences. Putative exons 1 and 3 of each six Dlx coding sequences have previously been amplified from genomic DNA and then assigned to each of the six gnathostome Dlx orthology groups [50]. The amplified exons 1 and 3 were then used to probe a BAC library of the catshark genome (see [54] for details) with radiolabelled probes. One clone was isolated and tested positive for the presence of Dlx5 and Dlx6. Another clone was tested positive for the presence of Dlx4 only. The BAC clone containing Dlx5 and Dlx6 was sequenced by standard shotgun sequencing method at the Genoscope (France).

**Long-range PCR**

Intergenic regions linking Dlx1 to Dlx2 (SCinter1–2) and Dlx3 to Dlx4 (SCinter3–4) were amplified using the Expand Long Template PCR System with System 3 (Roche). Primers were designed in the end of the coding sequence of the third exon (SCinter1–2: Dlx1.1 GAAACAGGTTAATGTTGCTGATTAGAGAACAGCG, Dlx1.2 CCCGCCGCTTTATTAGGGAACACTCTCGTC; SCinter3–4: Dlx3.1 CGATGAGCACTGCTCAAAACACCG, Dlx3.2 CACCTGCCACCTCCTGCTTTGTC). The following cycling conditions were used; for SCinter1–2, initial denaturation 92°C, 2’, followed by 10 cycles of denaturation (D) 92°C, 20’, annealing (A) 68°C, 30’, elongation (E) 68°C, 15’, followed by an additional 25 cycles of: D- 92°C, 20’, A- 68°C, 30’, E- 68°C, 15’ 20” and final E- 68°C, 7’; for SCinter3–4: initial D- 92°C, 2’, followed by 10 cycles D- 92°C, 20’, A- 64°C, 30’, E- 68°C, 15’, followed by an additional 20 cycles D- 92°C, 20’, A- 64°C, 30’, E- 68°C, 15’ 20” and a final E- 68°C, 7’. Starting volumes of PCR reactions were 35µL for SCinter1–2 and 200µL for SCinter3–4. PCR reactions were purified using the High Pure PCR Template Preparation Kit (Roche), precipitated using standard methods and resuspended in 25 µL 1xTE. Purified PCR reactions were polished and 5’-phosphorylated using the Expand Cloning Kit (Roche), purified again using the High Pure PCR Template Preparation Kit (Roche), precipitated using standard methods and resuspended in 10 µL 1xTE. 125 ng (SCinter1–2) or 300 ng (SCinter3–4) of purified, polished and 5’phosphorylated DNA was inserted into Vector I, packaged into the *E.coli* DH5α using the Expand Cloning Kit (Roche). Positive clones were selected, purified using the Wizard Plus Miniprep DNA Purification System (Promega) and screened for clones with the correct insert size by restriction digest with *Ndel* using standard methods.

A primer walking approach was then used to sequence both intergenic sequences, leading to a full sequence of about 10 kb between Dlx1 and Dlx2, and a partial intergenic sequence of about 9.6 kb between Dlx3 and Dlx4 (two stretches of poly-G blocked further sequencing in the centre of the intergenic sequence, 5 kb could be sequenced at the Dlx3 end, 4.6 kb at the Dlx4 end, poly-G in central position).
VISTA Alignments
The sequenced intergenic sequences were aligned against their orthologs in mouse (MMinter12, MMinter34 and MMinter56) and zebrafish (DRinter1a2a, intergenic region between dlx1a and dlx2a; DRinter3b4b, intergenic region between dlx3b and dlx4b; DRinter56i, intergenic region between dlx3a and dlx6a) retrieved from the Ensembl genome browser data (respectively NCBI M37 and Zv9). The alignment was done using the mVISTA tool from the Vista Genome Browser [78], with a threshold of 70% similarity over 100 bp.

Catshark Embryo Staging
Catshark (Stylophonus canaliculatus) embryos were obtained from the Station de biologie marine (Roscoff, France, CNRS and MNHN). All embryos were maintained at 17°C in sea water until dissection and staging after [79]. They were dissected and then fixed for 48 hours at 4°C in a phosphate buffered saline solution containing 4% paraformaldehyde. Embryos were then dehydrated in methanol and stored at −20°C.

Catshark Dlx Probes and in situ Hybridization
Dlx probes were designed as described in [50]. Antisense RNA digoxigenin-UTP probes were transcribed using SP6 or T7 RNA polymerases (Roche), according to the orientation of the insert in the plasmid. In situ hybridisations on catshark embryos were performed according to standard methods with modifications described in [80]. Proteinase K treatments (10 μg/ml) were adapted for different embryonic stages: 15 min at room temperature for young embryos (before stage 20), 30 min for older embryos. The colour detection step was performed using the NBT-BCIP reaction. Embryos were post-fixed in 4% paraformaldehyde. Embryos were then dehydrated in over 100 bp, conserved regions within the coding sequences are (arrow). The alignment was done using the mVISTA tool from the Vista Genome Browser, conservation level is shown as a curve along the alignment. Beyond the chosen threshold (70% similarity over 100 bp), conserved regions within the coding sequences are shown in blue, transcribed non-coding regions in white, untranscribed non-coding regions in red. Primers used to amplify the conserved putative regulatory regions are indicated on top of the alignments (arrowheads).

Supporting Information
Figure S1 mVISTA alignments of the catshark Dlks intergenic sequences with their mouse and zebrafish orthologs. Each of the catshark intergenic sequences, SCinter1–2 (top), SCinter2–3 (middle) and SCinter3–5–6 (bottom), are aligned against the mouse (alignments 1) and the zebrafish (alignments 2) orthologous regions. The transcribed sequence of each gene is indicated (arrow). The alignment was done using the mVISTA tool from the Vista Genome Browser, conservation level is shown as a curve along the alignment. Beyond the chosen threshold (70% similarity over 100 bp), conserved regions within the coding sequences are shown in blue, transcribed non-coding regions in white, untranscribed non-coding regions in red. Primers used to amplify the conserved putative regulatory regions are indicated on top of the alignments (arrowheads).

Acknowledgments
We thank Sylvie Mazan for access to the catshark BAC and cDNA sequences.

Author Contributions
Conceived and designed the experiments: VBB DC PL ME MDT. Performed the experiments: MDT JP CJM VBB. Analyzed the data: MDT VBB. Contributed reagents/materials/analysis tools: MD ME. Wrote the paper: MDT VBB.

References
7. Qin M, Bellone A, Ghattas I, Meneses JJ, Christensen L, et al. (1997) Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches of the zebrafish Dlx5–6 BAC clone, or of the subcloned Dlx1–Dlx2 intergenic sequence obtained by long-range PCR were designed against sequence 3–10 bp upstream of the conserved sequence; sc156i-R = GCCATG-GGACTGAGTCATCATT; sc156i-F = CATTGGCCCTACCAATTG (397 bp amplified); sc112a-R = AAACCGGCCTAACTCAGAG; sc112a-