uOttawa

L'Université canadienne
Canada's university

FACULTÉ DES ÉTUDES SUPÉRIEURES ET POSTOCTORALES

uOttawa

L'Université canadienne
Canada's university

FACULTY OF GRADUATE AND POSDOCTORAL STUDIES

Ashish Kumar Maurya
AUTEUR DE LA THÈSE / AUTHOR OF THESIS

M.Sc. (Biology)
GRADE / DEGREE

Department of Biology
FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Regulatory Mechanisms Controlling Distal-less Related Gene Expression in Lamprey

TITRE DE LA THÈSE / TITLE OF THESIS

Dr. M. Ekker
DIRECTEUR (DIRECTRICE) DE LA THÈSE / THESIS SUPERVISOR

CO-DIRECTEUR (CO-DIRECTRICE) DE LA THÈSE / THESIS CO-SUPERVISOR

EXAMINATEURS (EXAMINATRICES) DE LA THÈSE / THESIS EXAMINERS

Dr. L. Bonen

Dr. H. Rundle

Dr. W. Willmore

Gary W. Slater
Le Doyen de la Faculté des études supérieures et postdoctorales / Dean of the Faculty of Graduate and Postdoctoral Studies

Regulatory mechanisms controlling *Distal-less* related gene expression in lamprey

By

Ashish K. Maurya

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
University of Ottawa
In partial fulfillment of the requirements for the
M.Sc. degree in Biology
Ottawa-Carleton Institute of Biology

Thèse Soumise à la
Faculté des Études Supérieures et Postdoctorales
Université d'Ottawa
En vue de l'obtention d'une
Maîtrise en Biologie
L'Institut de Biologie d'Ottawa-Carleton

Your file Votre référence
ISBN: 978-0-494-25804-0
Our file Notre référence
ISBN: 978-0-494-25804-0

Acknowledgements

Foremost, I would like to thank Dr. Marc Ekker for providing supervision, funding and lab-space for the current study. Thank you Gary Hatch for teaching most of the techniques used in the lab and also for answering numerous questions of mine. Thanks to Dr. Luc Poitras for helpful discussions and advice on generation of the restriction map.

Thanks to Martin Paquette for help in shotgun sequencing of the cosmids. Thanks to Kyle Martin for help with zebrafish microinjections and raising fry for the intergenic and C3 *Eco*R1 8.1 *Egfp* constructs.

Thanks to Dr. Marie-Andree Akimenko for letting me use her lab for mouse work and getting the cosmids from RZPD. Thanks to Dr. Sylvie Retaux (CNRS-Gif-sur-Yvette, France) for providing lamprey genomic DNA. Thank you Dr. Barry L. Wanner and the *E.coli* genetic stock centre at Yale University for providing the pKD46 red recombinase plasmid. Thanks to Dr. Bill Willmore, Dr. Thomas Moon and Dr. Christofre Martin for being on my committee.

Thanks to members of the Ekker lab for their support, advice and friendship. Lastly, my sincere thanks to my family and friends for their patience and support as well as Dr. Ekker for his excellent counsel.

Abstract

Vertebrate *Dlx* genes code for homeodomain (a conserved DNA binding domain) transcription factors involved in the development of the forebrain, craniofacial skeleton, dentition, sensory organs and limbs. Vertebrates have at least 6 *Dlx* genes whereas invertebrates have only one *Dlx* homologue (called the *distal-less* or *Dll* gene) (Stock, Ellies et al. 1996). Vertebrate *Dlx* genes are organized as pairs of convergently transcribed genes. The intergenic region between each of these bigenes contains at least two highly conserved sequences (300-500bp each) that act as enhancers of transcription of the flanking *Dlx* genes. Activity of these enhancers when tested with reporter genes (*GFP* and *LacZ*) in transgenic zebrafish and mice recapitulates the endogenous *Dlx* expression patterns.

Here we are investigating *Dlx* regulation in lamprey, an agnathan species and a sister group of gnathostomes. In doing this we hope to (1). trace back the origin of the conserved *Dlx* intergenic enhancers found in vertebrates and (2). address the question, can the differences in regulation of lamprey *Dlx* genes with that of gnathostomes account for some of the morphological differences seen in lampreys (like absence of jaws)?

We have shown that lamprey have at least four *Dlx* genes (*A, B, C, D*), two of which are arranged in the bigenic configuration seen in other vertebrates. We were unable to find the presence of the conserved intergenic enhancers by Southern hybridization and PCR experiments within this lamprey *Dlx* intergenic region. We have shown that this region can target expression of reporter genes in the forebrain, branchial arches and trunk somites when tested in zebrafish embryos in a reporter gene assay, suggesting many of the *Dlx* regulatory mechanisms are conserved between lamprey and gnathostomes.

Résumé

Les gènes *Dlx* codent pour des facteurs de transcription à homéodomaine (domaine de liaison à l'ADN conservé) impliqués dans le développement du cerveau antérieur, du squelette crano-facial, des dents, des organes sensoriels et des membres. Les vertébrés ont au moins six gènes *Dlx*, alors que chez les invertébrés on ne retrouve qu'un homologue des *Dlx*, appelé *Distal-less* ou *Dll*. Les gènes *Dlx* des vertébrés sont organisés en paires de gènes transcrits de façon convergente. La région comprise entre chaque paire de gènes (région intergénique) contient au moins deux séquences hautement conservées (300-500 pb chacune) qui agissent comme des éléments activateurs de la transcription des gènes *Dlx* adjacents. L'étude de l'expression de gènes rapporteurs (*GFP* et *LacZ*) lors d'expériences de transgénèse chez le poisson zèbre et la souris, a montré que l'activité de ces éléments activateurs récapitule le patron d'expression endogène des gènes *Dlx*.

Le but de cette thèse est d'étudier la régulation des gènes *Dlx* chez la lamproie, un poisson de la famille des agnathes, groupe-frère des gnathostomes. Nous pensons ainsi pouvoir (1) retracer l'origine phylogénétique des séquences activatrices conservées présentes dans la région intergénique des gènes *Dlx* de vertébrés et (2) adresser la question suivante : Est-ce que les différences dans la régulation des gènes *Dlx* chez la lamproie et les gnathostomes peuvent être à l'origine de certaines spécificités morphologiques observées chez la lamproie, comme l'absence de mâchoire?

Nous avons démontré que la lamproie possède au moins quatre gènes *Dlx* (*A, B, C et D*) dont deux sont organisés suivant la même configuration que celle observée chez les autres vertébrés (paire de gènes transcrits de manière

convergente). Suite à des analyses d'hybridation de type «Southern» et de PCR sur la région intergénique des gènes *Dlx* de lamproie, nous avons été incapables de déterminer la présence d'éléments activateurs conservés. Néanmoins, nous avons démontré que la région intergénique pouvait cibler l'expression d'un gène rapporteur au niveau du cerveau antérieur, des arcs branchiaux et des somites du tronc lorsque testé dans par transgénèse chez le poisson zèbre.

Table of contents

Table of Figures

1. Introduction.

The *Distal-less* (*Dll*) gene, first identified in *Drosophila* codes for a homeodomain transcription factor (vertebrate orthologues work as activators of transcription (Zerucha, Stuhmer et al. 2000)). The homeodomain is a 60 amino acid protein domain that confers DNA binding and recognition ability. It is found in a large family of proteins in eukaryotes that are involved in gene regulation and development. Homeoboxes (180bp DNA sequence) that code for the homeodomain were first identified in two homeotic genes of *Drosophila* namely, *Antennapedia* and *Ultrabithorax* (McGinnis, Levine et al. 1984; Gehring 1987). Homeobox genes are well known for patterning the body plan of both invertebrates and vertebrates. Invertebrates like *Drosophila* have a single *Distal-less* gene whereas in higher vertebrates (mice, zebrafish and humans) there are at least 6 *Distal-less* genes (called *Dlx* in vertebrates).

Distal-less was named, because of its role in the development of distal appendages in *Drosophila* (Cohen, Bronner et al. 1989). *Drosophila* null mutants for *Dll* die as embryos and they lack the rudimentary larval limbs.

In vertebrates, mutations in single *Dlx* genes do not show profound effects on development as in *Drosophila*. Homozygous mutants are viable and show only subtle defects in development (Qiu, Bulfone et al. 1995; Qiu, Bulfone et al. 1997; Acampora, Merlo et al. 1999; Depew, Liu et al. 1999). This may be attributed to the fact that pairs of *Dlx* genes are expressed in highly overlapping patterns during development and the loss of one *Dlx* gene is probably compensated by other *Dlx* genes. Consistent with this hypothesis, *Dlx* double mutants in mice are not viable and show drastic phenotypes like severe defects in central nervous system (CNS), distal limbs and jaws (Qiu, Bulfone et al. 1997; Depew, Lufkin et al. 2002).

1.1. Drosophila *Dll*: expression and function.

Various duplicated *Dlx* genes in vertebrates combined with their varied spatiotemporal expression patterns provide a very complex system to begin understanding the function of *Distal-less*. *Drosophila* with its single *Dll* gene and a much simpler body plan seems a good starting point. Much of our understanding about the genetic networks involving *Dll* comes from studies done on *Drosophila*.

Dll is the earliest known genetic marker for limb formation and is dynamically expressed throughout the development of limbs in *Drosophila* (Cohen, Bronner et al. 1989; Panganiban 2000).

Expression and function of *Dll* in the developing leg of *Drosophila* has been extensively studied and suggests a role in patterning of the proximo-distal axis of the leg, where cells strongly expressing *Dll* take distal identity. Some of the evidence for this comes from viable combinations of *Dll* alleles of increasing severity. Weak allelic combinations lead to fusion of distal leg segments, intermediate combinations show loss of distal leg segments, whereas strong combinations lead to loss of both distal and medial leg segments (Sunkel and Whittle 1987; Cohen, Bronner et al. 1989; Panganiban and Rubenstein 2002).

The appendages (antennae, mouth parts, halteres, legs, analia) of *Drosophila* originate from a small group of cells known as the imaginal discs. Each of these imaginal discs initially comprises of 10 to 30 cells (Panganiban 2000). Although initially *Dll* can be seen expressed in the imaginal disc in both the proximal and the distal regions, the expression later gets restricted to more distal regions as the limb develops (Cohen, Bronner et al. 1989). Using mitotic clonal analysis, *Dll* mutant cells were generated in the distal leg and were found to separate from the disc epithelium and form small vesicles (Gorfinkiel, Morata et al. 1997; Wu and

Cohen 1999). Later these cells were found to migrate towards the proximal regions of the leg, implicating *Dll* in cell adhesion in the distal leg (Wu and Cohen 1999).

In the antenna, *Dll* has been shown to specify the proximo-distal identity (Sunkel and Whittle 1987; Cohen, Bronner et al. 1989). Expression of *Dll* is also seen in other appendages like the mouth parts and analia (Panganiban and Rubenstein 2002). In vertebrates, *Dlx* is implicated in development of olfactory placode and otic vesicle which are homologous (at least in function) to *Drosophila* antenna (functions as ears and nose of fly) and mouth parts, suggesting similar, conserved roles for *Distal-less*.

In addition to appendages, *Dll* expression in flies is also seen in the peripheral nervous system (Panganiban and Rubenstein 2002). *Dll* null mutants show absence of antennal, maxillary and labial sense organs (Cohen, Bronner et al. 1989). *Dll* null flies also show complete lack of the mechano-sensory vestigial organ called the Keilin's organs (Panganiban 2000).

Again, as in vertebrates *Dll* expression in flies can be seen in the central nervous system, specifically in the optic lobes and in the glia of ventral nerve cord (Cohen, Bronner et al. 1989). In vertebrates, *Dlx* seems to be involved in differentiation and migration of specific neuronal cell types in the ventral forebrain.

1.2. Regulation of *Dll* expression in Drosophila.

In *Drosophila* embryo, *Dll* expression requires activity of Wg (Wingless, member of Wnt family) and is repressed by Bone Morphogentic Protein (Bmp), Decapentaplegic (Dpp) and by Epidermal growth factor (Egf) signaling pathways (Panganiban 2000).

In contrast, during larval stages, *Dll* expression is positively controlled by both Dpp and Wg signaling (Diaz-Benjumea, Cohen et al. 1994; Neumann and Cohen 1997). In addition *Dll* is also known to help maintain its own expression through positive auto-regulatory inputs (Lecuit and Cohen 1997).

Dll, when expressed ectopically in proximal regions of ventral appendages, results in duplication of legs and antenna. In the abdomen *Dll* is negatively regulated by two homeotic genes, *Ultrabithorax* and *Abdominal-A*, which interact with a distal enhancer of *Dll* to repress appendage development (Vachon, Cohen et al. 1992).

1.3. Vertebrate *Dlx* genes: Genomic organization and Evolution.

Most vertebrates possess at least six *Dlx* genes, forming six orthologous groups. These six genes are arranged as three convergently transcribed bigene clusters on distinct chromosomes (Figures 1.1 & 1.2). These *Dlx* bigenes are separated by a relatively short intergenic region with sizes ranging from 3kb in fugu to 15kb in humans (Ghanem, Jarinova et al. 2003). These six *Dlx* genes can also be organized into two subfamilies or clades based on the homeodomains they code (*Dlx* 1,4,6 clade and *Dlx* 2,3,5 clade). Each of the bigene clusters contains one member from each of the two subfamilies. This classification led to the suggestion that all vertebrate *Dlx* genes are derived from an ancestral *Dlx* bigene cluster (Stock, Ellies et al. 1996; Quint, Zerucha et al. 2000; Zerucha and Ekker 2000) or in other words, the tandem duplication that formed the *Dlx* bigene cluster was a single event that happened before the divergence of vertebrates. This indeed seems to be the case, as the sea squirt *Ciona intestinalis* has three *Dlx* genes two of which are arranged in the same bigene organization as found in invertebrates. In this thesis, I show that lamprey an agnathan (jawless vertebrate) also has two of its *Dlx* genes in the same bigenic configuration.

Figure 1.1: Genomic organization and position of mouse *Dlx* genes along with the conserved intergenic cis-regulatory elements (enhancers).

Blue ellipses are regulatory elements. Brown blocks are *Dlx* genes belonging to the *Dlx1, 6, 4* clade and grey blocks are *Dlx* genes from the *Dlx2, 3, 5* clade.

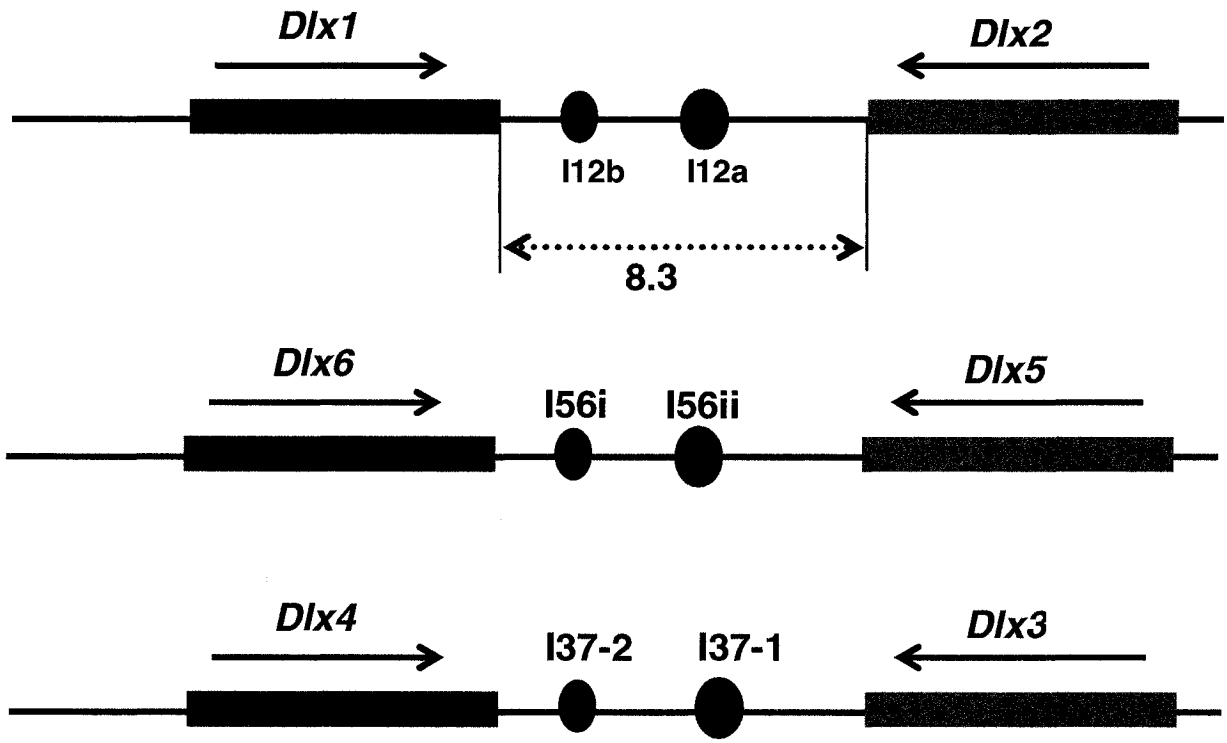
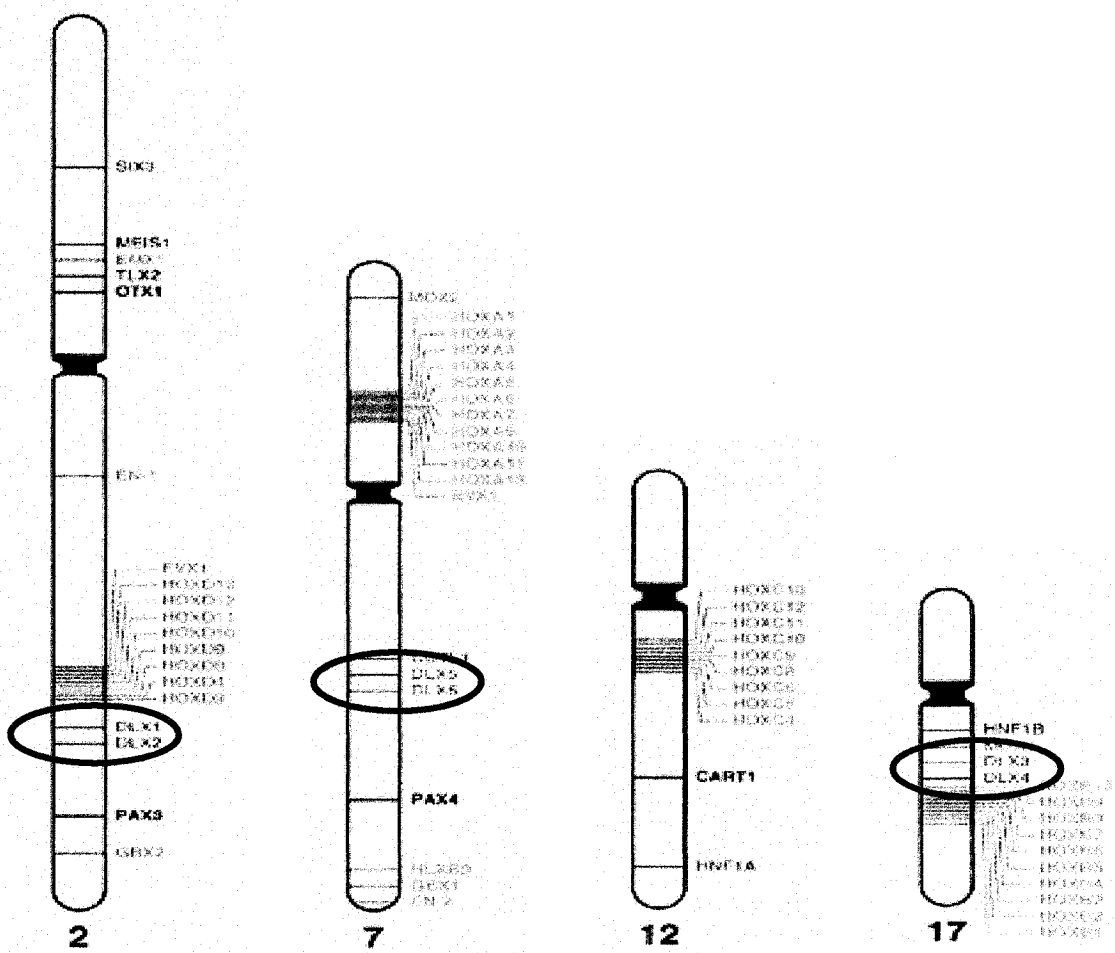

Dlx1
Dlx2
I12b
I12a
8.3
Dlx6
Dlx5
I56i
I56ii
Dlx4
Dlx3
I37-2
I37-1

Figure 1.2: Position and orientation of *Hox* clusters with that of *Dlx* bigenes on human chromosomes.

Dlx bigenes are shown in pink and are encircled, whereas *Hox* clusters are shown in brown (modified from (Banerjee-Basu and Baxevanis 2001).


SIX3
MEIS1
TLX2
OTX1
EN-1
PAX3
GBX2
2
DLX5
DLX6
PAX4
7
CART1
HNF1A
12
HNF1B
DLX3
DLX4
17

To track the genome duplication and gene loss events that led to the present set of *Dlx* genes in vertebrates, linkage of *Dlx* genes to that of *Hox* gene clusters has been utilized. In *C. elegans* (considered ancestral state, see Figure 1.3), the single *Dll* gene is present on the same chromosome as that of the single Hox gene cluster. In vertebrates too, all the *Dlx* clusters studied so far are distantly linked to a Hox gene cluster (*Dlx*1/2 to *Hoxd*, *Dlx*5/6 to *Hoxa*, *Dlx*3/7 to *Hoxb* (Rossi, Faiella et al. 1994; Simeone, Acampora et al. 1994; Nakamura, Stock et al. 1996; Stock, Ellies et al. 1996; Amores, Force et al. 1998)). Adjacent duplication of the single ancestral *Dlx* gene, followed by two rounds of genome duplication and a subsequent loss of the *Dlx* pair linked to *Hoxc* could account for the present complement of mammalian *Dlx* genes (Ellies, Stock et al. 1997; Neidert, Virupannavar et al. 2001). In the lineage leading to teleost fish, there seems to be an additional genome duplication event that took place after their divergence from lobe-finned fish, which led to two additional *Dlx* genes in zebrafish and fugu (Figure 1.3). One of the two unlinked *Dlx* genes in zebrafish is distantly linked to *Hox* cluster and is believed to have generated in the same duplication event that led to the additional *Hox* clusters in teleosts (Amores, Force et al. 1998).

In *Drosophila* the single *Dll* gene is not linked to the *Hox* cluster and it is thought that this may be a result of translocation of the *Dll* gene from the chromosome containing the *Hox* cluster to the present location.

Figure 1.3: Proposed model for the evolution of *Distal-less*-related genes.

The number and genomic organization of *Distal-less* related genes is shown in different metazoan species. Wherever known their linkage to a *Hox* cluster is displayed. Orange ovals represent enhancer elements found in the intergenic region of vertebrate *Dlx* bigenes. The cross mark where ever seen shows *Distal-less* is not linked to a *Hox* cluster. The orientation of *Hox* and *Dlx* clusters to each other is not exact.

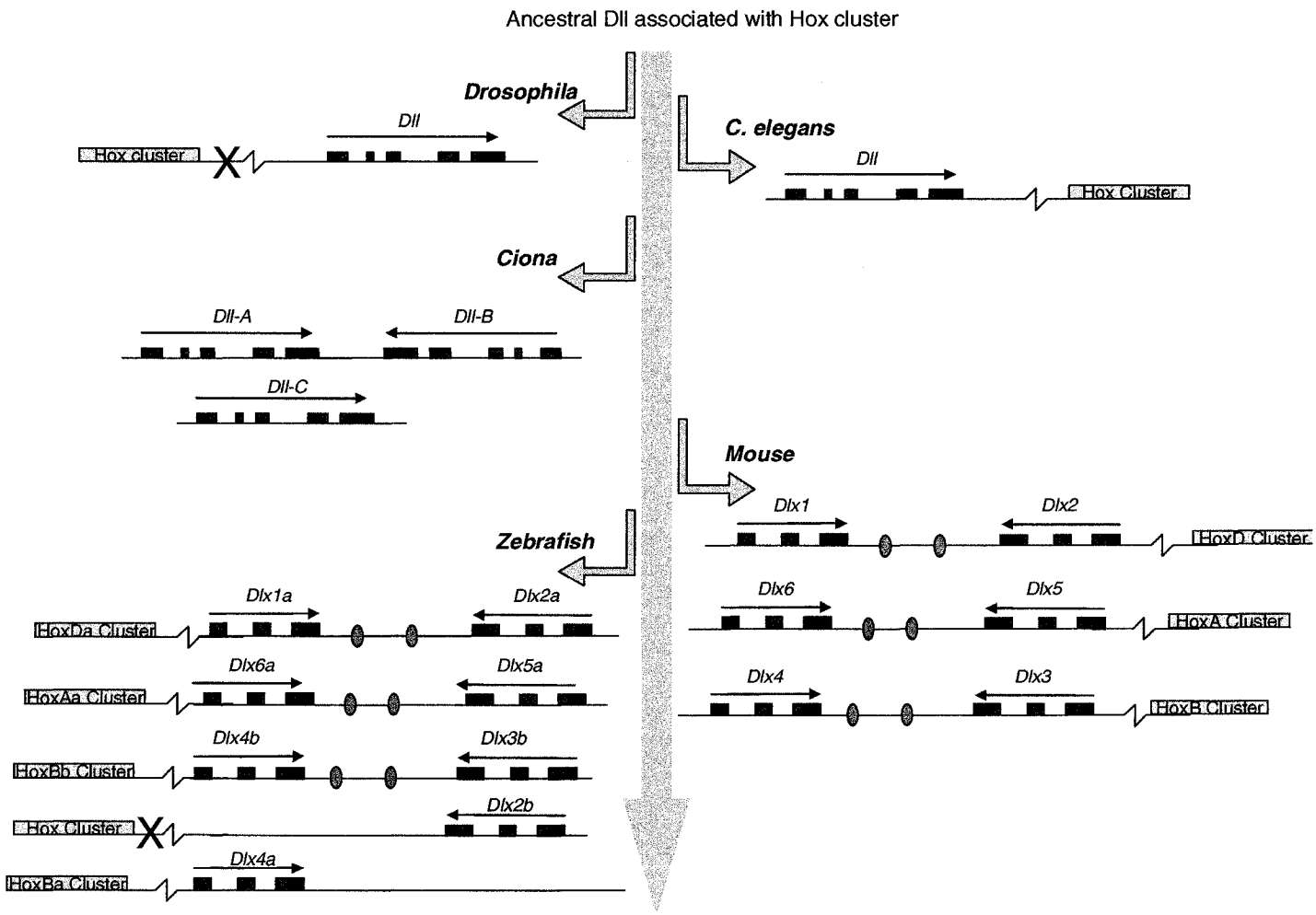
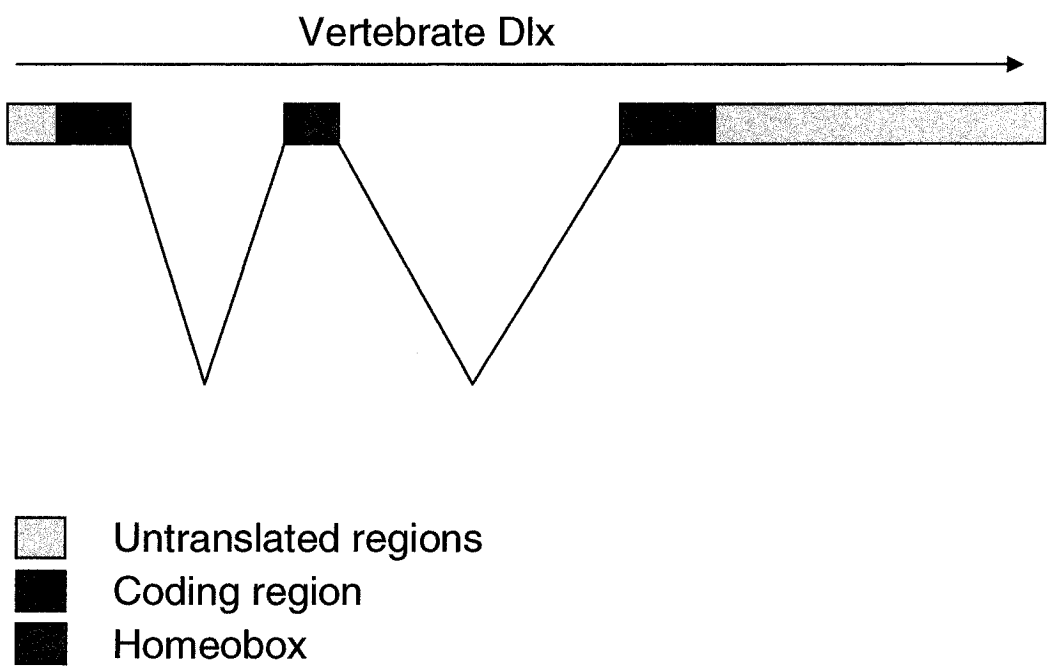

Ancestral Dll associated with Hox cluster
Drosophila
Dll
Hox cluster
C. elegans
Dll
Hox Cluster
Ciona
Dll-A
Dll-B
Dll-C
Mouse
Dlx1
Dlx2
HoxD Cluster
Zebrafish
Dlx1a
Dlx2a
HoxDa Cluster
Dlx6
Dlx5
HoxA Cluster
Dlx6a
Dlx5a
HoxAa Cluster
Dlx4
Dlx3
HoxB Cluster
Dlx4b
Dlx3b
HoxBb Cluster
Dlx2b
Hox Cluster
Dlx4a
HoxBa Cluster

Figure 1.4: Structure of a typical vertebrate *Dlx* gene (human *Dlx1* shown here).

All vertebrate *Dlx* genes examined so far seem to have three exons and two introns with conserved splice sites. The homeobox sequence (red region) is split between the 2nd and 3rd exon.


Vertebrate Dlx
Untranslated regions
Coding region
Homeobox

Exon-Intron organization of vertebrate *Dlx* genes: Most vertebrate *Dlx* genes have three exons interrupted by two introns (see Figure 1.4). In all vertebrate *Dlx* genes, the second and third exons are interrupted within the homeobox at the codons coding for the 46th and the 47th amino acid in the homeodomain (TQTQ <intron> VKIW). Similarly, the first and second exons are interrupted by an intron at the codon coding for the 24th amino acid upstream of the homeodomain. The homeodomain is used as reference because it is a feature common to all *Dlx* genes and is the most conserved part of the protein. Both the first and the third exons in all vertebrate *Dlx* genes contain untranslated regions.

1.4. Vertebrate *Dlx* gene expression and function.

All six *Dlx* genes in vertebrates show highly overlapping patterns of expression in the forebrain, branchial arches, otic vesicle, olfactory placode and in the limbs. The highest degree of overlap in expression is seen amongst physically linked *Dlx* genes. For example, expression of *Dlx3a* and *Dlx4a* is almost indistinguishable throughout zebrafish development (Ellies, Stock et al. 1997). Even though expression of linked *Dlx* genes is highly similar, there do exist documented differences. For example, the zebrafish migrating neural crest cells from the hindbrain express *Dlx2* and not the linked *Dlx1* (Akimenko, Ekker et al. 1994; Ellies, Stock et al. 1997).

The earliest expressing *Dlx* genes are *Dlx3b* and *Dlx4b* in zebrafish development (Akimenko, Ekker et al. 1994; Ellies, Stock et al. 1997). They are expressed in two stripes of cells extending bilaterally on either side of the midline during gastrulation. This region of expression later gets restricted to the prospective olfactory placode and otic vesicle (Akimenko, Ekker et al. 1994; Ellies, Stock et al. 1997).

Although there exist minor documented differences in expression of orthologs in various species, only the common features will be discussed here for the sake of simplicity and relevance.

1.4.1. Forebrain:

In mice, *Dlx* genes are expressed mainly in the ectodermal derivatives like nervous system and surface ectoderm. In the neural tube, their expression is highly restricted to the forebrain (Price, Lemaistre et al. 1991; Robinson, Wray et al. 1991; Dollé, Price et al. 1992; Bulfone, Kim et al. 1993; Simeone, Acampora et al. 1994; Yang, Zhang et al. 1998; Eisenstat, Liu et al. 1999). Four out of the six mouse *Dlx* genes (*Dlx*1/2/5/6) are expressed in two domains of the forebrain, i.e. telencephalon and diencephalon. *Dlx3* and *Dlx4* have not been found in the forebrain of any species examined so far (zebrafish and mice) (Zerucha and Ekker 2000).

Within the forebrain, *Dlx* genes seem to be expressed in a temporal sequence which suggests regulatory control between the *Dlx* genes themselves. *Dlx2* can be detected earliest in subsets of ventricular neuroepithelium cells, which is followed by *Dlx1*, *Dlx5* and then *Dlx6* in the most subventricular zone cells (Liu, Ghattas et al. 1997; Eisenstat, Liu et al. 1999; Zerucha, Stuhmer et al. 2000).

Previously, our lab has shown that I56i and I56ii enhancer elements (highly conserved enhancer elements in the *Dlx5/6* intergenic region, see Figure 1.1) are sites of this cross-regulatory interaction. Both of these elements enhanced transcription of a reporter gene in the presence of *Dlx2* protein in cell culture (Zerucha, Stuhmer et al. 2000) giving support to the temporal regulatory model suggested between the four *Dlx* genes expressed in the forebrain.

1.4.2. Branchial Arches:

All six mammalian *Dlx* genes are expressed in the branchial arches in complex spatio-temporal fashion (Dollé, Price et al. 1992; Bulfone, Kim et al. 1993; Akimenko, Ekker et al. 1994; Robinson and Mahon 1994; Simeone, Acampora et al. 1994; Qiu, Bulfone et al. 1997; Yang, Zhang et al. 1998). They are expressed in the ectomesenchyme derived from migratory cranial neural crest cells, which later populates the branchial arches to give rise to most of the facial tissues (skeletal and connective) (Depew, Lufkin et al. 2002). Within the branchial arches, expression of *Dlx* genes seems to occur in nested fashion which suggests their role in proximo-distal (PD) axis formation of the arches (Panganiban and Rubenstein 2002). The *Dlx1/2* pair is expressed all along the PD axis of the branchial arches, *Dlx5/6* is expressed midway to the distal tip and *Dlx3/4* expression is restricted to most distal regions of the PD axis. In addition they seem to be following the same temporal sequence in expression as seen in the forebrain (*Dlx2*, *Dlx1*, *Dlx5* and then *Dlx6*) (Panganiban and Rubenstein 2002). Later in development *Dlx* is expressed in differentiating skeletal tissues. *Dlx* genes are also expressed in both the mesodermal and ectodermal parts of the developing teeth (Thomas, Tucker et al. 1997; Zhao, Stock et al. 2000; Depew, Lufkin et al. 2002).

Other regions of *Dlx* gene expression include the neuronal precursors and subsets of neurons in the developing retina, appendages like apical ectodermal ridge of limb bud and the genital eminence.

1.5. *Dlx* gene functions as inferred by gene inactivation.

Targeted mutations in the mouse *Dlx1* gene lead to abnormal enteric nervous system and minor defects in neural crest derived skeletal components in the proximal regions of the first and second branchial arches. *Dlx2* mutant mice

show deformations in tissues derived from the first and second branchial arches. They also show abnormal differentiation of the olfactory bulb neurons. Consistent with the temporal regulatory model suggested for the *Dlx* genes (In the order *Dlx2, 1, 5, 6*) *Dlx1* expression is reduced in *Dlx2* mutant mice in the first 2 branchial arches and in the medial ganglionic eminence.

Paralogous genes with highly overlapping patterns of expression like *Dlx* present a problem in inferring their individual roles by targeted mutations because of partial functional redundancy or compensatory effects from paralogs, when one of them is knocked out.

The partial and mild phenotypes of single *Dlx* genes in mice led researchers to make mice that were deficient in both *Dlx1* and *Dlx2*. This double mutant mice show more profound dysmorphologies of the craniofacial bones and complete lack of maxillary molars. In the forebrain of these mice there are seen striking abnormalities in formation of striatal subventricular zone neurons, differentiation of striatal matrix neurons and migration of neocortical interneurons from the subcortical telencephalon. There is also seen a complete loss of *Dlx5,6* expression in the forebrain of *Dlx1/2* double mutant mice, consistent with the temporal regulatory model suggested amongst the *Dlx* genes. Even though *Dlx1* and *Dlx2* are expressed in the distal regions of the branchial arches, the derived structures are unaffected in the double mutant. Compensatory activities from products of other distally expressed *Dlx* genes (*Dlx3, 5, 6, 4*) have been suggested for the lack of distal phenotype.

Dlx5 mutant mice display multiple defects in craniofacial structures, severe malformation of vestibular structure, proximal mandibular arch skeleton and hypoplasty of nasal capsules (Acampora, Merlo et al. 1999; Depew, Liu et al. 1999). A subset of *Dlx5* mutant mice also develop exencephaly. No apparent

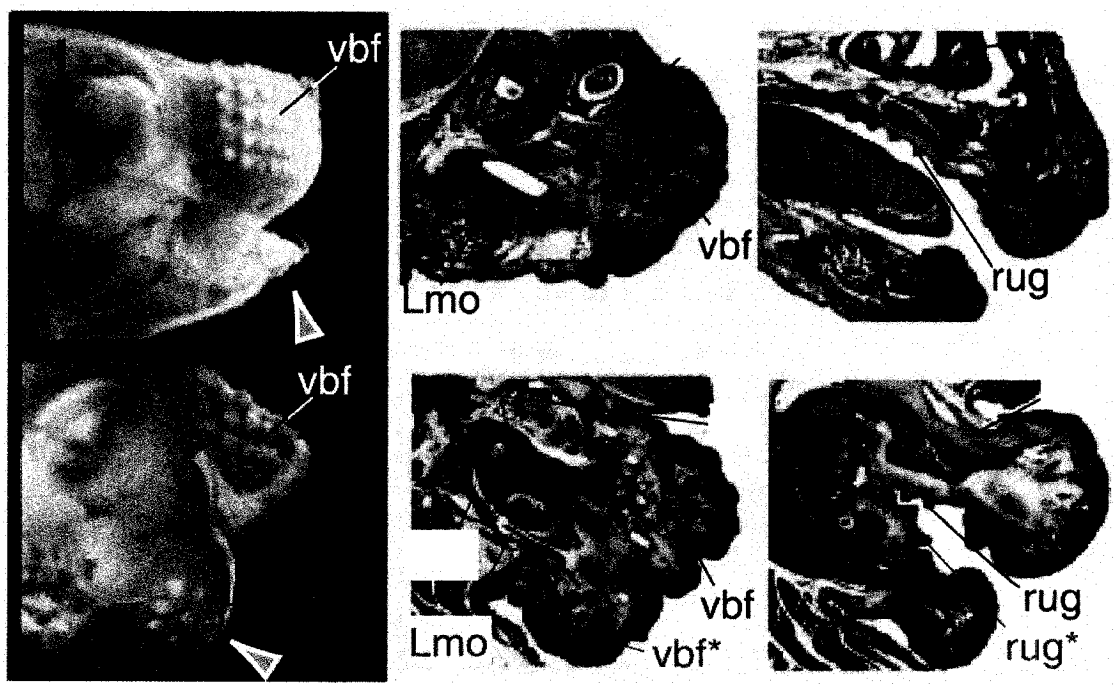
morphological phenotype was observed in the forebrain, again suggesting compensatory function by other *Dlx* genes.

Dlx5/6 double mutant mice show multiple craniofacial and ear defects, these mice fail to form Meckel's cartilage, mandible and calvaria (Robledo, Rajan et al. 2002). The craniofacial and ear defects observed in *Dlx5/6* double mutant are much more severe than those observed in *Dlx5*-deficient mice, suggesting that *Dlx5* and *Dlx6* have both unique and redundant functions. In addition to these dysmorphologies, the *Dlx5/6* double mutant mice display a homeotic mirror image transformation of the lower jaw into an upper jaw (Depew, Lufkin et al. 2002). In the double mutant one can easily see the set of whiskers appearing on the lower jaw, which are only present on the upper jaw in wild-type mice (see Figure 1.5). More detailed analysis by RNA *in-situ* hybridization for molecular markers of the lower jaw confirms this homeotic transformation.

It still remains unclear if the more severe phenotypes observed in the double knockouts (when compared to single knockouts) are mainly due to overall quantitative reduction in *Dlx* protein levels or are there functional differences involved between paralogs.

Figure 1.5: Wild-type (upper) and *Dlx5/6-/-* mutant (lower) neonates.

Superficial ectoderm removed (Left panel); sectioned (middle and right panel), reveal proximal morphological features duplicated on the lower jaw [vibrissae (compare arrowheads) and rugae]. Lmo, lower molar; rug, rugae; vbf, vibrissae follicle; modified from (Depew, Lufkin et al. 2002).


vbf
vbf
Lmo
vbf
rug
Lmo
vbf
vbf*
rug
rug*

1.6. Evolution of regulatory mechanisms responsible for *Dlx* gene expression.

Several cis-acting regulatory elements have been identified around the *Dlx* bigene clusters in vertebrates based on high sequence similarity between mice, zebrafish and humans. At least two regions of remarkable conservation amongst vertebrates are found in each of the three intergenic regions (I12a, I12b in *Dlx1/2* bigene and I56i, I56ii in *Dlx5/6* bigene (Ghanem, Jarinova et al. 2003)and I37-1, I37-2 in *Dlx3/4* bigene (see Figure 1.1) (Sumiyama and Ruddle 2003)). These relatively short regions (300-500bp in size), when taken out of context and placed in front of a reporter gene, with a minimal promoter can target robust expression of the reporter gene in developing mice and zebrafish embryos, in patterns very similar to that of endogenous *Dlx* expression (Ghanem, Jarinova et al. 2003). From these reporter gene assays we know that these fragments work non-directionally and act as enhancers of transcription.

In all the vertebrate species examined so far (mice, human, zebrafish, fugu and Xenopus), the location and orientation of these intergenic enhancers in the bigenic clusters has been very well preserved. Paralogous enhancer elements (in terms of their location with respect to *Dlx* paralogs) in the bigenes show similar activities, again suggesting common ancestry of these sequences from a single ancestral bigenic cluster in the vertebrate ancestor (Zerucha, Stuhmer et al. 2000; Ghanem, Jarinova et al. 2003).

These intergenic enhancers when compared between vertebrate species (like zebrafish, fugu, mice and humans) show an even higher degree of sequence conservation than the homeoboxes of the flanking *Dlx* genes. Some of them are also among 481 "ultraconserved elements"(completely conserved blocks >200bp in size) in mammalian genomes (Bejerano, Pheasant et al. 2004) and have been

identified in numerous studies that try to identify regions of the human genome that are evolutionarily most constrained (Woolfe, Goodson et al. 2005).

1.6.1. Regulatory elements in the *Dlx*1/2 locus:

Within the *Dlx1/2* locus our lab had previously identified four enhancer elements common to all vertebrate species examined so far (fugu, zebrafish, mice, humans and Xenopus). Two of the most conserved among these four regions are the intergenic enhancers I12a and I12b (see Figure 1.6). The other two regions lie upstream of *Dlx1* and are called URE1 and URE2 (URE for upstream regulatory element).

I12a is about 550bp and has about 92% sequence identity in humans and zebrafish. When placed close to a reporter gene with a beta-globin minimal promoter, mice I12a can target expression to a subset of mesenchymal cells in mandibular of first and second branchial arches during development.

I12b is about 400 bp and shows about 78% sequence identity in humans and zebrafish. When mouse I12b is tested in an enhancer assay, it targets expression to the two domains in the forebrain, namely telencephalon and diencephalon, faithfully mimicking the endogenous *Dlx* expression.

Combined activity of both intergenic enhancer elements is shown in an E11.5 (embryonic day 11.5) mouse embryo, which has the *lacZ* gene under the control of the betaglobin minimal promoter and complete intergenic region (see Figure 1.7).

URE1 targets expression to neuronal precursors in the eye and URE2 targets expression to both forebrain and branchial arches (both of the enhancers are currently being characterized).

Figure 1.6: Conserved intergenic enhancer elements in the *Dlx1/Dlx2* intergenic region.

Schematic representation of the *Dlx1/Dlx2* intergenic region of five vertebrate species. The third exons of the *Dlx* genes are indicated. In addition to the I12a and I12b sequences, ovals labeled "c" represent a region of sequence conservation between the three teleost fish species (Ghanem, Jarinova et al. 2003).

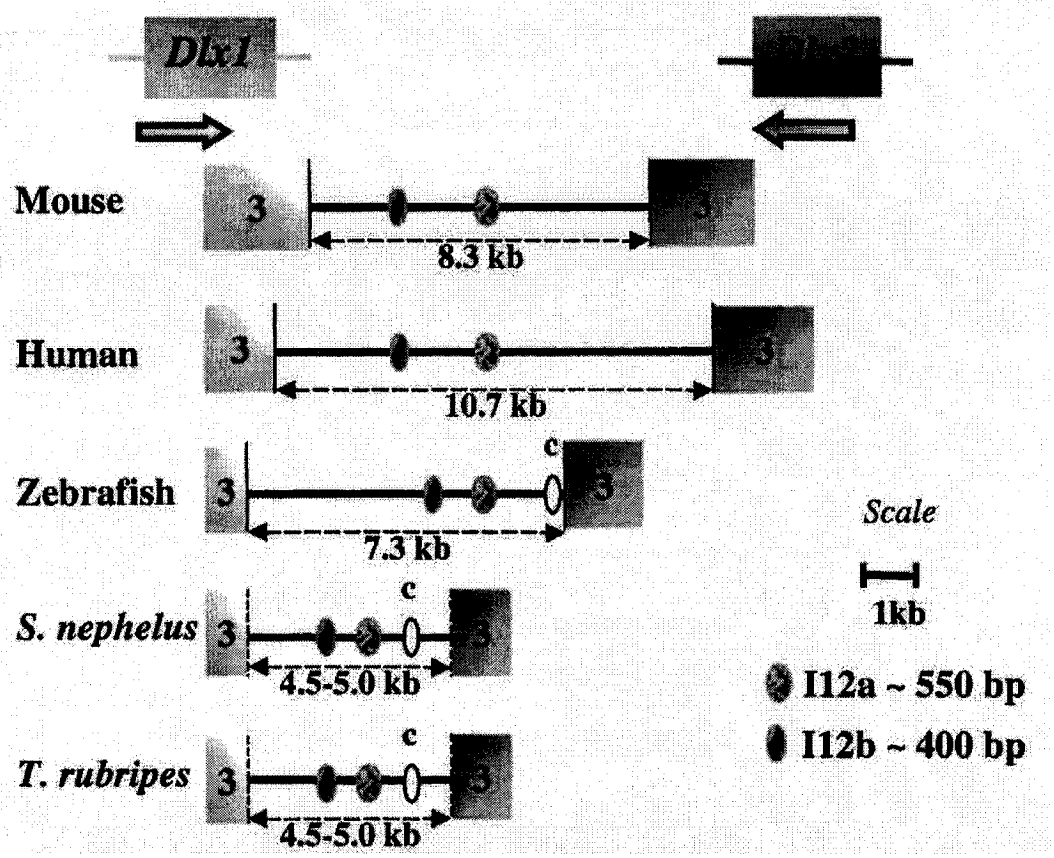
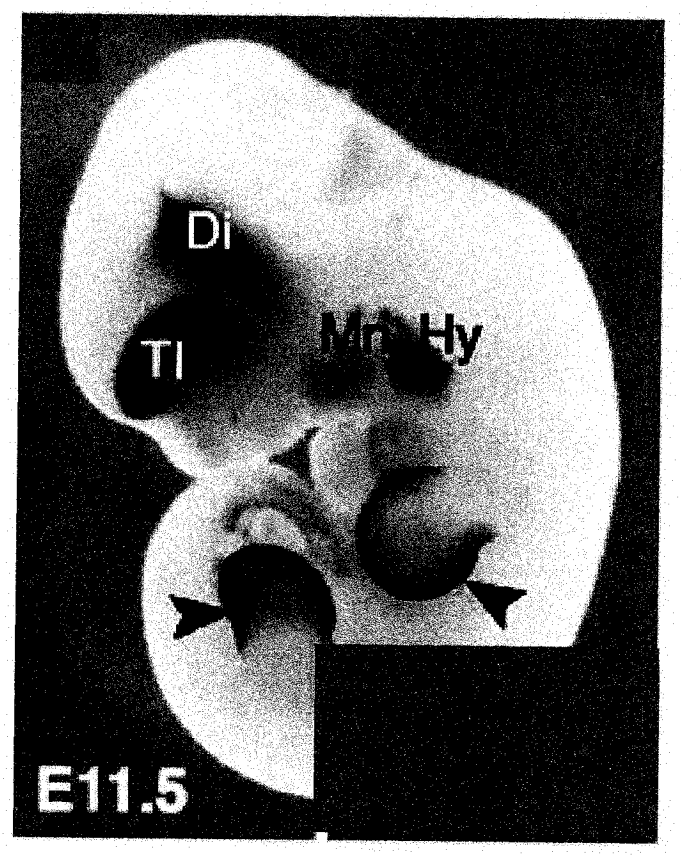

Dlx1
Mouse
3
8.3 kb
Human
10.7 kb
Zebrafish
c
7.3 kb
S. nephelus
4.5-5.0 kb
T. rubripes
Scale
1kb
I12a ~ 550 bp
I12b ~ 400 bp

Figure 1.7: Whole-mount beta-galactosidase staining of an E11.5 mouse transgenic embryo with a 13.5-kb lacZ reporter transgene containing the entire *Dlx1/2* intergenic region.

Expression is seen in diencephalon (Di) and telencephalon (Tl) of the forebrain, mandibular (Mn) and hyoid (Hy) branchial arches, and the limb bud apical ectodermal ridge (AER; arrowheads) (Park, Sperber et al. 2003).


Di
Tl
Hy
E11.5

1.6.2. Regulatory elements in the *Dlx5/6* locus:

Similarly, based on sequence comparison from a number of vertebrate species, two enhancer elements (I56i, I56ii) were found in the *Dlx5/6* locus. Both of these lie in the intergenic region between the bigene (see Figure 1.8).

I56i is about 440bp and shows 81% sequence identity between humans and zebrafish. When mice I56i is tested in an enhancer assay in mice, it targets the expression of reporter gene specifically to the forebrain (telencephalon, diencephalon) and the branchial arches (mandibular portion of the first and second arch) at E11.5 (see Figure 1.9).

I56ii is about 310bp long and displays 84% sequence identity between humans and zebrafish. In an enhancer assay, mice I56ii targets reporter gene expression to the two domains of the forebrain, telencephalon and diencephalons (see Figure 1.9).

Figure 1.8: Conserved sequences in the *Dlx5/Dlx6* intergenic region.

Schematic representation of the *Dlx5/Dlx6* intergenic region of five vertebrate species. The third exon of the *Dlx* genes are indicated. In addition to the I56i and I56ii sequences, ovals labeled iii, iv, and v represent regions of sequence conservation between a subset of the five species (Ghanem, Jarinova et al. 2003).

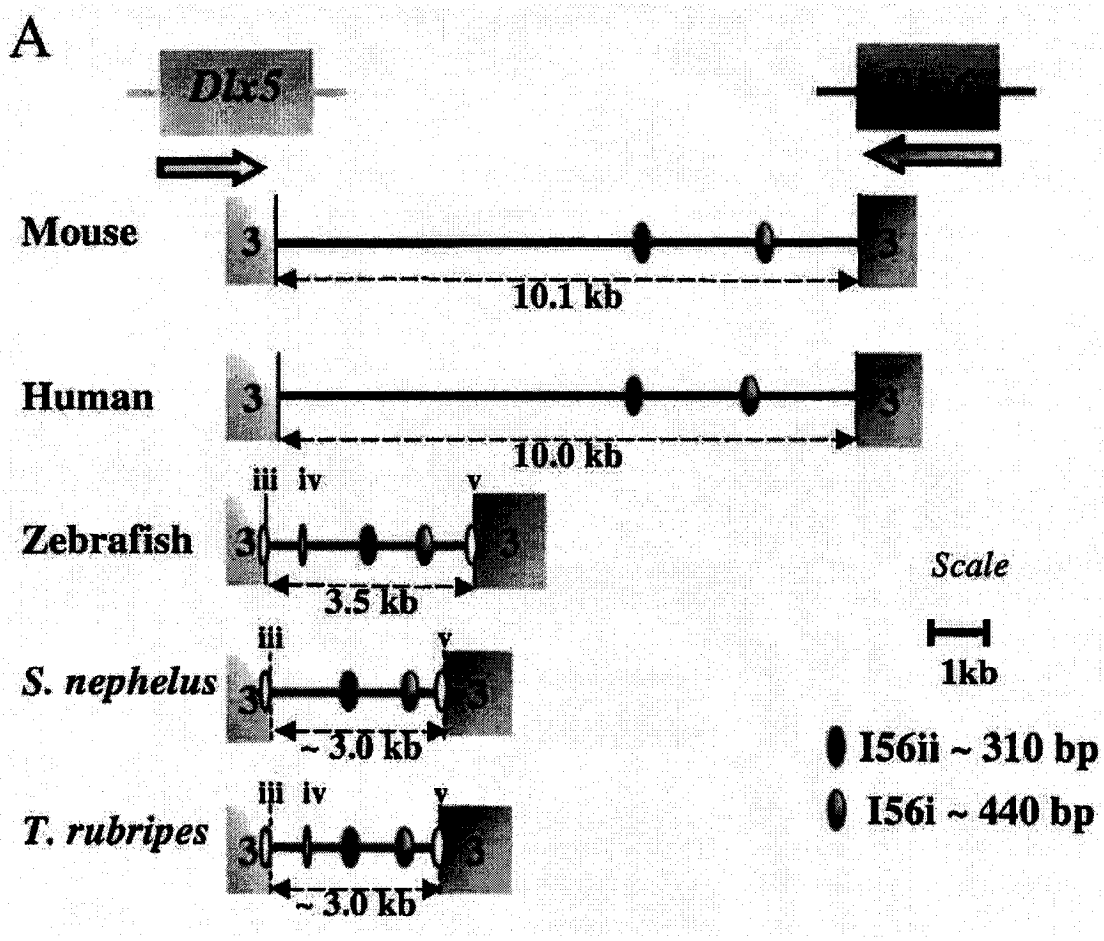
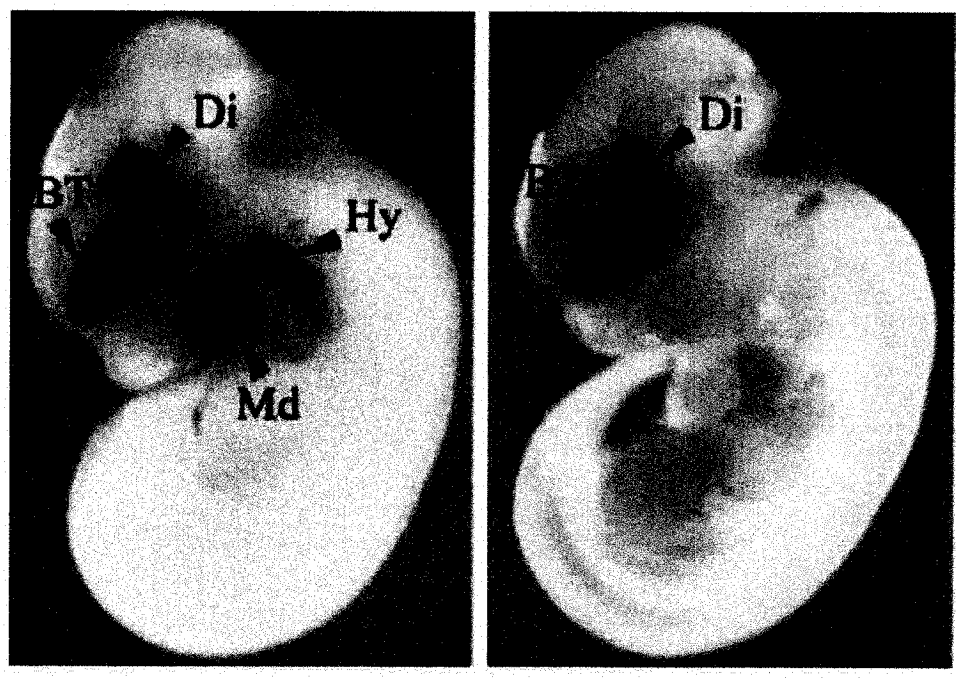

A
Dlx5
Mouse
3
3
10.1 kb
Human
3
3
10.0 kb
iii iv v
Zebrafish
3
3
3.5 kb
iii v
S. nephelus
3
3
~ 3.0 kb
iii iv v
T. rubripes
3
3
~ 3.0 kb
Scale
1kb
I56ii ~ 310 bp
I56i ~ 440 bp

Figure 1.9: Whole-mount beta-galactosidase staining of E11.5 mouse transgenic embryos with a lacZ reporter transgene containing I56i (left) and I56ii (right) enhancer elements.

Expression is seen in diencephalon (Di) and basal telencephalon (BT) of the forebrain, mandibular (Md) and hyoid (Hy) branchial arches. Modified from (Ghanem, Jarinova et al. 2003).


Di
Hy
Md
Di

1.6.3. Regulatory elements in the *Dlx3/4* locus:

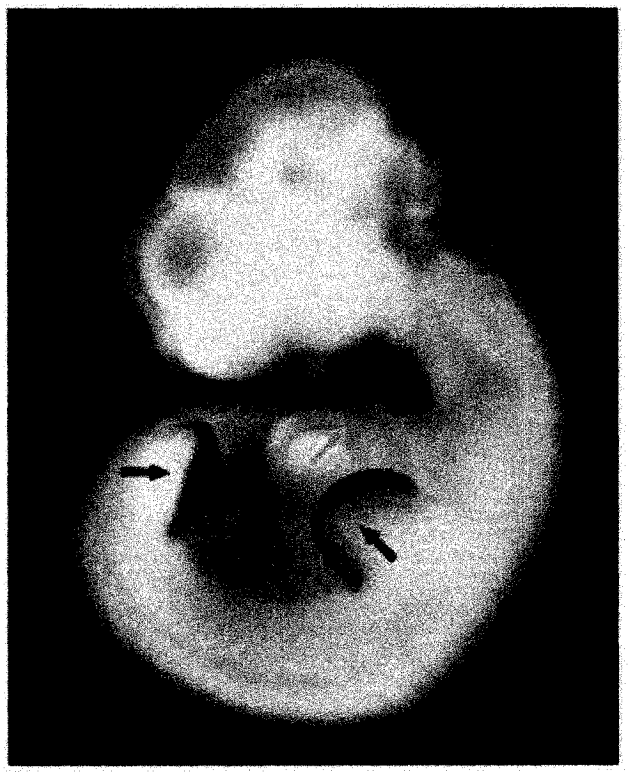
In the *Dlx3/4* locus, two regions of high sequence conservation are found in the intergenic region, the degree of conservation is much lower (90% and 88% sequence identity for I37-1, I37-2 between mice and humans) than that for the elements in the *Dlx1/2* and *Dlx5/6* locus (97-99% sequence identity for the four elements between mice and humans). The activity of these two fragments has not been tested individually, but activity of the whole *Dlx3/4* locus when tested in mice can target expression to branchial arches and the apical ectodermal ridge (AER, see Figure 1.10).

The relative position and orientation of all enhancers discussed above (in all three *Dlx* bigene clusters) are conserved in all vertebrates examined so far. Paralogous enhancer elements (in terms of location with respect to *Dlx* paralogs) show very little sequence conservation, which means soon after the acquisition of these elements they functionally diverged from each other (Ghanem, Jarinova et al. 2003).

Despite the high degree of sequence conservation in all vertebrates (humans, zebrafish, mice, rat, fugu, opossum, dog and chimp) none of the vertebrate *Dlx* enhancers can be identified in invertebrates like *Drosophila* based on sequence comparison. The sea squirt *Ciona*, which is closer to vertebrates than *Drosophila*, has a *Dll* bigene cluster in its genome, but none of these enhancers are identifiable in its intergenic region, leaving the evolutionary origin of these highly conserved vertebrate sequences obscure. Therefore, to determine the origin of these enhancers during vertebrate evolution, we are investigating *Dlx* regulation in lamprey, which is considered a vertebrate ancestor and is closer to vertebrates than *Ciona*. More specifically, this study will let us know if these highly conserved vertebrate sequences were acquired before or after the divergence of agnathans (jawless vertebrates) from the rest of vertebrates.

Figure 1.10: Whole-mount beta-galactosidase staining of an E11.5 mouse transgenic embryos with a *lacZ-Dlx3* reporter fusion construct containing a large region of the *Dlx3/7* locus.

This *Dlx3/7* locus of size 79kb includes the complete *Dlx3, Dlx7* genes and the intergenic region. Expression is seen in the first, second branchial arches and in the apical ectodermal ridge (arrows) (Sumiyama, Irvine et al. 2002).



1.7. Lampreys and Evolution of Jaws.

Lampreys are extant jawless vertebrates (agnathans, sister group of gnathostomes or jawed vertebrates) easily identified by their characteristic oral apparatus and simple axial morphology. Lampreys display many unique and many shared morphological characteristics when compared with gnathostomes. Morphological features like true bones, true teeth, paired fins with pectoral and pelvic girdles and most importantly jaws were acquired by vertebrates after they diverged from lampreys. Lampreys also share a number of morphological features with that of gnathostomes, like multiple brain divisions, neural crest and its derivatives, branchial arches, and a carliginous endoskeleton. Because of the shared, derived characteristics of lampreys compared to gnathostomes, lampreys have long been considered as the best living proxy of the vertebrate ancestor (Neidert, Virupannavar et al. 2001).

Investigating *Dlx* regulation in lamprey is therefore interesting in two ways, first, it will help us trace back the origin of the highly conserved *Dlx* intergenic enhancers in gnathostomes, and second, may provide us clues about the evolution of some of the gnathostome specific morphological features.

1.8. Lamprey *Dlx* genes and their expression patterns.

A study by Neidert et. al. identified four *Dlx* genes in sea lamprey (*Petromyzon marinus*) (Neidert, Virupannavar et al. 2001). They cloned and sequenced four *Dlx* cDNA clones and performed phylogenetic analysis of the obtained sequences. None of the four identified lamprey *Dlx* genes form clear orthologs to the six individual *Dlx* genes of other vertebrates. Therefore, they were named *DlxA, B, C* and *D*. The four lamprey *Dlx* genes do segregate into the two clades identified in vertebrates. *DlxA, B, C* fall into the *Dlx 2, 3, 5* clade whereas *DlxD* falls into the *Dlx 1, 7, 6* clade. This suggested that the lamprey genome may also contain a *Dlx* bigene cluster (as genes in all bigene clusters have one member

from each clade of *Dlx* genes). *Dlx* bigene clusters are found in all vertebrates examined so far and are assumed to have originated before the divergence of all vertebrates. This phylogenetic analysis also suggested that the genome duplication events that gave rise to lamprey and gnathostome *Dlx* gene complements occurred independently.

Neidert et. al. also performed RNA in-situ hybridization with the four identified *Dlx* genes on lamprey embryos (Neidert, Virupannavar et al. 2001). *DlxA* and *DlxD* transcripts were detected in the cranial neural crest cells (see Figure 1.11). They were also found throughout the dorsal part of the cranial neural tube. *DlxA, C, D* are all expressed in the migratory cranial neural crest. *Dlx A, C, D* are also seen expressed in the olfactory placode and otic vesicle (see Figure 1.12).

DlxA, D expression persists from the migratory and pre-migratory neural crest cells into the branchial arches (migratory neural crest cells populate the branchial arches to later form craniofacial structures within the branchial arches) (Neidert, Virupannavar et al. 2001). *DlxC* on the other hand is expressed only in the migratory neural crest but its expression also persists in the branchal arches. Within the branchial arches, the four lamprey *Dlx* genes (*DlxA, B, C & D*) are expressed all along the proximodistal axis, unlike gnathostomes, where an overlapping, nested expression is seen for the three *Dlx* bigenes. Similar to gnathostomes, *Dlx* gene expression within the branchial arches is seen in regions where future cartilage condensations would occur.

Figure 1.11: Lamprey *Dlx* gene expression in the cranial neural crest.

Whole mounts and transverse sections of the head regions of 9-day-old lamprey embryos labeled with *DlxA*, *DlxC*, and *DlxD* riboprobes (*DlxB* expression is not detected at this stage). *DlxA* and *DlxD* are both expressed in the dorsal aspect of the neural tube (*D*and *F*, arrowheads), as well as in ectomesenchyme consistent with migrating neural crest (arrows). *DlxC* is not detected in the neural tube, but is detected in head ectomesenchyme (*E*, arrows). Lines labeled by lowercase letters d, e, and f indicate sectioning planes. Modified from (Neidert, Virupannavar et al. 2001).

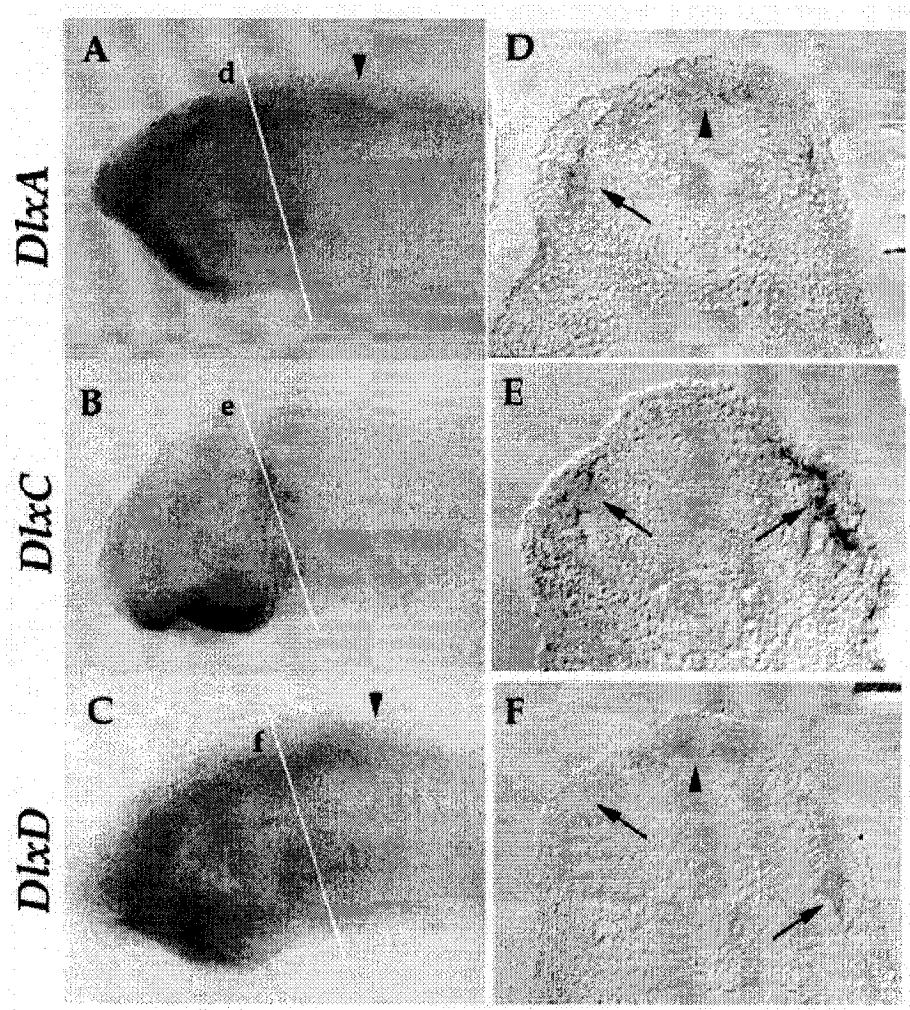
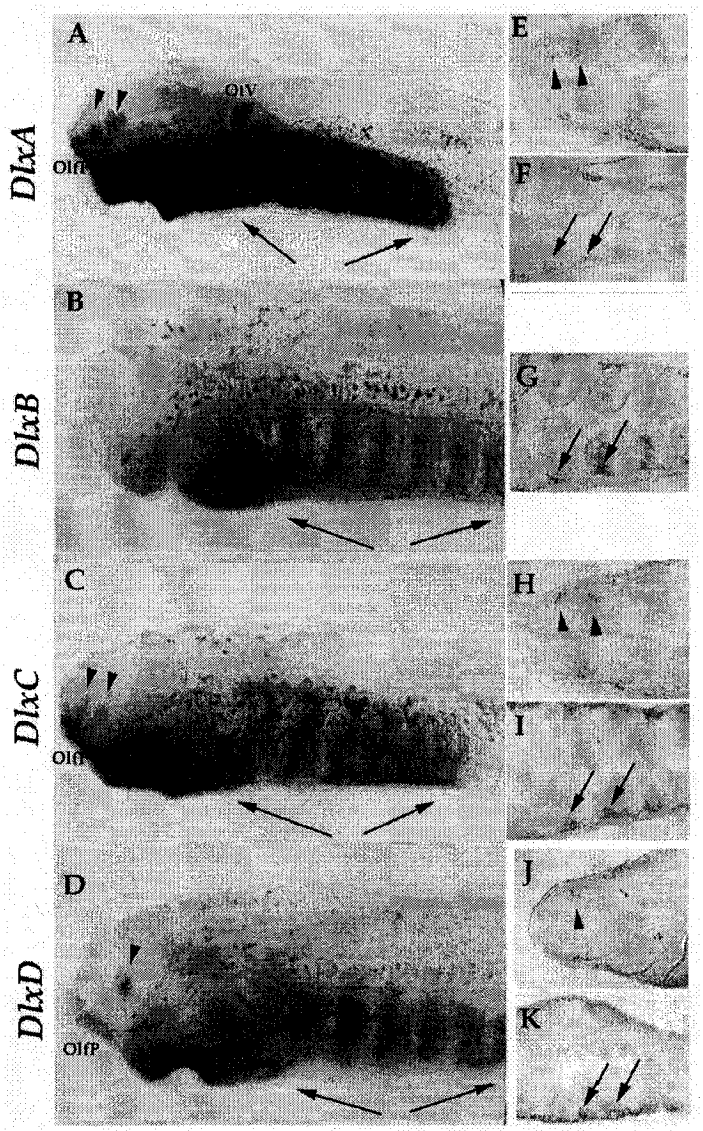

DlxA
DlxC
DlxD
A
B
C
D
E
F
d
e
f

Figure 1.12: Lamprey *Dlx* gene expression in the forebrain and branchial arches.

Whole mounts and horizontal sections of the head regions of 15-day-old lamprey embryos labeled with *DlxA*, *DlxB*, *DlxC*, and *DlxD* riboprobes. (*A–D*, *E–K*) All four lamprey *Dlx* genes are expressed in each branchial arch (*A–D*, arrows). Sections reveal that *Dlx* transcripts accumulate in rostrolateral quadrant of the arch mesenchyme (*F*, *G*, *I*, and *K*, arrows), the site of future cartilage condensations. *DlxA* and *DlxC* are expressed in two bilateral pairs of transverse stripes in the ventral diencephalon and telencephalon (*A*, *C*, *E*, and *H*, arrowheads), whereas *DlxD* is expressed in a single pair of transverse stripes in the ventral diencephalon (*D* and *J*). *DlxA*, *DlxB*, and *DlxC* are also expressed in the olfactory placode (OlfP) and the otic vesicle (OtV, not shown for *DlxC* and *DlxD*). *DlxB* is not detected in the forebrain, olfactory placode, or otic vesicle. Modified from (Neidert, Virupannavar et al. 2001).


A
E
OlfV
OlfP
F
DlxA
B
G
DlxB
C
H
DlxC
I
D
J
DlxD
K

In the forebrain of lamprey, *DlxA* and *C* are expressed in both Diencephalon and telencephalon (lamprey brains are much simpler than gnathostome brains but they do contain these two domains in the forebrain), whereas, *DlxD* is only seen expressed in the diencephalon. *DlxB* transcripts were not detected anywhere outside of the branchial arches.

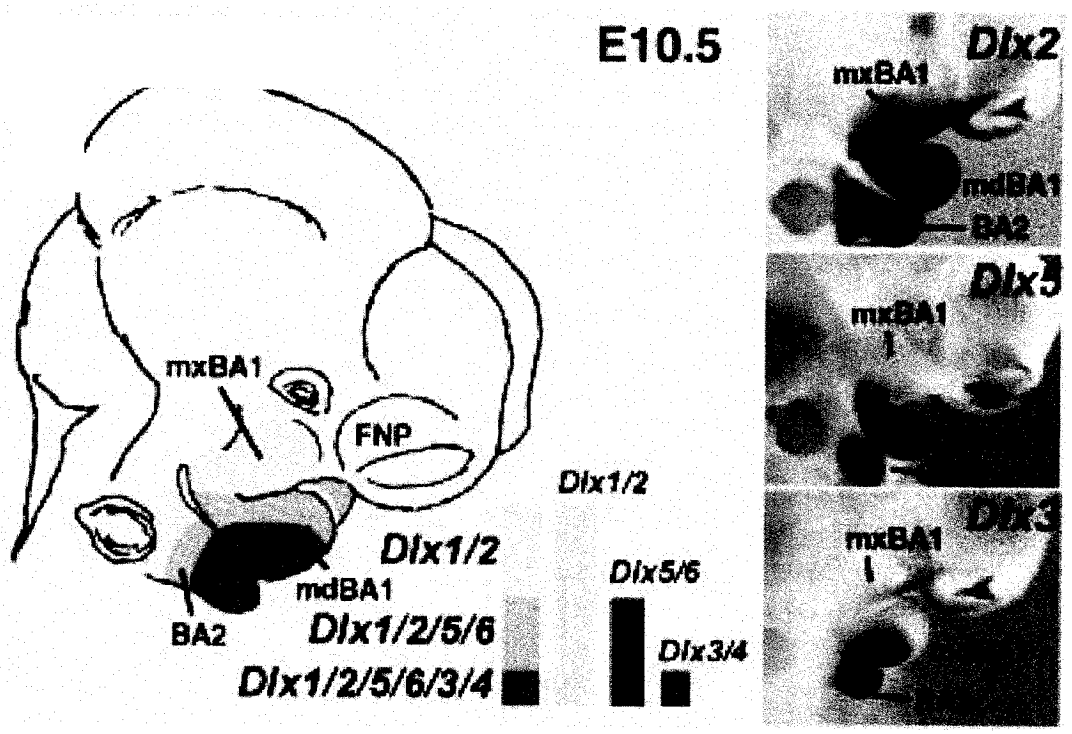
DlxA and *DlxD* transcripts are also detected in the trunk somites of lamprey, whereas none of the gnathostomes *Dlx* genes examined so far are expressed in trunk somites (Neidert, Virupannavar et al. 2001).

1.9. *Dlx* and Jaw acquisition.

The expression of *Dlx* genes in the branchial arches (segmentally repeated structures with proximo-distal polarity) of gnathostomes is of importance to the development and proximo-distal patterning of the jaw. In the first arch, the more proximal portion develops into the upper jaw and the more distal portion into the lower jaw. In the first arch, *Dlx* genes are expressed in nested pairs along the axis with *Dlx1/2* present along most of the axis, with *Dlx5/6* and *Dlx3/7* showing progressively more restricted distal expression (Depew, Lufkin et al. 2002) (see Figure 1.13).

Figure 1.13: *Dlx* gene expression in mouse branchial arches at embryonic day 10.5.

Left panel- Schematic showing overlapping expression patterns of the six *Dlx* genes in mice branchial arches. Right panel- Branchial arches of Embryonic day 10.5 mouse embryos labeled with *Dlx2, Dlx5, Dlx3* riboprobes. Modified from (Depew, Lufkin et al. 2002).


E10.5
mxBA1
FNP
Dlx1/2
mdBA1
BA2
Dlx1/2/5/6
Dlx1/2/5/6/3/4
Dlx1/2
Dlx5/6
Dlx3/4
Dlx2
mxBA1
mdBA1
BA2
Dlx5
mxBA1
Dlx3
mxBA1

The nested expression of *Dlx* genes in the branchial arches suggested that they might be determinants of proximo-distal (PD) identity of the arches. Therefore, it was thought if the distal determinants were abolished the distal regions should show proximal identity. This was exactly the case when our collaborator Dr. John Rubenstein (UCSF, San Francisco, USA), produced the *Dlx5,6*$^{-/-/-}$ double mutant mice. The loss of these two genes caused homeotic, mirror image transformation of the lower jaw into the upper jaw, suggesting *Dlx* genes do determine the proximo-distal identity of the branchial arches (Thomas, Tucker et al. 1997; Depew, Lufkin et al. 2002). The gain of regulatory mechanisms that provide the first branchial arch such polarity might have greatly facilitated the acquisition of biting jaws early in vertebrate evolution.

Lamprey *Dlx* genes are expressed in the branchial arches very strongly but the expression does not seem to be nested as seen in gnathostome vertebrates (Neidert, Virupannavar et al. 2001; Shigetani, Sugahara et al. 2002)(see Figure 1.12). Interestingly, lampreys also completely lack jaws. Instead of a jaw they have an oral apparatus that consists of a sucker like mouth, lined up with a series of spine like teeth (not true teeth). They use their oral apparatus to latch onto their prey and suck on their blood.

Homology between lamprey oral apparatus and gnathostome jaws (both of which are highly derived structures) is unclear and highly disputed. Therefore, *Dlx* expression in lamprey and gnathostome arches alone is inconclusive in saying that *Dlx* regulation was responsible for gnathostomes acquiring jaws. Testing lamprey *Dlx* regulatory elements in gnathosome genetic background would clearly show us if lamprey *Dlx* genes are expressed all over the PD axis of the arches and not in a distally restricted fashion as in gnathostomes or in other words if lamprey enhancers still possess the potential to target *Dlx* in a distally restricted fashion in gnathostome genetic background.

1.10. Statement of Problem.

The idea that morphological changes happen largely due to changes in cis-regulatory elements of developmental genes rather than changes in proteins that these genes code for, is gaining popularity amongst developmental biologists. Cis-regulatory elements of developmental genes provide a flexible node at which such changes can occur, mainly because they control individual genes rather than a large number of downstream targets (which is the case for trans-acting factors). Changes that lead to gain or loss of transcription factor binding sites within cis-regulatory elements can readily happen as these sites are relatively small (10-20bp) and consensus sequences can be defined even for the most well characterized binding sites (many of the nucleotide positions in a binding site are variable). Moreover, regulatory elements are modular in nature where one element controls expression to a subset of the overall expression, therefore changes in regulatory elements are thought to be more easily tolerated than changes in the trans-acting factor during evolution.

This idea seems attractive but very few studies that track such regulatory changes in genes to gain insights into morphological differences have been published (Gompel, Prud'homme et al. 2005). Tracking such morphological changes simply based on sequence divergence of conserved regulatory elements could be a good starting point. Here we want to identify and isolate *Dlx* regulatory elements in lamprey and compare their sequence and their activities to their gnathostomes counterparts.

Gene expression patterns in vertebrates suggest critical roles for *Dlx* in the development of forebrain, branchial arches, craniofacial derivatives, olfactory placodes, otic vesicle, inner ear, and limbs/fins. This is supported by individual and combinatorial *Dlx* knockouts in mice. These structures also happen to be major developmental innovations of vertebrates and are thought to be

responsible for their tremendous success during evolution, for example, neural crest and the derived sensory placodes are thought to have facilitated their transition from passive filter feeders to predators. Jaws, which may have equally contributed to such a transition is another such morphological feature that seems to have greatly contributed to the success of vertebrates.

In this thesis we want to identify and isolate cis-acting regulatory elements controlling *Dlx* expression in lamprey, a vertebrate ancestor. We also want to determine the structure and organization of *Dlx* genes, which has important implications for *Dlx* gene regulation. We are testing *Dlx* cis-acting regulatory elements in reporter gene assays in gnathostome genetic background (zebrafish and mice). From previous work in our lab, we know that there are extremely well conserved intergenic enhancers that control the expression of *Dlx* genes in gnathostomes. These enhancers are present in all teleost fish, amphibians and mammalian species examined so far and therefore are thought to have arisen before these groups of animals diverged from each other (350-400 million years ago). While investigating *Dlx* regulation in lamprey, we hope to trace back the origin of these remarkably well preserved *Dlx* intergenic enhancer sequences more precisely during vertebrate evolution.

Lampreys have much simpler brains when compared to gnathostomes and are the most primitive organisms alive to have cranial neural crest cells. Lamprey diverged from other vertebrates relatively soon after the neural crest arose and is thought to display primitive features lost or masked in gnathostomes (McCauley and Bronner-Fraser 2003). Lampreys have already been shown to have fewer number of migratory neural crest cells and the migratory routes these neural crest cells take seem much less restricted (when compared to gnathostomes), where crest cells from a single axial level populate multiple surrounding branchial arches with extensive mixing of these migratory cell populations

(McCauley and Bronner-Fraser 2003). Lampreys also completely lack jaws. *Dlx* has been shown to play indispensable roles in patterning both these structures. Regulatory mechanisms of *Dlx* genes have been implicated for the absence of jaws in lamprey (Neidert, Virupannavar et al. 2001).

Therefore, this project may provide insights into the state of neural crest during early vertebrate evolution and help us resolve if *Dlx* regulation was responsible for the acquisition of jaws in gnathostomes.

2. Materials and Methods.

2.1. List of oligonucleotides used.

Oligo Name	Oligo sequence	Primary use
Ljap Dlx-A LR1	GCTCTGCCGCCCCACACCTG	Pcr intergenic
Ljap Dlx-A LR2	CTGCAAGCCCCGCACCACCT	Pcr intergenic
LpdlxA fwd 3	CTCTGTTCACTTATGATTGC	Sequencing
LpDlxA fwd 4	TCCGAGGAAGTATGCAATAA	Sequencing
LpDlxA fwd 5	CGTAGCCATGGAGTCAACAT	Sequencing
LpDlxA Up-1	TGCCCAGCTTCTTGAACTTG	Sequencing
Lp DlxA up-2	GGGTTAATAAAGCAGTGGCC	Sequencing
Lp DlxA Gap Dn	AGTCCGTCTTGGTTGTTAGC	Sequencing
DlxA Up-3	GTGGTATGCAGCAGGAGACA	Sequencing
DlxA Gap2	GACTGACTGAGGTAACTGGC	Sequencing
DlxA-Up-4	AGCACTGCATGACGAGTCTC	Sequencing
DlxA-exon2-up	GCCAGGTACTGCGTCTGCTG	Sequencing
DlxA-exon2-Dn	CAGCAGACGCAGTACCTGGC	Sequencing
DlxA-exon1-Dn	AACGCCTACGAGGGCTTCAC	Sequencing
DlxA-exon-up-2	GCGGTGGTGATTAGAGAGATG	Sequencing
DlxA-exon2-Up2	ATGCCTGACGACGATGATAC	Sequencing
DlxA-exon1-Dn2	CGGACTTTATGAAGAGGGACG	Sequencing
DlxA-up5	ATCTGGTTAAATTATGTACGTTC	Sequencing
DlxA-exon1-Dn3	TACGAGGGCTTCACGCACGA	Sequencing
DlxA-Ex-1-up	CAGCGTGGGCGACGATTGCG	Sequencing
DlxA-Ex-1-Dn4	GCGTTCTCCTTTGTGTATTCCG	Sequencing
DlxA In-situ For	ACACCAGCAGCTCAGCTCTG	Pcr exon3 *DlxA*
DlxA In-situ Rev	GAGCTAGGAGCAGCAACAGG	Pcr exon3 *DlxA*
DlxA ex1 up2	AACCACCGCGAACGACTACG	Sequencing
DlxA-H1-Rev	AAGCTTTGGTCCAGTCCGTACAGGTC	Pcr homology arm1 *DlxA*
DlxA-H1-For	AAGCTTGGAAGCATACGACACAGCCC	Pcr homology arm1 *DlxA*
DlxA-H2-For	CTCGAGTCATCACCCGCAGGCGCAAT	Pcr homology arm2 *DlxA*
DlxA-H2-Rev	TCTAGATGAAGCCCTCGTAGGCGTTG	Pcr homology arm2 *DlxA*
PMAR Dlx B 3'	GCCGGTACCACGGCTGGTAC	Pcr part of exon3 of *DlxB*

PMAR Dlx B 5'	CCGGCGCTCCAAGTTCAAGC	Pcr part of exon3 of *DlxB*
PMAR Dlx B set 2-3'	GCCACGTGGACTTGCATCAG	Pcr part of exon3 of *DlxB*
PMAR Dlx B set 2-5'	GGCGAACTCGGCTCGGAGCA	Pcr part of exon3 of *DlxB*
Ljap Dlx-B LR1	GACTCCATGACCTGCGACTC	Pcr intergenic
Ljap Dlx-B LR2	CGCACCCCGTCGCCGACGCG	Pcr intergenic
Lp DlxB Rev	AGTCGCAGGTCATGGAATCG	Pcr part of exon1 of *DlxB*
LpDlxB-PCR-Dn	GTTGAGTGCGTATTACGCGG	Pcr part of exon1 of *DlxB*
LpDlxB-PCR-Up	GGCTCAGAGTCGTCCTTATC	Pcr part of exon1 of *DlxB*
DlxB-Dn-2	GTCTTGCGCATCTCCGGCAG	Sequencing
DlxB-Dn-3	GTCGCCGCGCTCCTTGGCGT	Sequencing
DlxB-Dn-2	GTCTTGCGCATCTCCGGCAG	Sequencing
DlxB-Dn-3	GTCGCCGCGCTCCTTGGCGT	Sequencing
PMAR DlxC 3'	ATCTGGTTTCAGAACCGGCG	Pcr part of exon3 of *DlxC*
PMAR DlxC 5'	CGGGTACTGAGCGTGCAGCT	Pcr part of exon3 of *DlxC*
PMAR DlxC 3'-2	GCCGTGCTGTTCGCAGGGCTGG	Pcr part of exon3 of *DlxC*
PMAR DlxC 5'-2	CAAACACGGCGAGATGATGCCC	Pcr part of exon3 of *DlxC*
LpDlxC up-1	CTCCATTGTGGACGGTTTAC	Sequencing
Lp DlxC Up-2	TGGTATGGTCAGCACCAAGC	Sequencing
Lp DlxC Dn-2	GACTTGGTGCAAGCTTATCG	Sequencing
DlxC Dn-2-2	ACATCAGCAGCATCAGCAGC	Sequencing
DlxC Up-2-2	CGTGACTCGATAGTTTCTGC	Sequencing
DlxC-up-3	AACATCTCGCCACCGTACAG	Sequencing
DlxC-Dn-3	CTACCACGAGACCACCTGCA	Sequencing
DlxC-exon2-up	TGCGTGACTCCCAGTGAGGC	Sequencing
DlxC-Up-4	CAGTGCCGATTGGTGAACGC	Sequencing
DlxC-Dn-4	TTCGCAGTGCCTTCCTCCGG	Sequencing
DlxC-exon2-Dn	GCCTCACTGGGAGTCACGCA	Sequencing
DlxC-exon1-Dn	TACCACCAGGGCTACCACCA	Sequencing
DlxC-exon-up-2	CAGTCTAGGCGTACACACAG	Sequencing
DlxC-H1-For	AAGCTTAGCAGGCACGGCAGCTCTAC	Pcr homology arm1 *DlxC*
DlxC-H1-Rev	GGTACCGATGAGTGCGCCTGGTGAAG	Pcr homology arm1 *DlxC*
DlxC-H2-For	CTCGAGGTCCACATCCACGGACTCCA	Pcr homology arm2 *DlxC*
DlxC-H2-Rev	TCTAGATACGCCTCGTAGGCGCTACC	Pcr homology arm2 *DlxC*
DlxC-H1-Rev-Hind	AAGCTTGATGAGTGCGCCTGGTGAAG	Pcr homology arm2 *DlxC*

DlxC In-situ For	TGCGAACAGCACGGCGCTCT	Pcr exon3 of *DlxC*
DlxC In-situ Rev	CGTTCACGGGAGCGACACTG	Pcr exon3 of *DlxC*
CD inter Rev	GTACGTGTGTATGTACGTGTGCAC	Sequencing
DlxC ex1 up	ATCCGGAGGTGTCTACGCAG	Sequencing
Cosmid C3 SP6 side	GCCAATGATGAAATCACCTGG	Pcr *DlxC* 2.5kb promoter
Dlx D Rev	TGAACACATCGAATCGTTCAAC	Sequencing
LpDlxD For-2	TTAGTGAACACTGGCTGCCT	Sequencing
LpDlxD Rev-2	TTGGCCATGCTCTCTTGATG	Sequencing
Lpdlx (D) Rev-3	TCATCGGCAGTAACAGTACC	Sequencing
Lpdlx (D) For-3	AAGTGGCGGTGACATCACCA	Sequencing
DlxD Dn-4	GCACCTATGAAGGATCACGC	Sequencing
DlxD Up-3	TGATACTATACACGCGCTCATA	Sequencing
DlxD-Up-4	TCAACCTCGGATTACAGGCC	Sequencing
DlxD-Dn-5	GATCGATTGCGTGGCATTGC	Sequencing
DlxD-exon2-up	AGCGCCAGGTACTGCGTCTG	Sequencing
DlxD-exon2-Dn	CAGACGCAGTACCTGGCGCT	Sequencing
DlxD-exon1-Dn	TACCTGAACCCCATCGGCAG	Sequencing
DlxD-exon-up-2	GCGCATAACATCGCTGAACA	Sequencing
DlxD-Up-5	GCAACATCGTTCGCGGATTG	Sequencing
DlxD In-situ For	CCCACTGGTACTCACCATCG	Pcr exon3 of *DlxD*
DlxD In-situ Rev	CGTCCTGCGCTGCGATCAG	Pcr exon3 of *DlxD*
Lj DlxD ex1 dn	CCATGAGCGCCTGCAGAATG	Sequencing
DlxD ex2 up2	CGTTGAAACACACGTCTCGC	Sequencing
I12a For	GAATTCGCTTGCAGAATAGAACTG	Pcr enhancer I12a
I12a Rev	GGTACCTGCGCAAGACAATTTCCTG	Pcr enhancer I12a
I12a For set2	GAATTCGGAGTCAAGATGTTATGCT	Pcr enhancer I12a
I12a Rev set2	GGTACCAAGCTGCAGCAATCATGTG	Pcr enhancer I12a
I56i Forward	CTCGAGAAAAATGTTTTCTTTT	Pcr enhancer I56i
I56i Reverse	CTCGAGGCATTATAATTTTGGT	Pcr enhancer I56i
Beta-globin-Fwd	AGGGCAGAGCCATCTATTGC	Pcr lacZ
LacZ-Rev	CGCTCATCCGCCACATATCC	Pcr lacZ
FH-fwd	GATCATGACCGCCGTAGG	Pcr fetal hemoglobin
FH-Rev	CATGAACTTGTCCCAGGCTT	Pcr fetal hemoglobin

Long oligonucleotides used for homologous recombination are listed below. The grey highlighted regions are homology arms and the remaining ~20bp at the end were for PCR amplification.

Amp Bg et

TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAG TTGCCTGACTCCCCGTCGTTGCGGCCGCGAATTCCTG

Amp cm et

ATGAGTATTCAACATTTCCGTGTCGCCCTTATTCCCTTTTTTGCGGCATTTTGCCTTCCTGTT TTTGCTCACCCAGAAACTGCTTTCGAATTTCTGCCAT

pKD46 amp-cm H1

TATCCGCTCATGAGACAATAACCCTGATAAATGCTTCAATAATATTGAAAAAGGAAGAG TACGAGGCCCTTTCGTCTTCG

pKD46 amp-cm H2

TACCAATGCTTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAT TGCTTTCGAATTTCTGCCA

2.2. Southern Hybridization Experiments.

2.2.1. Southern Blotting:

After running the digests on an agarose gel in TBE, the gel was treated with depurination solution (0.25M HCl) for 10-20 minutes to fragment large pieces of DNA and to facilitate transfer onto the membrane. After a brief wash in distilled water, the gel was treated with denaturation solution (1.5M NaCl, 0.5M NaOH) for 30 minutes followed by a wash with distilled water and treatment with neutralization solution (1M Tris-HCl, 1.5M NaCl) for another 30 minutes. The DNA from the gel was then transferred to a HyBond-N membrane using the capillarity transfer method. A wick was setup in 20X SSC (3M NaCl, 0.3M NaCitrate), on which 3 sheets of 3MM Whatman paper was setup. The gel was placed upside down onto the Whatman paper followed by nylon membrane on the gel. On top of this setup, 10 cms of dry paper towels were stacked and a glass

plate with approximately 750gms of weight was put on the paper towels. The gel was surrounded with parafilm to prevent the paper towels to come in contact with the wet paper below the gel. To prevent evaporation of the transfer solution sides of the tray were sealed by plastic wrap. This setup was left overnight for transfer. The membrane was then baked for two hours at 80°C to fix the DNA onto the membrane.

2.2.2. Probe Preparation for Southern Hybridization:

Probes were prepared by two methods. Random priming procedure was used to prepare probes from larger double stranded DNA and for oligonucleotide probes end labeling by phosphorylation reaction was used to incorporate radioactivity.

Labeling by Random Priming:

For making dsDNA radioactive probe, 50-100 ng of DNA template was mixed with 2 µg of pdn6 (random hexanucleotide), 1 X React2 buffer (Invitrogen), 1.5 mM dNTP mix lacking dATP, 50 µCi of 32-P-alpha dATP and 10 units of Klenow (Invitrogen) and adjusted to a final volume of 25 µl with water. This mix was incubated for 1hr at 37°C. The reaction was stopped by adding 1 µl of 0.5M EDTA. The probe was purified by passing through Sephadex G-25 column made in a 1 ml syringe blocked by glass wool plug at the bottom. The labeled probe was collected by spinning for 2 min at 1500g.

End labeling of oligonucleotides:

Ten picomoles synthetic oligo, 1X T4 polynucleotide kinase reaction buffer (70 mM Tris-HCl, 10 mM MgCl2, 5 mM dithiothreitol), 5 µC of γ32 P (Amersham) and 10 units of T4 polynucleotide kinase, in a final volume of 25µl were incubated at 37 °C for 1 hr. The reaction was terminated by adding EDTA (1 µl of

0.5M solution). The probe was purified by ammonium acetate, ethanol precipitation and resuspended in 100 µl of TE buffer.

2.2.3. Southern Hybridization:

Blots or colony lifts were placed in 50ml Falcon tubes. Prehybridization was done in Church-medium (0.36 M Na_2HPO_4, 0.14 M NaH_2PO_4) containing 0.5 mg/ml of denaturated salmon sperm DNA for 2 to 6 hours at 65°C. The medium was then removed, and replaced with fresh Church-medium. Twenty µl of the radioactive probe was incubated with the membrane, and the hybridization was done over night at temperatures ranging from 50 to 65°C, depending on the length of the probe.

Blots were washed three times with 1X SSC, 0.1% SDS. They were then removed from the Falcon tubes and washed on a shaker for 15 minutes in 1X SSC, 0.1% SDS. A second wash was then performed for another 15 minutes with 0.1X SSC, 0.1% SDS. Blots were then exposed to BIO RAD Imaging Screen - K overnight and scanned with BIO RAD Molecular Imager® FX.

2.2.4. Stripping Southers blots of probe:

The DNA blots were stripped of radioactivity by washing in boiling hot 0.1% SDS solution and left to cool at room temperature. The blots were then washed with 1X SSC solution and stored at 4°C until further use.

2.3. Turbo screening of colony lifts.

This protocol was used when a large number of bacterial colonies were screened for a particular DNA sequence using southern hybridization. Nitrocellulose membrane of appropriate size and shape were placed on agar plates with colonies. Holes were made through the agar and membrane in 3 corners for

orientation after the membrane became wet. The membrane was removed with help of foceps and placed colony side up on Whatman filter paper soaked in 2X SSC, 5% SDS solution for 2 minutes. The membrane was then removed and placed in a microwave oven and heated for 5 minutes at full setting. The dry membranes were treated as southern blots for further analysis.

2.4. Screening cosmid genomic library.

The cosmid library macroarray (RZPD) was screened by southern hybridization according to the procedure described above. Oligonucleotide pairs described in the following table were used to amplify parts of third exon excluding the homeobox sequence of *Dlx* genes from *Lampetra fluviatilis* genomic DNA kindly provided by Dr. Sylvie Retaux (CNRS- Gif-sur-Yvette, France). The primers were designed from cDNA sequences available from NCBI for sea lamprey (*Petromyzon marinus*) *Dlx* genes, with following accession numbers AY010116 (*DlxA*), AY010117(*DlxB*) and AY010118 (*DlxC*) (Neidert, Virupannavar et al. 2001). *DlxD* sequences were obtained from cosmid C3 by primer walking sequencing method. Cosmid C3 was one of the cosmids obtained for *DlxC* and contains part of the *DlxD* gene. The primers initially used for sequencing *DlxD* regions from cosmid C3 were designed from cDNA sequence available for *DlxD* from *Petromyzon marinus* (Accession number - AY010119 (Neidert, Virupannavar et al. 2001)). The obtained *DlxD* sequences were used to design a pair of primers to PCR amplify the third exon of *DlxD* excluding the homeobox sequence of *DlxD*.

The amplified sequences were subcloned with the help of QIAGEN PCR Cloning Kit. The identity of the amplicons was assessed initially by size followed by sequencing with T7 primer. The subcloned fragments were released by *Eco*RI digestion, purified and radioactivity labeled using Random Priming Technique described above.

DlxA	PMAR Dlx A 5′	GCGCTCCAAGTTCAAGAAGC	*Petromyzon marinus*
	PMAR Dlx A 3′	TGCTGGTGCTGCAGGTACCA	*Petromyzon marinus*
DlxB	PMAR Dlx B 5'	CCGGCGCTCCAAGTTCAAGC	*Petromyzon marinus*
	PMAR Dlx B set 2-3'	GCCACGTGGACTTGCATCAG	*Petromyzon marinus*
	LpDlxB-PCR-Dn	GTTGAGTGCGTATTACGCGG	*Petromyzon marinus*
	LpDlxB-PCR-Up	GGCTCAGAGTCGTCCTTATC	*Petromyzon marinus*
DlxC	DlxC In-situ For	TGCGAACAGCACGGCGCTCT	*Lampetra fluviatilis*
	DlxC In-situ Rev	CGTTCACGGGAGCGACACTG	*Lampetra fluviatilis*
DlxD	DlxD In-situ For	CCCACTGGTACTCACCATCG	*Lampetra fluviatilis*
	DlxD In-situ Rev	CGTCCTGCGCTGCGATCAG	*Lampetra fluviatilis*

2.5. Zebrafish embryo microinjection.

Circular DNA at a concentration of 100-400 ng/µl was diluted in 100mM of KCl and 0.05% of phenol red. Microinjection needles (ID:0.5mm, OD:1mm, Borosilicate with filament, Sutter Instruments) were pulled using a P90 micropipette puller (Browing and Flame). Zebrafish were bred and embryos were collected and arranged on an agarose plate for microinjection as described by - The Zebrafish Book (Westerfield 1995). One or two cell stage embryos were microinjected using a Narashingi IM300 microinjector. The amount of DNA solution injected was adjusted manually from the microinjector such that the

diameter of the injected solution (red in colour) is approximately one-fifth of the embryo's diameter.

The injected embryos were maintained in petri dishes at 28.5°C and primary embryos were examined for *Egfp* expression at 1.5, 2, 3 days post fertilization (dpf). *Egfp* expression was detected by using a fluorescent microscope (microscope: SMZ 1500, Nikon; Camera: Retiga 1300, Q-Imaging) equipped with *Egfp* filters and mercury lamp.

The embryos which showed *Egfp* expression were raised until maturity for making stable transgenic lines. This will be done by breeding each primary fish with a wild type and screening for *Egfp* expression at around 2 dpf under a fluorescent microscope.

2.6. Transient transgenesis in mice and genotyping.

The linearized plasmid constructs were prepared following GeneCAPSULE™ nucleic acids extraction method and were injected to fertilized oocytes at a concentration of 5ng/µl following standard procedures (Hogan, Constantini et al. 1986). The injections were performed by Adrianna Gamarotta, animal care technician at Ottawa Health Research Institute, Ottawa. The presence of transgenes was tested in embryos by PCR on genomic DNA prepared from placentas. PCR was performed using the following set of oligonucleotide primers (for sequence see section 2.1), Beta-globin-Fwd and LacZ-Rev, which give approximately a 700 bp PCR product. Part of Fetal hemoglobin gene was used as an internal positive control for the PCR reactions with the following oligonucleotides (for sequence see section 2.1), FH-fwd and FH-Rev, which give a 300bp PCR product. PCR amplification was performed using the PTC-100, Peltier Thermal Cycler (MJ Research). PCR conditions used were as follows; 2µl of genomic DNA preparation, 2.5µl of 10X PCR reaction buffer, 2.5µl of dNTPs

(2mM), 1µl of forward and reverse primers (10pM) and 1µl of Taq DNA Polymerase in a final volume of 25µl. The initial denaturation step was performed at 94°C for 4 min. This was followed by 30 cycles with a denaturation step at 94°C for 1 min, an annealing step at 60°C for 1 min, and an extension step at 72°C for 1 min. The final extension step was performed at 72°C for 8 min.

2.7. Genomic DNA preparation from placental tissue.

Placental tissue was separated from embryos and frozen until used for DNA extraction. 500 µl of Proteinase K solution (50 mM Tris (pH 8.0), 100 mM EDTA (pH 8.0), 1% SDS, 100 mM NaCl, 350mg Proteinase K (fresh)) was added to the tissue, mixed and incubated overnight at 55°C with intermittent mixing. Next morning, 150 µl of 6M NaCl solution was added and the tubes spun for 10 min at 13000 rpm. Supernatant containing the DNA was transferred to a fresh tube and 500 µl of isopropanol was added. The precipitated DNA was spun down and washed once with 70% ethanol. The DNA pellet was resuspended by incubating in 200µl of TE buffer at 37°C overnight.

2.8. Beta-galactosidase staining for mice embryos.

Embryos injected with construct were harvested from foster mother, dissected in 1 X PBS (0.1 M phosphate buffer saline, 141g Na2HPO4 anhydrous, 8 ml 85% phosphoric acid adjusted to 10L final volume with water) to remove the extra-embryonic tissues. Placentas were removed and kept frozen for screening by PCR for transgene presence. After many washes with PBS, the embryos were fixed for 30 min at 4°C in FIX solution (1% formaldehyde, 0.2% gluteraldehyde, 0.02% NP-40, 1 X PBS). Followed by three, 15 min washes with 1 X PBS. For staining, the embryos were incubated in staining solution (1mg/ml X-gal, 5mM $K_3Fe(CN)_6$, 5mM $K_4Fe(CN)_6$, 2mM $MgCl_2$, 0.02% NP-40, 1XPBS) overnight at 30°C. Next morning, stained embryos were washed several times in PBS and stored in 1 X PBS with 10mM EDTA.

2.9. Electrocompetent cells preparation.

A single colony of E.coli was used to inoculate a 5ml Luria Bertani (LB) media starter culture at 37°C shaker (225 rpm). After this starter culture was saturated (14-16 hours), 2 ml of it was used to inoculate a 200ml LB media culture. Cells were harvested after they reached an OD (600nm) of 0.4-0.6. The culture was chilled on ice-water bath for 15 min with intermittent shaking. After this point the cells were maintained at temperature between 0 and 4°C. The cells were then spun at 5000 g for 5 min in an Avanti J-series Beckman Coulter Centrifuge, the supernatant was discarded and the cells were gently resuspended in 50 ml of double distilled pre-chilled water. This was repeated twice and finally the cells were resuspended in about 1ml of 10% autoclaved glycerol (also pre-chilled). The cell suspension was transferred to a 1.5 ml microtube and spun for 30 sec in a table top refrigerated centrifuge (at 0°C). The volume of cells was estimated (approx 400µl) and the cell suspension was adjusted for a final volume equal to twice the volume of cells.

2.10. Transformation of *E.coli* by electroporation.

Electroporation cuvettes (1mm gap, BioRad) were washed several times with double-distilled water and 70% ethanol followed by cooling them on ice for at least 15 min. Electrocompetent cells were thawed on ice and DNA to be transformed mixed with the cells. Cells were added to the cold cuvette and electroporated at standard settings on the electroporator (Biorad Micropulser). 1ml LB was immediately added to the cuvette and the cells were incubated at 37°C for 1hr with shaking. The cells were spun down and plated on suitable antibiotic LB plates.

2.11. Homologous recombination using DY380 *E.coli* strain.

The recombinogenic *E.coli* DY380 (obtained from Dr. Neal Copeland, US National Cancer Institute, Frederick, USA) were made electrocompetent using the method described in section 2.9 except, the cells were always incubated at 32°C instead of 37°C and recombination was induced by incubating the culture at 42°C for 15 min after the cells reached an OD (600nm) of 0.4-0.6. Plasmid/cosmid to be modified was electroporated in DY380. Single colonies were picked and were confirmed for the presence of the plasmid/cosmid by restriction digestion of isolated plasmid DNA. These DY380 cells containing the plasmid/cosmid were made electrocompetent and recombination was induced by incubating at 42°C for 15min before the washes with ice-cold water (section 2.9). 300ng of PCR products harbouring homology arms (to plasmid or cosmid to be modified) and antibiotic selection gene were electroporated into induced, plasmid/cosmid containing DY380 cells and plated on selection for antibiotic resistance provided by the resistance gene in PCR products.

2.12. Homologous recombination with pKD46 plasmid in DH5α (*E. coil* strain).

pKD46 plasmid: This plasmid contains genes for the bet, exo and gam proteins from lambda genome under the control of arabinose inducible promoter. pKD46 also has a temperature sensitive origin of replication *oriR101* and is subsequently lost from the bacteria if grown at 37°C or above. Therefore, cells containing the pKD46 plasmid were always grown at 30°C except when pKD46 was no longer required (after the desired modification of the cosmid was done)

DH5α *E.coli* strain was first transformed with the red recombinase plasmid pKD46 (obtained from the E.coli stock centre, Yale University) followed by the cosmid that had to be modified. Individual colonies containing both the plamid

and the cosmid were grown to saturation in a 3ml LB culture overnight at 30°C. This culture was used to inoculate a 50 ml culture with appropriate antibiotics. This 50 ml culture was grown to an OD (600nm) of 0.3 followed by an hour long induction by L-arabinose at a final concentration of 10mM. After the 1 hr induction the cells were harvested and made electrocompetent. To perform homologous recombination, the linear targeting construct was electroporated into these bacteria as described in section 2.10 and plated on LB-agar plates with antibiotic carried by the linear targeting vector. The cells after electroporation pulse were incubated at 37°C to get rid of pKD46 plasmid.

2.13. Shotgun sequencing of Cosmids.

2.13.1. Shotgun Library construction

Twenty to fourty μg of DNA from cosmids A, C2 and D in 100μl of water was sonicated using a probe sonicator, at an output of 3 for 6 seconds. Conditions for efficient fragmentation in the desired size range were initially optimized. Fragments were then treated using 5 units of T4 DNA polymerase (Invitrogen) to fill in the 3′ ends, and 5 units of Klenow (Invitrogen) to fill in the 5′ ends for 30 mins at 37°C. The end treated DNA was extracted by Phenol: chloroform: isoamyl alcohol extraction (Sambrook, Fritsch et al. 1989). The purified DNA was further treated by 20 units of T4-polynucleotide kinase (Invitrogen) in 2mM ATP for 30 min at 37°C to restore phosphate groups at the ends.

2.13.2. Size selection

The fragmented and end-repaired DNA was then run on a 0.8% agarose gel. Fragments from 0.8 to 1.5Kb were gel extracted and purified using the QIAGEN Gel Extraction Kit and resuspended in 30μl of water.

2.13.3. Vector DNA preparation and Ligation

The vector, pBluescript, was linearized using EcoRV, and was treated with CIAP to hydrolyse the 5′ phosphate groups. Ligation reactions were setup with the following concentrations and conditions; 50ng of the vector, 150ng of the sonicated DNA, 1X ligase buffer and 5U ligase in a final volume of 10µl were incubated overnight on a PCR machine at 16°C.

One µl of the ligation products was then electroporated into 50µl of electro-competent *E. coli* DH5α cells. The cells were then grown in Liquid Broth (LB) at 37°C for one hour on a shaker at 250 rpm. They were then plated on LB-AMP-IPTG-Xgal plates. Fragments inserted in the *Eco*RV site were likely to disrupt the *lacZ* reading frame in pBluescript producing white colonies in an IPTG-Xgal LB plate.

2.13.4. Size determination of insert by PCR

White colonies were picked and grown in 1.5ml of LB over night in 96 well culture racks. PCR was performed using 2µl of the freshly grown bacteria as template. The final concentrations for the PCR were; 1X Taq buffer, 0.2µM M13 Forward primer, 0.2µM M13 Reverse primer, 200µM dNTPs, 1.0U Taq, in a final volume of 20µl. The thermal cycling parameters were; 1 cycle at 95°C for 5 minutes, 1 cycle at 4°C for 2 minutes,1 cycle at 94°C for 5 minutes, followed by, 30 cycles of 30 seconds at 94°C, 30 seconds at 55°C, 1 minutes at 72°C. This was then followed by a cycle at 72°C for 5 minutes, and cooling at 4°C.

Clones containing an appropriate sized insert were then re-grown in 5 ml of LB culture overnight. Plasmid DNA was isolate using the Wizard Plus© Plasmid Purification Kit (Promega). Elution was done in water with a final volume of

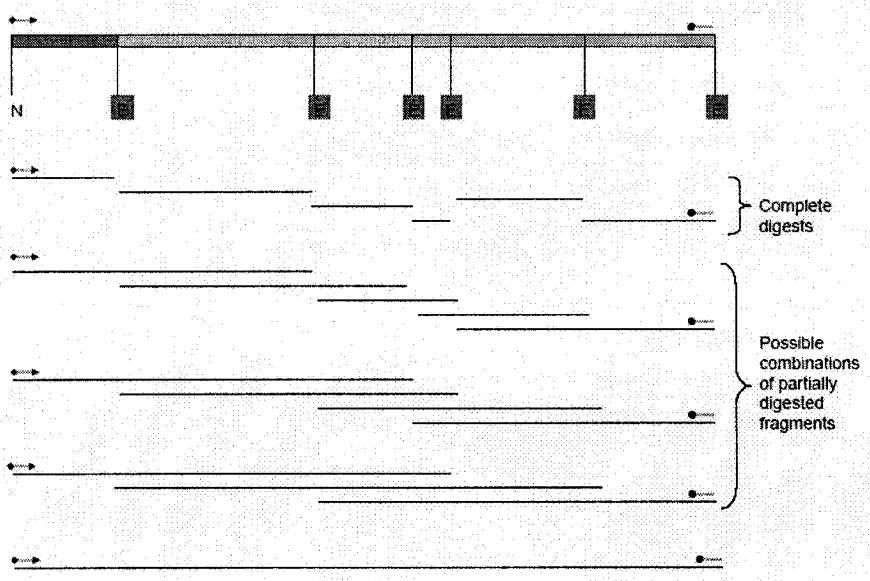
75µl. Double digests on the purified plasmid DNA were done with XhoI and XbaI to confirm the presence of an insert.

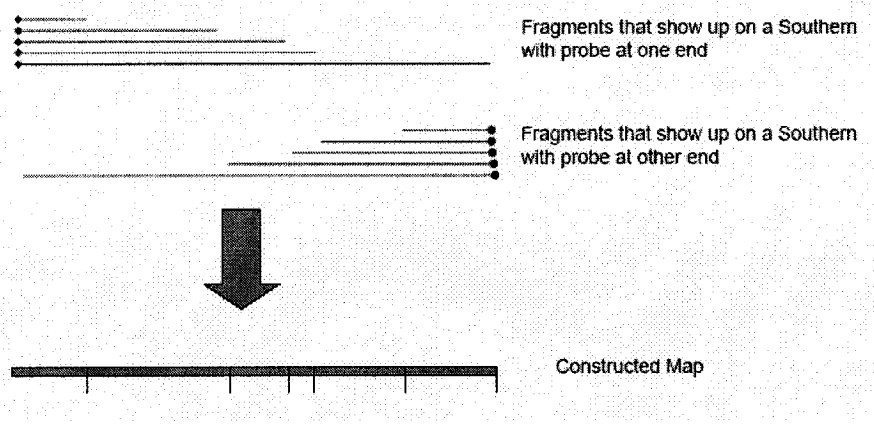
The purified DNA was then sent to McGill University and Genome Québec Innovation Center for sequencing using 3730XL DNA Analyzer systems. Contigs were assembled using Contig Express in VectorNTI 10.1.1 software package from Invitrogen.

2.14. Mapping using partial digests.

The Cosmid to be mapped was linearized with *Not*I followed by a partial digestion with *Eco*RI and *Xho*I separately with varying amounts of digestion. The time and amount of enzyme was calculated before for obtaining partially digested cosmid. These fragments were separated on a 0.5% agarose gel and blotted on a nitrocellulose membrane for Southern hybridization experiments. As shown in the schematic in figure 2.1, if this blot is hybridized with a probe at one end of the linearized cosmid, only the partially digested fragments that contain this probe will show on the Southern. From the difference in size of adjacent bands on such a Southern one can infer the restriction map from one end of the cosmid. This can be repeated from the other end to map restriction sites from the other end.

Figure 2.1: Schematic displaying the concept of restriction mapping of the cosmid by Southern on partially digested linear cosmid.


N
Complete digests
Possible combinations of partially digested fragments


Fragments that show up on a Southern with probe at one end
Fragments that show up on a Southern with probe at other end
Constructed Map

3. Results.

3.1. Isolation & characterization of cosmids containing lamprey *Dlx* genes.

Primers were designed using available cDNA sequences of *Dlx* genes from the sea lamprey (*Petromyzon marinus*). Part of third exons (excluding the homeoboxes) of all four *Dlx* genes were PCR amplified from genomic DNA of *Lampetra japonica* (Japanese River Lamprey), subcloned and used to screen a cosmid genomic library (from RZPD) of *Lampetra fluviatilis* (European River Lamprey). Five cosmids were isolated from the library containing three *Dlx* genes (one for *DlxA*; 3 for *DlxC* and one for *DlxD*).

3.1.1. *DlxA*:

For the single cosmid of *DlxA* we sequenced in and around the *DlxA* gene by primer walking method. Primer walking method involves sequencing using a primer on the cosmid and from the sequence obtained another primer is designed to sequence further downstream on the same cosmid. This is repeated several times depending on the amount of sequence to be obtained and the sequence reads are aligned against each other to obtain an assembly. We have obtained about 12kb sequence out of which 4.6 kb lies downstream of *DlxA*. From this 4.6 kb sequence we found sequences resembling another gene downstream of *DlxA* which is not a *Dlx* gene and therefore *DlxA* is probably not linked to any other *Dlx* gene (all linked vertebrate *Dlx* genes are directly linked to each other in tail to tail orientation).

Within this 4.6 kb downstream intergenic region we were unable to find sequences similar to *Dlx* enhancers from other vertebrates (Note: most of the vertebrate enhancers are found downstream of *Dlx* genes). This intergenic region was compared to the three *Dlx* loci from mice using BlastZ algorithm used in the

PIP-Maker (percent identity plot maker) program (Schwartz, Zhang et al. 2000). This particular algorithm has been optimized for comparison of long genomic sequences and has been used extensively to identify regulatory elements by phylogenetic footprinting. In addition to sequence comparisons, we specifically looked for *Dlx* binding sites in this region. The only conserved feature amongst the vertebrate *Dlx* enhancers (I12b, I56i and I37-2) is the presence of two extremely conserved *Dlx* binding sites separated by about 50 bp. This feature was not present in the sequences we obtained from cosmid A.

3.1.2. *DlxB*:

All four cosmids that we obtained for *DlxB* are probably false positives, because we were unable to get any information about *DlxB* from them. The PCR fragments of *DlxB* that were used to obtain these cosmids could never be amplified from any of the *DlxB* cosmids. Neither were we able to obtain any sequence reads from primers specific for *DlxB*.

3.1.3. *DlxC*:

The three cosmids obtained for *DlxC* were all verified to contain the regions with which the library was screened for (by PCR and also by southern hybridization). Cosmid C1 starts at third exon (contains most of third exon) of *DlxC* and contains about 10kb region downstream of *DlxC* (see figure 3.1). Size of cosmid C1 which is about 15kb is unexpectedly small compared to standard cosmid size ranges (about 30-45 kb). This could be a result of rearrangements or deletions within the cosmid and therefore cosmid C1 was excluded from further investigation. Cosmid C2 is about 35 kb in size and contains more than 10kb region upstream of *DlxC*. Also upstream of *DlxC* at the end of cosmid C2 we have sequences similar to vertebrate vitaminD receptor gene. This means that we have sequences upstream of *DlxC* until the next gene, and there is a good chance that we have all the upstream regulatory elements within this region. In mice and zebrafish

genomes the vitamin-D receptor gene is nowhere close to any of the *Dlx* genes. Cosmid C2 contains the complete *DlxC* gene and no part of the adjacent *DlxD*. Cosmid C3 which is about 38 kb in size contains the complete *DlxC* gene and about 2.5 kb region upstream of *DlxC* (see figure 3.1).

3.1.4. *DlxD*:

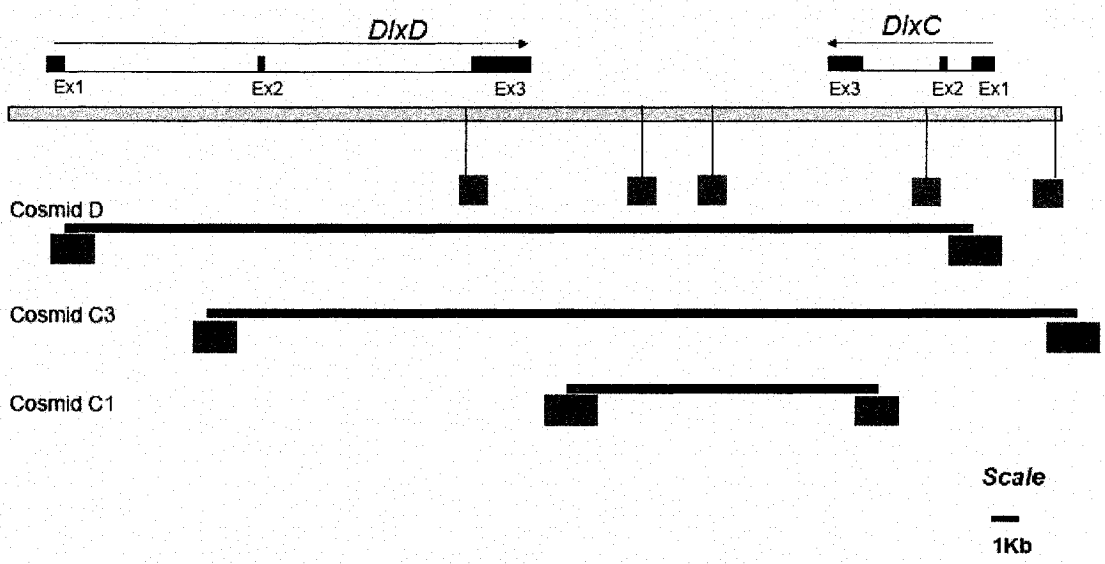
Cosmid C3 also contains parts of *DlxD* gene. From PCR and Southerns on the digested cosmids, we know that cosmid C3 contains the second and the third exon of *DlxD* but not the first exon (sequencing with primers in first exon did not work) (see figure 3.1). The absence of the first exon was confirmed by direct sequencing on the cosmid with primers in the second exon going upstream (the cosmid ends before the first exon). Using sequencing by primer walking on cosmid C3, we have obtained about 7kb of sequence in and around *DlxC* and about 5kb sequence from *DlxD* gene region. Cosmid D, obtained for *DlxD* is about 43 kb in size and contains a large region of the *DlxC* gene. With the help of Southerns, PCR and sequencing we know that cosmid D contains the second and third exon of *DlxC* (see figure 3.1). We were unable to get sequence reads near the first exon from this cosmid using oligonucleotides designed from *Petromyzon marinus DlxD* first exon (NCBI accession number - AY010119). Therefore, currently none of the cosmids contain the first exon of *DlxD*.

3.2. Genomic organization of lamprey *Dlx* genes.

Sequence reads from T7 end of cosmid C3 match *DlxD* sequences, direction of this match tells us that *DlxD* is going downstream into the cosmid C3. By PCR, southern and sequencing we also know that this cosmid contains the second and third exon of *DlxD*. Thus, *DlxC* has its tail towards *DlxD* or in other words, 3′ regions of *DlxC* are closer to *DlxD* compared to the 5′ regions.

Figure 3.1: Regions of the *DlxC/D* locus contained within the cosmids C1, C3 and D.

Orientation of the insert is shown by T7 and SP6 promoter sequences from the multicloning site of the backbone. *Eco*RI restriction sites (Grey boxes) are only shown in the region within cosmid C3. Size of *DlxD* intron1 may be more than 7kb (as shown in figure) as none of the cosmids contains the first exon of *DlxD*.


DlxD
DlxC
Ex1
Ex2
Ex3
Ex3
Ex2
Ex1
Cosmid D
Cosmid C3
Cosmid C1
Scale
1Kb

From Southern and PCR experiments we know that cosmid D contains the third exons of both *DlxC* and *D*. Further from sequences obtained from cosmid D, we know that *DlxD* goes downstream from one end of this cosmid. Therefore, *DlxC* and *DlxD* are directly linked to each other and form a convergently transcribed bigene cluster as found in most other vertebrates (see figure 3.1).

This has important implications for understanding evolution of regulatory mechanisms of *Dlx* genes. It has been proposed and widely accepted that one of the factors responsible for the clustering of developmental genes in animal genomes is due to shared enhancer elements. We know from previous work done in our lab that within the *Dlx* bigenic clusters there are shared enhancer elements. These bigenic clusters are always accompanied by these intergenic enhancers in vertebrates and vice versa. Therefore, the lamprey *DlxC/D* intergenic region seems the most likely place where one may find regulatory elements controlling expression of the flanking *DlxC/D* genes.

As mentioned before in section 3.1.1, sequencing data from cosmid A indicates that there is no *Dlx* gene immediately downstream of *DlxA*.

3.2.1. Exon intron structure of lamprey *Dlx* genes:

Lamprey *Dlx* cDNA sequences (available from NCBI) were compared to the genomic sequences obtained from the cosmids to determine their exon-intron organization. *DlxA, C* and *D* all have three exons interrupted by two introns as found in other vertebrates. We also know that the second splice site (which is in the middle of the homeobox sequence) for *DlxA*, C and D is conserved with that of other vertebrates. There is not enough sequence conservation near the first splice sites to say that it is conserved with that of other vertebrates, but it does occur at roughly the same number of nucleotides from the homeobox as in other vertebrates (splice site is at 24 codons before the start of homeobox sequence).

3.3. Detection of highly conserved vertebrate *Dlx* enhancers in lamprey.

To detect the presence of conserved *Dlx* arch enhancers, we performed Southern hybridizations with I12a (mouse) and I56i (zebrafish) enhancer sequences on digested cosmids. The Southern hybridizations were done at 65°C (the melting temperature of both the probes is much higher) and the washes were done at room temperature twice for 15 minutes each with 1X SSC and then twice for 10 min each with 0.1 X SSC. After the first wash with 0.1 X SSC the blot was checked with a Geiger-Müller counter (Model 3, Ludlum measurements Inc.) at 0.1 times setting and if there was too much or too little radioactivity the washes were changed accordingly. We were unable to detect these enhancers on the digested cosmids using the above mentioned Southern experiments.

We also used primers (I12a For, I12a Rev, I12a For set2, I12a Rev set2, I56i Forward, I56i Reverse, see section 2.1 for sequences) in conserved regions of these enhancers to PCR amplify them from these cosmids. Annealing temperature used for these PCRs was generally 2-3°C degrees lower than the melting temperature of the lower melting oligonucleotide. These PCRs also were unable to detect the presence of the two enhancers (I12a, I56ii).

We have also tried comparing vertebrate *Dlx* enhancer sequences (mouse I12a, I12b, I56i, I56ii, I37-1, I37-2) using megablast to sequence reads available from the *Petromyzon marinus* genome sequencing project at the NCBI trace archive. The trace archive at present contains about 10.2 million shotgun sequence reads from the lamprey genome. Assuming each sequence read produces 300bp (conservative estimate) of reliable sequence, the trace archive contains approximately 3 billion base pairs. This amounts to 1.5 times coverage of the sea lamprey genome which is about 2 billion base pairs

(http://www.genome.gov/19516773, (Gregory 2005)). We were unable to find sequences similar to the six intergenic *Dlx* enhancers from mouse in the available lamprey genome sequence using MegaBlast.

3.4. PCR amplification of *Dlx* intergenic region.

To quickly determine the linkage between lamprey *Dlx* genes, we tried to amplify the intergenic region (3-15 kb in vertebrates) between *DlxD* and one of *DlxA, B* and *C* gene. This was based on phylogenetic analysis of lamprey *Dlx* homeodomains. *DlxD* groups with *Dlx*1, 6, 7 clade and *DlxA, B, C* group with *Dlx2, 3, 5* clade (Neidert, Virupannavar et al. 2001) and as all vertebrate *Dlx* bigenes have one member from each of the two clades. It was assumed *DlxD* is linked to one of *DlxA, B* or *C* gene.

Two primers were designed going downstream of each lamprey *Dlx* gene in the third exon. Long PCRs were done using one oligonucleotide from the end of *DlxD* and the other oligonucleotide from the end of one of *DlxA*, B or C with isolated cosmids as the template. It was assumed that the intergenic region would be somewhere between the sizes found in fugu and mouse i.e. 3 - 12 kb. This assumption was based on relative genome sizes of these species and the somewhat observed proportional length of the *Dlx* intergenic regions (*Lampetra fluviatilis* genome is less than half the size of human genome (Gregory 2005)). None of these PCRs gave us bands larger than 4 kb. The 4kb band was ruled out as the intergenic region as a 4kb band was also seen in one of the negatives done with just one oligonucletide.

After we found that *DlxC* and *DlxD* are linked with an intergenic region of size 11 Kb, we tried to amplify this region by PCR. Primers at the end of third exon of *DlxC* and *DlxD* going downstream of both genes were used to do this PCR. We only saw very faint bands close to the correct size. These bands were gel

extracted and used as template for another PCR with the same primers. This did not improve the PCR reaction. (*DlxC* primers: DlxC-Dn-3, DlxC-Dn-4, DlxC-Dn-4; *DlxD* primers: DlxD Dn-4, DlxD-Dn-5, DlxD In-situ For. See section 2.1 for sequences)

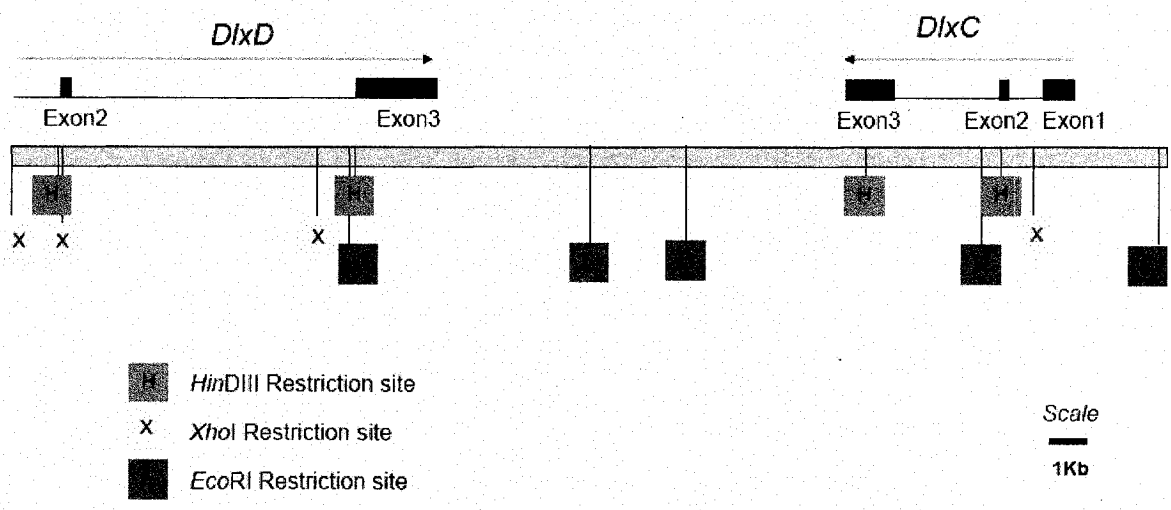
3.5. Restriction Map of *DlxC-D* locus (cosmid C3).

Out of the four lamprey *Dlx* genes only two, *DlxC* and *DlxD* are arranged in the same bigenic organization as in gnathostomes. In gnathostomes, this bigenic configuration is always accompanied by highly conserved enhancer elements in the intergenic region. It is hypothesized that it is the sharing of these enhancer elements that has maintained the two *Dlx* genes together for more than 400 million years (evolutionary distance between humans and zebrafish) (Zerucha, Stuhmer et al. 2000; Ghanem, Jarinova et al. 2003). No enhancer elements were detected near the two unpaired *Dlx* genes in teleosts (fugu, zebrafish and tetraodon) by sequence comparisons with mammals. Hence, there seems to be a good chance that lamprey *DlxC/D* intergenic region may also contain similar sequences. Therefore, to study this locus we decided to further characterize the lamprey *DlxC/D* locus.

For cosmid C3 that contains the complete *DlxC* gene, the intergenic region and a large portion of the *DlxD* region (up to the first intron), we have been able to make a preliminary restriction map, which has *Eco*RI, *Hin*dIII and *Xho*I restriction sites mapped onto it (see figure 3.2). This restriction map has helped us subclone pieces of DNA from the *DlxC,D* locus for making transgenic animals (to test enhancer activity with a reporter gene driven by a minimal promoter) and also for sequencing. Some of these fragments (*DlxC* promoter region and *DlxC*/D intergenic region) are being used in reporter gene assays in mice and zebrafish.

Figure 3.2: Restriction map of Cosmid C3 along with position and orientation of the *DlxC* and *DlxD* genes.

The exon-intron structure of the two genes is also displayed with green boxes representing exons.


DlxD
DlxC
Exon2
Exon3
Exon3
Exon2
Exon1
H
X
HindIII Restriction site
XhoI Restriction site
EcoRI Restriction site
Scale
1Kb

The map was generated using a combination of techniques like PCR, Southern hybridization (on complete and partially digested cosmids), subcloning and sequencing. Specific information on the methodology is provided in the following sections.

3.5.1. Restriction mapping by subcloning cosmid C3 *Eco*RI fragments:

The cosmid was digested with *Eco*RI and the fragments (9.2, 8.1, 6.7, 5.1, 2.5 kb in size, not including the backbone, Lawrist 7) were subcloned individually by purifying each fragment using gel extraction, followed by subcloning into pBluescript. White colonies on X-gal, IPTG-LB plates were picked and checked by restriction digestion for insert on plasmid preps.

As these smaller subcloned fragments contained fewer restriction enzyme sites, many of the restriction sites were mapped based on single and double digests followed by size estimation of fragments using agarose gel electrophoresis.

3.5.2. Restriction mapping by Southern on digested cosmid:

A number of southern hybridization experiments were done on digested cosmid clones using probes from either the *Dlx* gene regions or the ends of the cosmid to identify fragments containing particular sequences. Here, I present a few such southern hybridization experiments and show how useful restriction mapping information was inferred from them.

Figure 3.3 shows Southern hybridization on digested cosmids (A, C3 and D) with T7 promoter sequence as probe. The T7 sequence is located on one side of the cosmid insert in the multicloning site of the backbone (Lawrist 7). For *Eco*RI which cuts in the cosmid backbone (close to T7, but further into the vector backbone (see figure 3.4)), we can say that the most immediate *Eco*RI restriction site into the insert occurs at 3.4, 8 and 3.1 kb for cosmids A, C3 and D

respectively. Similarly, the closest *Xho*I sites into the insert from the T7 side are at 1.3, 0.9 and 2.4 kb for cosmids A, C3 and D respectively (see figure 3.4). *Bam*HI and *Hin*dIII do not cut within the vector backbone and cannot be mapped with just the Southern. This was also done for the SP6 promoter sequence for mapping restriction sites at the other end of the insert (see figure 3.4).

Figure 3.5 shows a Southern hybridization experiment done on the above mentioned blot with an oligonucleotide probe (DlxC Ex1 Dn, see section 2.1 for sequence) from the first exon of *DlxC*. From this southern we knew that cosmid-D does not contain this region of *DlxC* and the 5Kb *Eco*RI fragment of cosmid-C3 contains the *DlxC* first exon.

Figure 3.3: Southern hybridization experiment done on digested cosmids with T7 promoter as probe.

Left Image: Cosmids A, C3 and D digested with *Eco*RI (E), *Xho*1 (X), *Bam*HI (B) and *Hin*dIII (H) separated on a 0.6% agarose gel. Right Image: Southern hybridization on the gel shown to the left with T7 promoter sequence as probe. Positive bands are marked by white ovals on the gel image. Two ladders were used; GeneRuler™ DNA Ladder Mix (L) and Lambda DNA/HindIII Marker (λ), both from Fermentas.

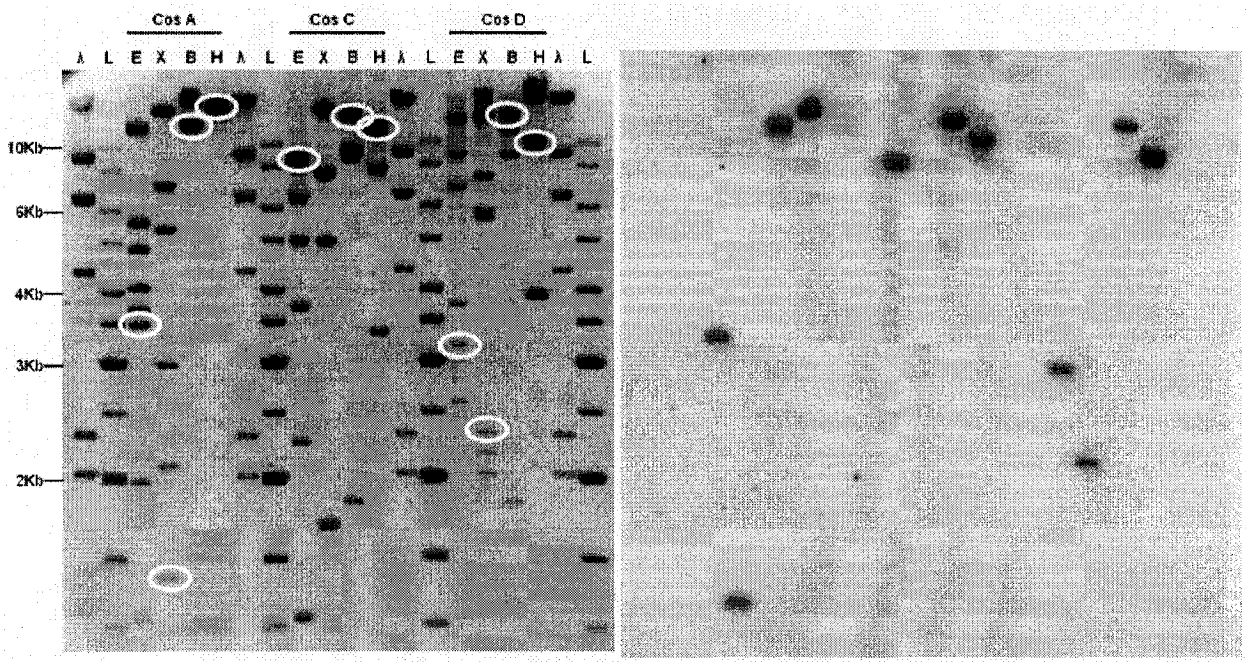

Cos A
Cos C
Cos D
λ L E X B H λ L E X B H λ L E X B H λ L
10Kb
6Kb
4Kb
3Kb
2Kb

Figure 3.4: *Eco*RI, *Xho*I restriction sites closest to the ends of cosmids A, C3 and D are displayed.

The sizes of the fragments were estimated using figure 3.3 and are expressed in Kb. The solid line represents the insert, whereas the dotted line represents the MCS.

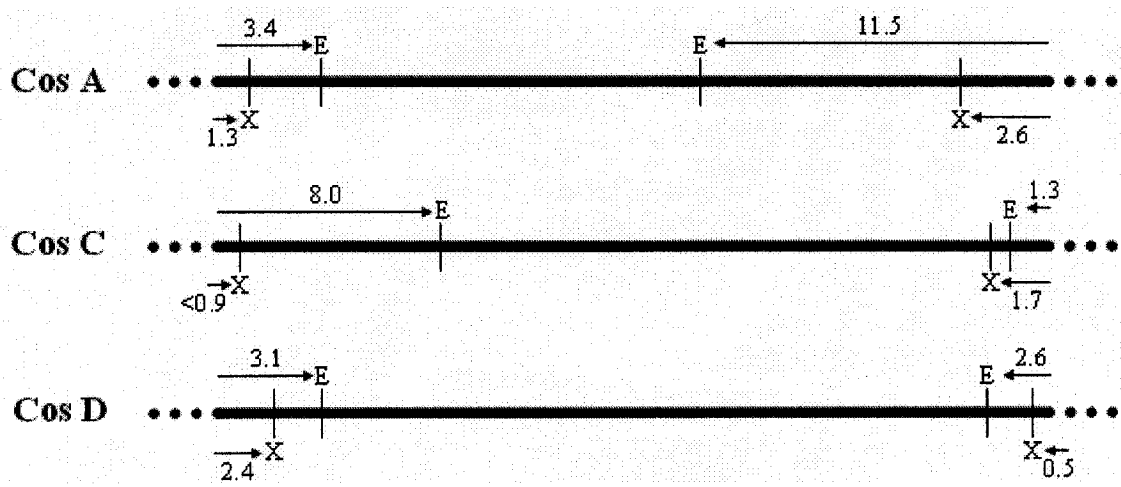

Cos A
3.4
E
E
11.5
1.3
X
X
2.6
Cos C
8.0
E
E
1.3
<0.9
X
X
1.7
Cos D
3.1
E
E
2.6
2.4
X
X
0.5

Figure 3.5: Southern hybridization experiment done on digested cosmids with oligonucleotide probe in the first exon of *DlxC*.

Left Image: Cosmids A, C3 and D digested with *Eco*RI (E), *Xho*1 (X), *Bam*HI (B) and *Hin*dIII (H) separated on a 0.6% agarose gel. Right Image: Southern hybridization on the gel shown to the left with an oligonucleotide probe in *DlxC* first exon (*DlxC* ex1 Dn, see section 2.1). Positive bands are marked by white ovals on the gel image.

Two ladders were used; GeneRuler™ DNA Ladder Mix (L) and Lambda DNA/HindIII Marker (λ), both from Fermentas.

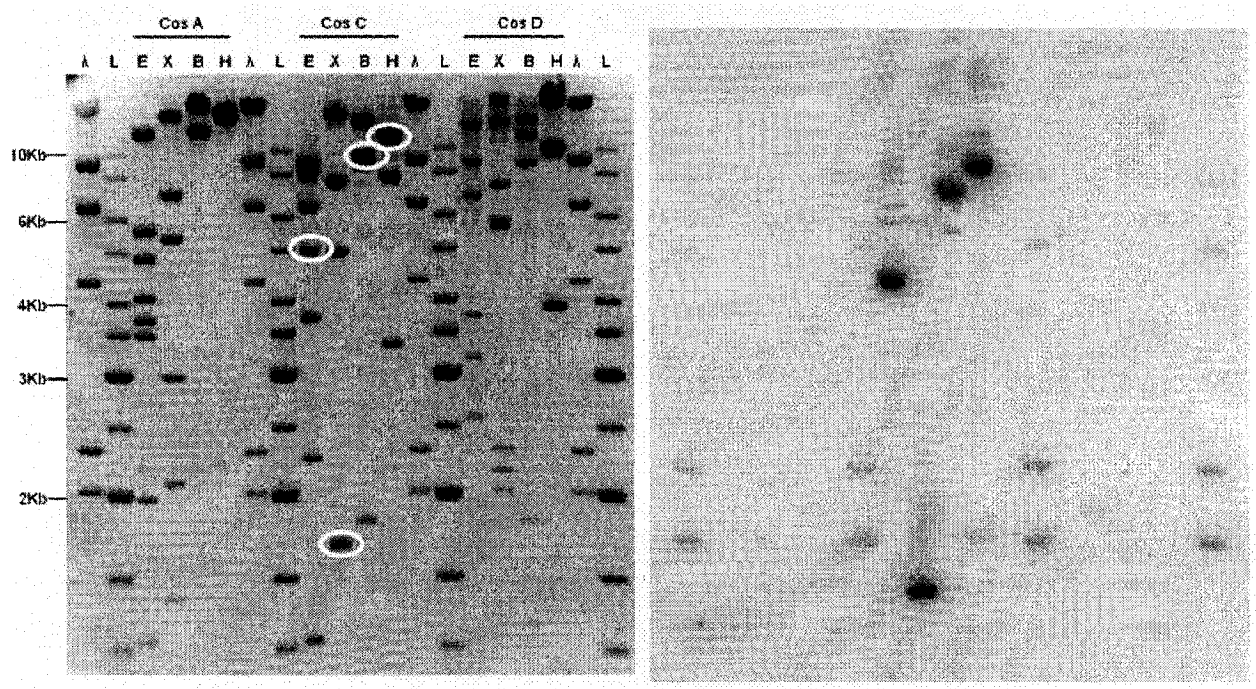
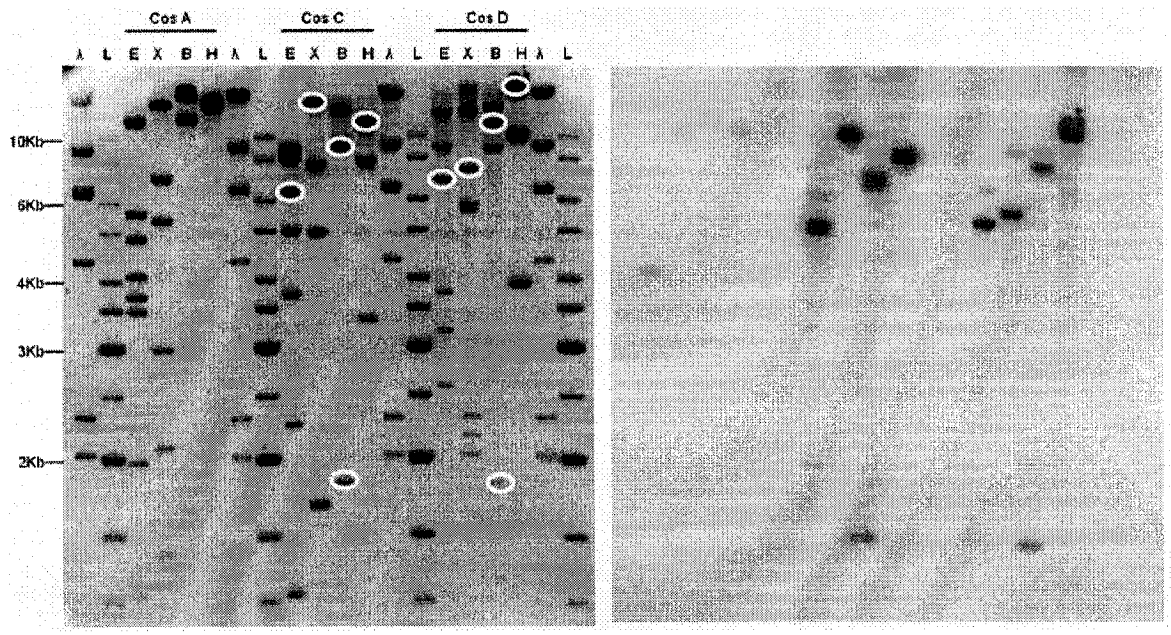

Cos A
Cos C
Cos D
λ L E X B H λ L E X B H λ L E X B H λ L
10Kb
6Kb
4Kb
3Kb
2Kb

Figure 3.6: Southern hybridization experiment done on digested cosmids with oligonucleotide probe from the third exon of *DlxD*.

Left Image: Cosmids A, C3 and D digested with *Eco*RI (E), *Xho*1 (X), *Bam*HI (B) and *Hin*dIII (H) separated on a 0.6% agarose gel. Right Image: Southern hybridization on the gel shown to the left with an oligonucleotide probe in the third exon (3' UTR) of *DlxD* (*Dlx*D In-situ Rev, see section 2.1 for sequence). Positive bands are marked by white ovals on the gel image.

Two ladders were used; GeneRuler™ DNA Ladder Mix (L) and Lambda DNA/HindIII Marker (λ), both from Fermentas.


Cos A
Cos C
Cos D
λ L E X B H λ L E X B H λ L E X B H λ L
10Kb
6Kb
4Kb
3Kb
2Kb

3.5.3. Mapping by partial digestion of linearized cosmid followed by Southern hybridization with probe at one end of cosmid:

To put these *Eco*RI pieces in order and to combine the restriction sites obtained from the smaller sub-cloned *Eco*RI fragments, the cosmid was linearized with *Not*I followed by partial digestion with *Eco*RI. Digested fragments were separated on a 0.5% agarose gel. Southern hybridization was performed with a probe at one end of the cosmid. The order of *Eco*RI fragments was inferred by the size difference of neighboring fragments that showed on the Southern starting from the lowest sized fragment (for schematic of concept see section 2.14, figure 2.1; for the actual Southern and inferred mapping information see figure 3.7 and 3.8 respectively). This was done from both sides of the cosmid. This technique is limited by the fact that as we go above 10-12 kb sized fragments, accurate size estimation (within 1-2 kb range) becomes difficult. Therefore, we also sequenced the ends of the *Eco*RI subcloned fragments. This when combined with the sequence obtained from the cosmid by primer walking enabled us to piece together a restriction map for the whole cosmid-C3.

Figure 3.7: Southern hybridization experiment done on linearized cosmids partially digested with *Xho*I and *Eco*RI with T7 oligonucleotide probe.

This was done for mapping restriction sites from one end of the cosmid (see section 2.14 and figure 2.1 for concept)

Left Image: Linearized (with *Not*I) cosmid C3 partially digested with *Xho*I and *Eco*RI and separated on a 0.5% agarose gel. Degree of partial digestion increasing from left to right. Right Image: Southern hybridization on the gel shown to the left with T7 promoter oligonucleotide probe (lies at one end of the cosmid after linearizing). The 8, 25 kb *Xho*I bands and the 5, 6, 13 kb *Eco*R1 bands are considered positive on the southern.

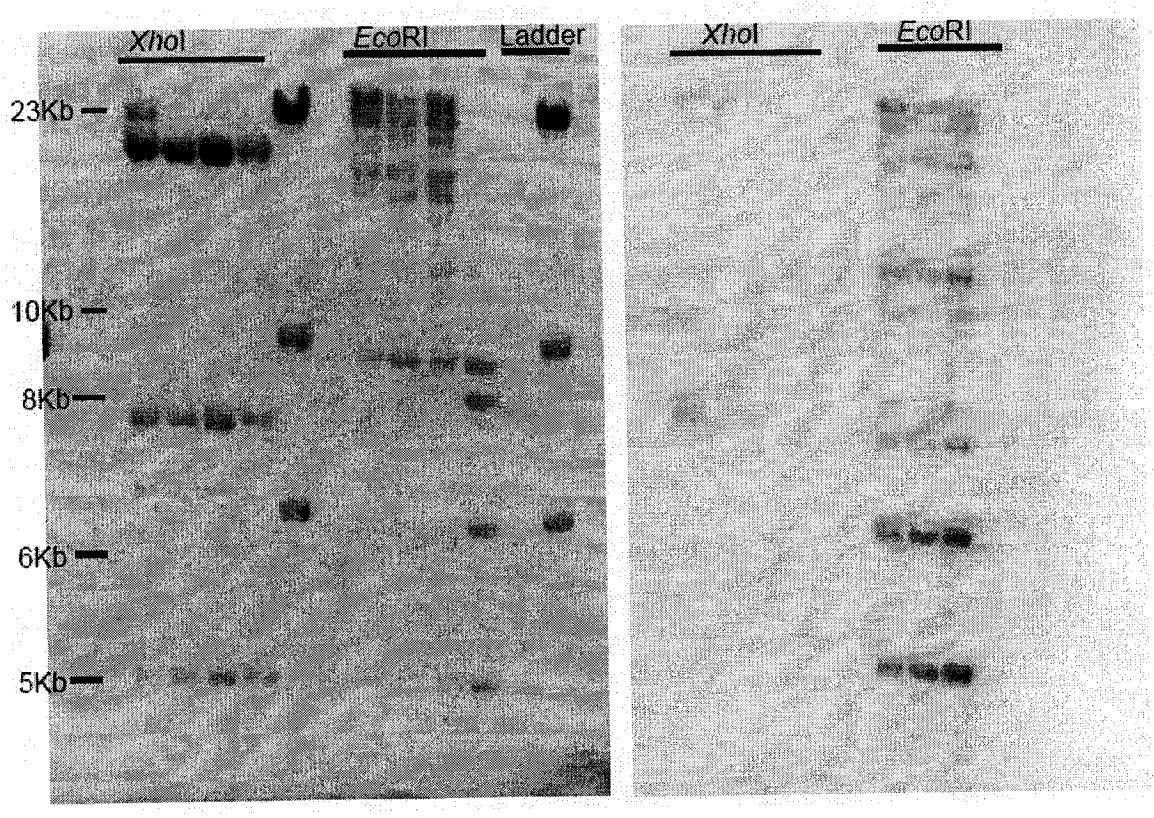
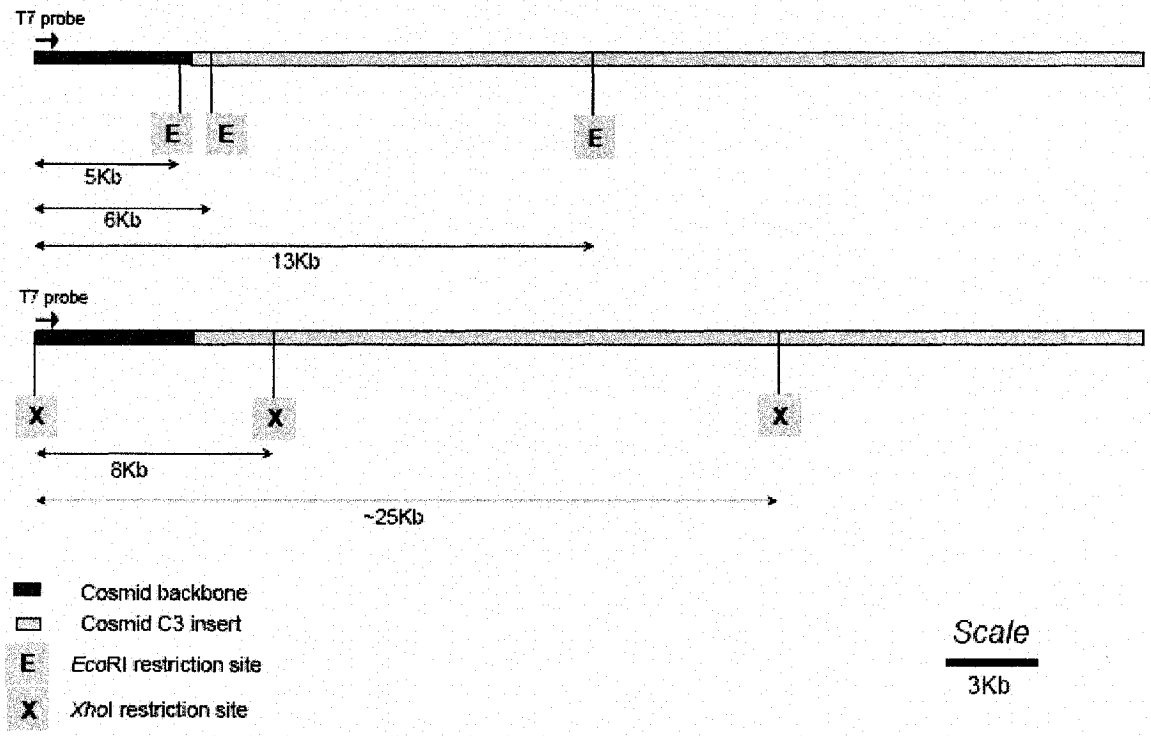

XhoI
EcoRI
Ladder
23Kb
10Kb
8Kb
6Kb
5Kb
XhoI
EcoRI

Figure 3.8: Schematic showing how information from the Southern blot shown in figure 3.7 was inferred to map some of the restriction sites, starting from one end.


T7 probe
E
E
E
5Kb
6Kb
13Kb
T7 probe
X
X
X
8Kb
~25Kb
Cosmid backbone
Cosmid C3 insert
E EcoRI restriction site
X XhoI restriction site
Scale
3Kb

3.6. Shotgun sequencing of *DlxA, DlxC* and *DlxD* containing cosmids.

It is possible that there are enhancers similar in sequence to that of other vertebrates in lamprey, but our PCR and Southern experiments were unable to detect them simply because of significant sequence divergence. Therefore, we have started shotgun sequencing of some of our cosmids. We have selected comid A, C2 and D for complete shotgun sequencing. These three cosmids were selected as they overlapped least and span the most DNA flanking the three *Dlx* genes. In shotgun sequencing technique, we fragment the complete cosmid randomly by sonication, treat the end of these fragments, size select the fragments (0.8-1.5kb in our case), subclone these smaller fragments into a plasmid vector and sequence a large number of these subclones. The obtained sequences are then aligned amongst each other to obtain sequence assemblies.

We currently have 76, 128, 104 sequence reads from cosmid A, C2 and D shotgun subcloned fragments. These reads assuming generate 500bp of good sequence approximately amount to 1X, 1.8X and 1.2X coverage of the cosmids A, C2 and D. As discussed in section 3.1, sequencing done by primer walking method gave us about 12, 7, 5 kb of sequence from the *DlxA, C, D* genes respectively. Figure 3.9 displays the regions of the *DlxC/D* locus that have been sequenced and whose identities are known.

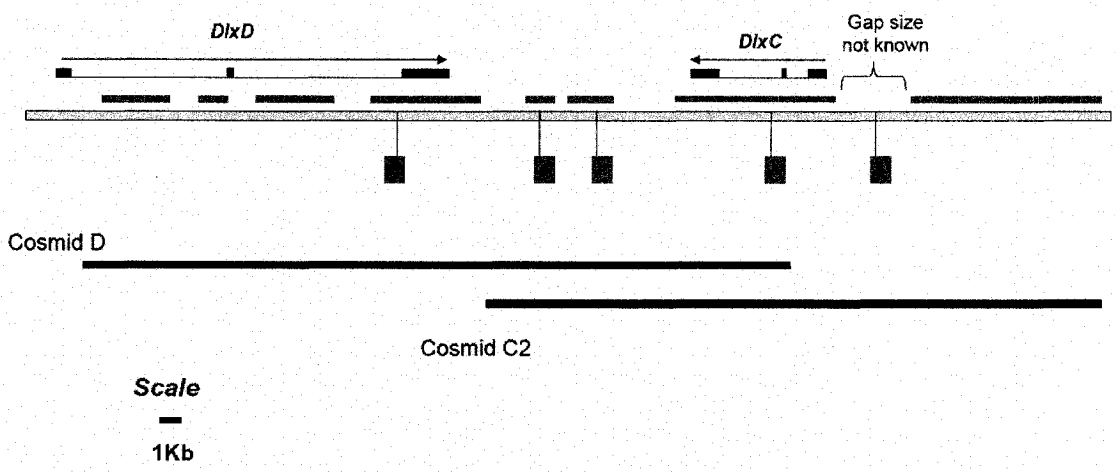
There is also a possibility that none of the vertebrate *Dlx* enhancers are present in lamprey. Obtaining the sequence in and around the lamprey *Dlx* genes would still be informative, as comparing these sequences with that of a closely related species may help identify cis-regulatory elements. The genome of another lamprey species i.e. *Petromyzon marinus* is being sequenced by the WUSTL (Washington University at St. Louis) genome sequencing centre. However, we have recently found that the first release of the *Petromyzon marinus* (which was

due Dec 2005) genome sequence has been postponed indefinitely due to assembling difficulties.

From the contigs and sequence reads generated by shotgun sequencing of our cosmids, we tried to search for the vertebrate intergenic enhancer like sequences using the find function in the contig express program (vector NTI, Invitrogen). We used approximately 12 bp queries (both sense and antisense) from many different regions of the intergenic enhancers. We were unable to find any homologous sequences from the shotgun reads and assemblies obtained from the cosmid.

Figure 3.9: The *DlxC, DlxD* locus displaying the regions (orange) that have been sequenced.

Sequences were obtained from cosmid C2 and D. Regions spanned by the cosmids are also displayed by green lines. The grey E-boxes are *Eco*RI sites within cosmid C3.

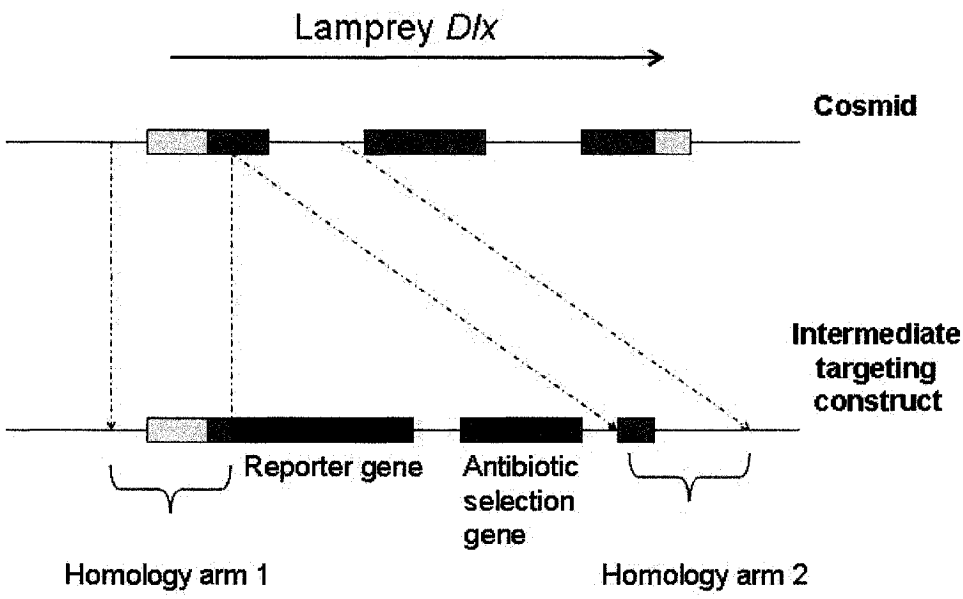

DlxD
DlxC
Gap size
not known
Cosmid D
Cosmid C2
Scale
1Kb

3.7. Homologous recombination in Bacteria to insert reporter genes at lamprey *Dlx* loci.

In order to test the DNA flanking lamprey *Dlx* genes for enhancer activity, we tried to insert *Egfp* (for zebrafish) and *lacZ* (for mice) reporter genes in frame in the first exon of lamprey *Dlx* genes. This can be done by homologous recombination in a modified strain of *E.coli* (DY380), which can be induced to increase homologous recombination rates by simply incubating the bacteria at 42°C (Yu, Ellis et al. 2000). This bacterial strain was produced by inserting a defective lambda prophage containing the genes required for homologous recombination from phage lambda into DH10B *E.coli* strain (Yu, Ellis et al. 2000).

For performing the homologous recombination in DY380 a linear targeting vector containing the reporter gene, an antibiotic selection gene flanked by two regions of homology (250-300bp each) to the site of insertion are required (See figure 3.10). This linear targeting vector alone cannot be propagated in DY380 unless it gets incorporated into the cosmid. The incorporation is done by homologous recombination and clones are selected for antibiotic resistance carried by the linear targeting vector.

Figure 3.10: Position and orientation of the various components in the linear targeting construct.


Lamprey *Dlx*
Cosmid
Intermediate targeting construct
Reporter gene
Antibiotic selection gene
Homology arm 1
Homology arm 2

The sequence near the first exon of *Dlx* genes (*DlxA* and C) was determined by primer walking method from the cosmids (A & C3) and 250-300bp homology arms amplified and subcloned in the desired orientation flanking the reporter gene (*lacZ* and *Egfp*) and ampicillin resistance gene. Design of primers for the amplification of the first arm was critical as the reading frame of the *Dlx* gene in the first homology arm had to be in frame with the reading frame of the reporter gene.

Currently, we have the linear targeting constructs of *lacZ* and *Egfp* reporter genes for *DlxA* and for *DlxC*. Homology arms of all four linear targeting constructs have been sequenced to confirm that the reading frames of the reporter genes and the *Dlx* genes are the same and are in the right orientation.

In both *DlxA* targeting constructs, codons for 28 amino acids from the *DlxA* gene have been added in front of the reporter genes. In both *DlxC* targeting constructs, codons for 31 amino acids from the *DlxC* gene have been added in front of the reporter genes.

To perform the homologous recombination, the first step involves transforming DY380 with our cosmids. We have found out our cosmids are unstable in this strain, probably because of the defective lambda prophage that these bacteria contain in their genomes (Yu, Ellis et al. 2000). This prophage shares 500bp of homology with our cosmid backbone (lawrist 7, which is also derived from lambda). We were once able to transform these bacteria with our cosmids, but as soon as we induced them for recombination, undesired recombination events resulting in rearrangements within the cosmid took place. We know this by comparing the restriction pattern of the cosmid before and after the induction.

These undesired rearrangements could also be due to high amounts of repetitive DNA within the cosmid insert. We know that the sea lamprey's (*Petromyzon marinus*) genome (being sequenced), has large amounts of repetitive DNA in its genome causing problems in its assembly from shotgun data (personal communication: from Dr. C. Amemiya, Benaroya Research Institute at Virginia Mason, Seattle).

Recently we have had some success with another method to perform homologous recombination in bacteria to engineer plasmids and one of the cosmids. In this method the "recombinases" (gam, bet and exo proteins coded by lambda genome) required for homologous recombination are expressed from a low copy plasmid pKD46 (Datsenko and Wanner 2000). Homologous recombination functions can be transiently induced by growing the bacteria in 10mM L-arabinose. The pKD46 plasmid also contains parts of the lambda genome but there is no homology between the cosmid backbone and pKD46 plasmid. Using the pKD46 system we were able to insert an *Egfp-ampicillin* gene cassette in frame into the first exon of *DlxA* on cosmid A. This was initially confirmed by restriction digestion followed by sequencing using a primer going upstream from the *Egfp* gene on the modified cosmid.

3.8. Homologous recombination in bacteria on plasmids.

In order to show that homologous recombination in bacteria (DY380 system) works well in our hands, we attempted to do it on plasmids.

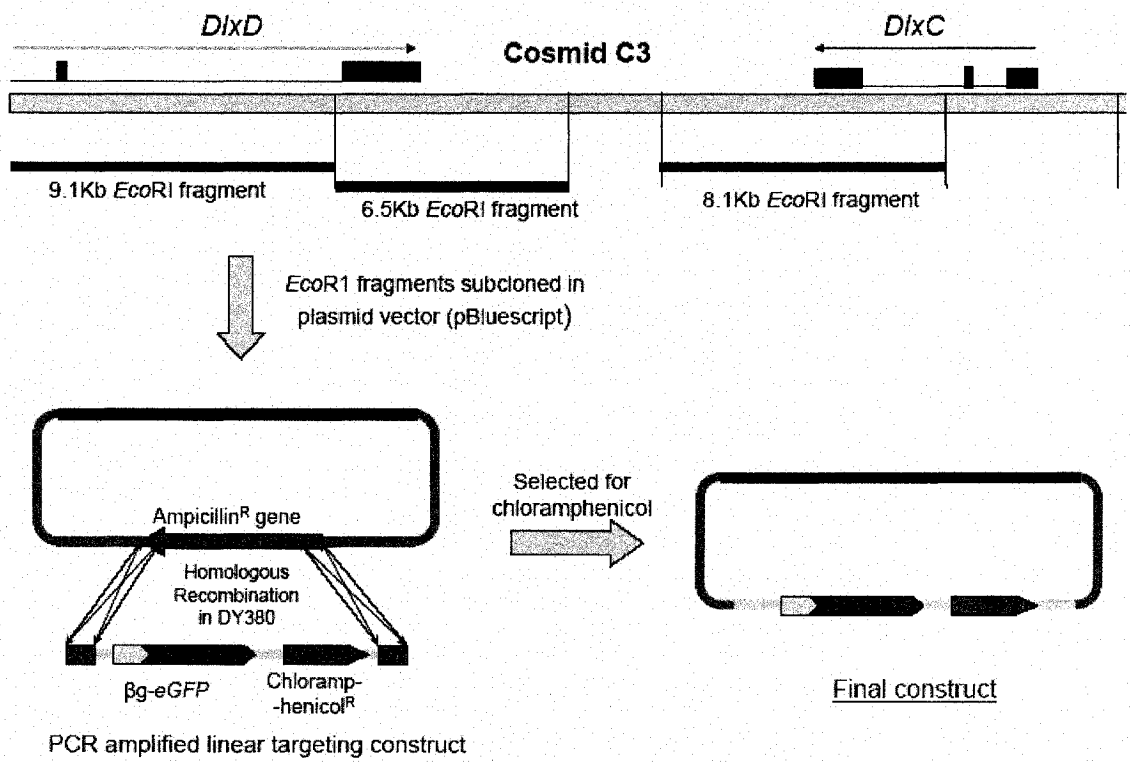
We replaced the ampicillin gene on pBluescript-KS (Stratagene) plasmid (containing the *Eco*RI fragments of cosmid C3, previously described in section 3.5.1) with a betaglobin-*Egfp*-chloramphenicol gene cassette using homologous recombination (see figure 3.11). When microinjected in zebrafish embryos these

constructs will help us locate enhancer activity to specific fragments of cosmid-C3.

To make the linear targeting construct beta-globin minimal promoter along with *Egfp* gene was subcloned into the *Not*I restriction site of pDrive (Qiagen) plasmid right next to the chloamphenicol gene previously amplified and subcloned by TA cloning (TA-cloning kit, Qiagen). The final targeting construct was made by PCR amplification of the betaglobin-egfp-chloramphenicol cassette by long (100bp) synthetic oligonucleotides (Amp Bg et and Amp cm et, sequences provided in section 2.1) containing 80bp of homology at their 5′ ends to the ampicillin resistance gene on pBluescript plasmid (see figure 3.11).

Figure 3.11: Schematic showing strategy for homologous recombination in bacteria on plasmids.

Plasmid shown contains EcoRI fragments of cosmid C3. Homologous recombination in bacteria was used to insert a betaglobin minimal promoter-*Egfp*-chloramphenicol gene cassette into the plasmid backbone (replacing the ampicillin resistant gene).


DlxD
Cosmid C3
DlxC
9.1Kb EcoRI fragment
6.5Kb EcoRI fragment
8.1Kb EcoRI fragment
EcoR1 fragments subcloned in plasmid vector (pBluescript)
Ampicillin^R gene
Homologous Recombination in DY380
βg-eGFP
Chloramp--henicol^R
PCR amplified linear targeting construct
Selected for chloramphenicol
Final construct

3.9. Reporter gene constructs.

3.9.1. Lamprey *DlxC/D* intergenic constructs:

Based on the restriction map we recently subcloned a 14 kb *Hin*dIII fragment from cosmid C3 that contains the complete *DlxC/D* intergenic region (see figure 3.12). Several attempts were made to gel extract the desired fragment after digesting the cosmid, followed by subcloning in plasmid pBluescript (Stratagene), but without success. Later we digested the cosmid, purified all the digested fragments together, subcloned all the *Hin*dIII fragments into pBluescript. Approximately 2000 colonies were screened for the desired fragment by southern hybridization on colony lifts with a *DlxC* third exon probe (DlxC-IS-Rev, see section 2.1 for sequence). Approximately 10-15 positive clones were identified a few checked by PCR and restriction digestion and finally one of them was checked by sequencing.

This 14kb *Hin*dIII fragment starts in the third exon of *DlxC* and goes downstream of *DlxC* until the third exon of *DlxD*, where it ends just before the 3' end of homeobox sequence (see figure 3.12). This 14kb fragment was excised from pBluescript and subcloned into a modified SP72 vector at the *Hin*dIII restriction enzyme site in multicloning site just upstream of *Egfp* (see figure 3.12). The modified SP72 vector contains human betaglobin minimal promoter in front of the *Egfp* gene. The betaglobin minimal promoter by itself does not drive expression of the reporter gene that can be detected in the transgenic animals (mice and zebrafish). It provides a transcriptional start site and is the site where the transcriptional machinery (RNA Polymerase II and the general transcription factors) of the cell assembles. If an enhancer sequence is placed adjacent to such a minimal promoter with a reporter gene, it usually would enhance the expression of the reporter gene so that it would be visibly detectable in transgenic animals.

The 14kb fragment was also subcloned into p1230 vector at *Hin*dIII restriction enzyme site upstream of the *lacZ* gene with beta-globin minimal promoter (see figure 3.12). This construct is being used for transient transgenesis in mice embryos.

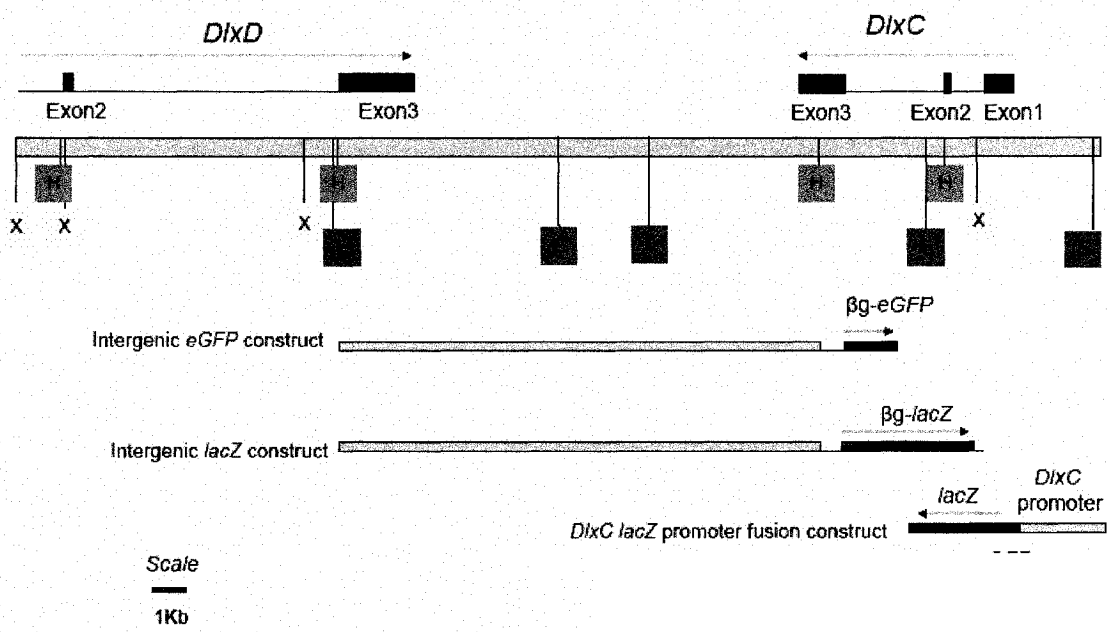
3.9.2. *DlxC* promoter construct:

A 2.5kb promoter region upstream of the start codon of *DlxC* was PCR-amplified from cosmid C3 (Oligonucleotides: Cosmid C3 SP6 side, DlxC-H2-Rev, see section 2.1 for sequences) and subcloned in front of *LacZ* gene (previously subcloned in pCR2.1 (Invitrogen)) at the *Kpn*I restriction enzyme site, in the right orientation. This 2.5 kb PCR amplified fragment also contains about 31 amino acid codons from the *DlxC* gene. Subcloning strategy was such that the reading frames of *DlxC* and *lacZ* genes remained same. This construct would also be used to produce primary transgenic mice.

3.9.3. Cosmid C3 *Eco*RI fragment constructs:

As described in section 3.8 and displayed in figure 3.12, we inserted betaglobin-*Egfp-chloramphenicol* gene cassette into plasmids containing the 6.5, 8.1, 9.1 Kb *Eco*RI fragments of cosmid C3. The 6.5 and 8.1 kb fragments contain a large part of the intergenic region and these two constructs when tested in zebrafish would help us locate enhancer activity in smaller regions of the intergenic region. The 9.1kb construct contains the complete *DlxD* second intron which is approximately 7kb in size. This construct would be tested in zebrafish if the intergenic region does not display enhancer activity. The rationale is, if the intergenic region does not contain enhancers, then other regions like this 9kb fragment would. This 9kb fragment spans more than one-fourth of the whole cosmid C3.

Figure 3.12: Reporter gene constructs made using *Hin*dIII intergenic fragment and PCR amplified *DlxC* promoter region.


DlxD
DlxC
Exon2
Exon3
Exon3
Exon2
Exon1
X
X
X
X
βg-eGFP
Intergenic eGFP construct
βg-lacZ
Intergenic lacZ construct
lacZ
DlxC promoter
DlxC lacZ promoter fusion construct
Scale
1Kb

3.9.4. Cosmid A Egfp construct:

The modified pKD46 plasmid with chloramphenicol resistant gene in place of ampicillin was used to test if this system can be used to modify the cosmids. *E.coli* strain DH5α was transformed with pKD46-cm plasmid followed by cosmid A (containing *DlxA*). Colonies were grown at 30°C and selected on both kanamycin and chloramphenicol plates. Single colonies were picked and checked for both pKD46-cm and cosmid A by restriction digestion. Positive bacteria were induced for homologous recombination functions by growing them in 10mM arabinose for 1hr just before the cells were made electrocompetent. The *Egfp* linear targeting construct was electroporated and plated on kanamycin and ampicillin plates. Positive colonies were checked with restriction digestion. One of the positive clones was also sequenced to check the reading frames of *DlxA* and *Egfp*. The whole modified cosmid A was used for transgenesis in zebrafish.

3.10. Activity of the lamprey *DlxC/D* intergenic region in zebrafish and mice embryos.

The complete *DlxC/D* intergenic region when placed adjacent to *Egfp* driven by the human beta-globin minimal promoter (see figure 3.12), can target expression of *Egfp* in a tissue specific manner in zebrafish. Of about 150 injected embryos that showed some *Egfp* expression, 30-40 % show expression in the forebrain, 20-40 % show expression in the branchial arches, 80-90 % show expression in trunk somites and a much smaller proportion of primaries show expression in the hindbrain, heart, blood cells and in the tail fin (see figure 3.13). *Egfp* expression was detected from 1.5dpf and continued till after 4dpf in these embryos.

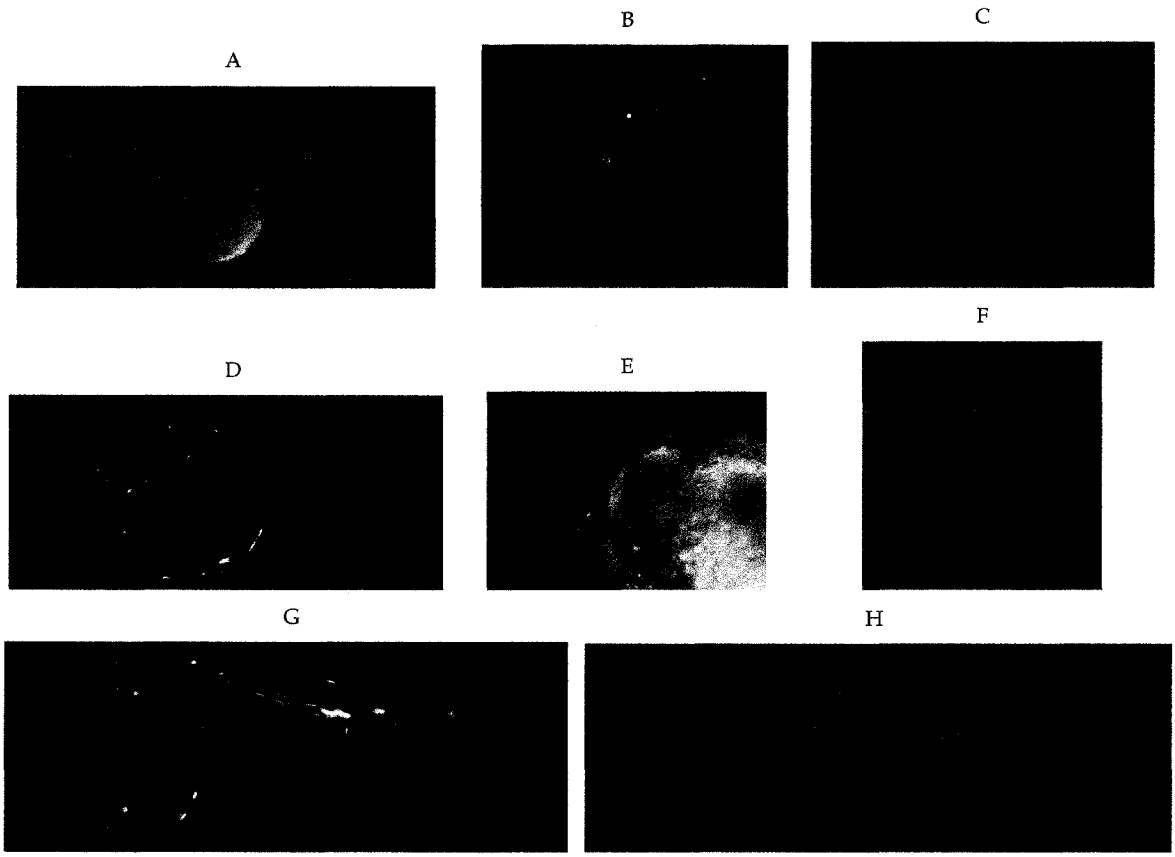
Transient transgenesis in zebrafish is highly mosaic and only a preliminary qualitative idea about expression of the transgene can be obtained. Therefore, we are in the process of making stable transgenic lines with this construct. Making transgenic lines involves producing large number of primary transgenic animals

(100-300), raising them to adulthood (takes about 3-4 months), breeding them one by one with wild type zebrafish followed by screening of progeny for *Egfp* expression under the fluorescent microscope.

The *lacZ* construct with the lamprey intergenic region (described in section 3.9.1) was injected in mouse 1-2 cell staged embryos and were harvested at embryonic day 11.5. In total, 20 embryos were analyzed by beta-galactosidase staining and also by PCR for the presence of *lacZ* gene in genomic DNA from placentas. Three placentas were found positive by PCR, whereas none of the embryos showed beta-galactosidase activity.

Figure 3.13: Primary transgenic, 2dpf zebrafish embryos injected with *Dlx*C/D intergenic *Egfp* construct (see figure 3.12).

A, B, C, D display *Egfp* expression in parts of the branchial arches. A, B, D, G H show expression in the trunk somites. A, C, D, E, F show expression in the forebrain. In C, it looks like the expression id in two domains in the forebrain, telencephalon and diencephalons.


A
B
C
D
E
F
G
H

3.11. Activity of Cosmid-C3 8.1kb *Eco*RI fragment in zebrafish embryos.

The cosmid-C3 8.1Kb *Egfp* construct discussed in section 3.9.3 was used to make primary transgenic zebrafish embryos. Out of approximately 80 embryos that displayed any egfp expression, almost 80-90% displayed it in the tail close to the prospective anal fin (most likely in the apical ectodermal ridge). Unlike the intergenic construct, no consistent expression in the forebrain, branchial arches or trunk somites was seen with this construct. This means that the cis-regulatory elements in the intergenic region that display enhancer activity in the forebrain, branchial arches and trunk somites are not within this 8.1kb fragment (see figure 3.11 and section 3.8).

3.12. Activity of Cosmid-A *Egfp* construct in zebrafish embryos

The complete circular cosmid A with *Egfp* in frame with the first exon of *DlxA* was used to make primary transgenic zebrafish. Approximately 200 embryos were microinjected with this construct out of which approximately 80% displayed *Egfp* expression at 1.5dpf. About 80% of the expressing embryos displayed expression in the apical ectodermal ridge of either the prospective pectoral, anal and/or caudal fins (see figure 3.15). About 20-30% embryos showed expression in the branchial arches and in the trunk somites.

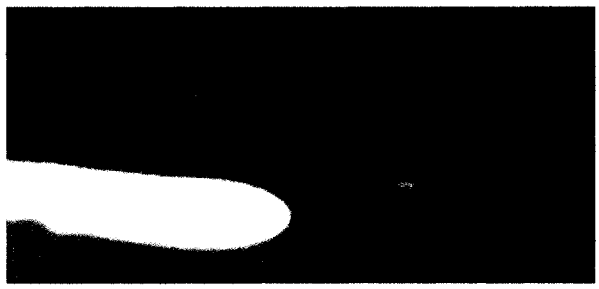
The lamprey *DlxA* construct displays strong and consistent activity in the zebrafish pectoral fin, a structure completely absent in lamprey. From this we can conclude that the genetic changes responsible for lampreys not having pectoral fins lies upstream of *DlxA* in the genetic cascade involved in limb/fin development.

Figure 3.14: Primary transgenic, 2dpf zebrafish embryos injected with Cosmid-C3 8.1 kb EcoRI fragment construct.

Expression is seen in the epidermal cells of the trunk (A) and also in the apical ectodermal ridge near the anal fin (in B and C).

A

B



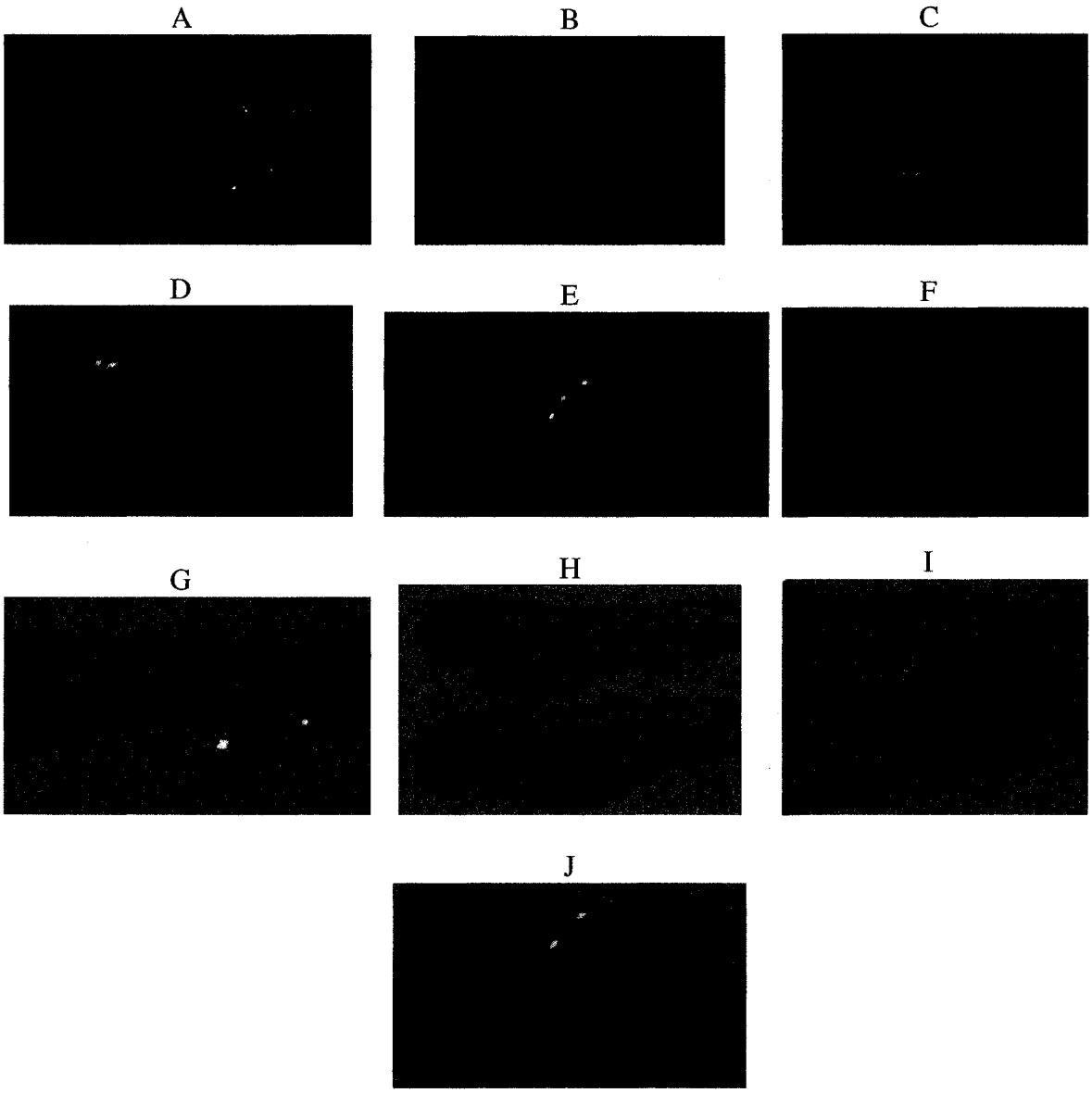
C

Figure 3.15: Primary transgenic zebrafish embryos injected with the Cosmid-A *Egfp* construct and GFP detected under a fluorescent microscope.

Cells A, B, C, D, E, F are 2 day post fertilization embryos

Cells G, H, I, J are 3 day post fertilization embryos

Cells A, B, G show expression in the apical ectodermal ridge of the anal and caudal fin. Cell F shows expression in trunk ganglia. Cells C, D, E, H show expression in parts of branchial arches. Cells E, J show expression in the pectoral fin.


A
B
C
D
E
F
G
H
I
J

4. Discussion.

The burst of character acquisition seen in vertebrates coincides with genome scale duplication events that lead to a greater number of developmental genes in vertebrates. Some of these characters include a tripartite brain, neural crest and sensory placodes, cranium, pharyngeal skeleton, true bones, true teeth, paired fins with pectoral and pelvic girdles, and jaws. Many of these features (like true bones, true teeth, paired fins with pectoral and pelvic girdles, and jaws) are absent in lampreys, a sister group of gnathostome vertebrates. The multiple duplicated *Dlx* genes have been shown to play critical roles in the development of most of these morphological features in gnathostomes. Here we are trying to understand how *Dlx* genes are regulated in the lamprey and to determine whether the differences in regulation of lamprey *Dlx* genes compared to those of gnathostomes provide insights into the acquisition of such morphological features (like jaws).

In this thesis, I have shown that lampreys have at least four *Dlx* genes two of which are arranged in the same tail to tail bigenic configuration found in other vertebrates. We could not find evidence for linkage of the remaining two lamprey *Dlx* genes to other yet uncharacterized *Dlx*. The lamprey *Dlx* bigenic cluster (*DlxC* and *D*) consists of approximately 11kb of intergenic region, which is relatively large when compared to those in other vertebrates (3-7 kb in teleost fish and 7-15 kb in mammals). We have also shown that the lamprey *Dlx* genes (A, C and D) have the same exon-intron structure as found in other vertebrates. We also found that introns of lamprey *Dlx* genes seem to be much larger than expected from that of other vertebrates. The first and second intron of *DlxD* are around 7Kb in size.

The presence of a *Dlx* bigene in lamprey further supports the idea that all vertebrate *Dlx* genes are derived from a single ancestral *Dlx* bigenic cluster (Stock, Ellies et al. 1996).

4.1. Homologous recombination in bacteria to insert reporter genes at lamprey *Dlx* loci.

Constructs made by homologous recombination to insert reporter genes at *Dlx* loci on cosmids are more likely to contain most of the *Dlx* regulatory elements as they contain a larger genomic region. These constructs would give us a more complete understanding of *Dlx* regulation in lamprey compared to testing of smaller fragments (promoter and intergenic regions) for transcriptional activity in transgenic animals. However, the latter experiments in combination with the homologous recombination constructs would give us valuable information about the location and interaction of enhancer elements with promoter regions.

To perform homologous recombination to modify BACs, PACs and cosmids, our lab uses the DY380 *E.coli* bacterial strain (Yu, Ellis et al. 2000). DY380s contain a defective lambda prophage that provides the homologous recombination functions to this strain. Expression of 'recombinases' is tightly controlled by a temperature sensitive repressor, resulting in induction of homologous recombination at higher temperatures (37-42°C). We were never able to stably propogate our cosmids in DY380, the first step involved in modification of our cosmids.

We have recently obtained a red recombinase plasmid pKD46 from the *E.coli* genetic stock centre at Yale University (Datsenko and Wanner 2000). This plasmid has been successfully used to engineer the *E.coli* genome using homologous recombination with short homology arms (Datsenko and Wanner

2000). We were recently able to stably propagate cosmid A along with the red recombinase plasmid pKD46 in *E.coli* (DH5α). We were also able to insert an *Egfp* and ampicillin resistance gene cassette in frame with the first exon of *DlxA* into this cosmid using the pKD46 plasmid recombination system.

The instability of lamprey cosmids in DY380 (*E. coli* strain, as discussed in section 3.7) could be due to at least two reasons: (1). Due to a 500bp homology between the cosmid backbone and the defective lambda prophage which provides the "recombinases" necessary for homologous recombination in DY380 genome or (2). Due to the fact that the cosmids contain excessive repetitive sequences that result in undesired rearrangements of the cosmids in the presence of the "recombinases" from DY380. We already know that the lamprey genome (*Petromyzon marinus*, sea lamprey) contains a larger proportion of repetitive DNA, which has caused the assembly of the shotgun data for the whole genome sequencing project very problematic (personal communication: from Dr. C. Amemiya, Benaroya Research Institute at Virginia Mason, Seattle).

The successful modification of cosmid A with the pKD46 system suggests the problem was most likely due to the 500bp homology of the cosmid backbone and the DY380 genome (not present in pKD46 system). There is also a possibility that the stability problem was amplified in DY380 due to higher amounts of recombinases in this system combined with the large amounts of repetitive sequences within the cosmid. We will know this was not the case if we are able to modify some of the other cosmids using this system (cosmid C2 and C3).

4.2. Evolutionary origin of vertebrate *Dlx* enhancers.

We were unable to detect the presence of the conserved intergenic arch enhancers (I12a, I56i) on the cosmids containing lamprey *Dlx* genes, at least by by PCR and Southern hybridization. Therefore, we propose that these sequences

emerged after gnathostomes diverged from agnathans. One may argue that these sequences may have been acquired before the divergence of lampreys but were lost in the lineage leading to lampreys. Chances of this happening seem less as in all the gnathostome species examined so far (human, mouse, rat, opossum, zebrafish, fugu, tetraodon, and frogs), these enhancer sequences have not only been retained but have been extremely well conserved over 400 million years of evolution.

To give further support to this statement one would need to investigate the presence of these vertebrate enhancer sequences in hagfish (only other extant agnathan). Although the phylogenetic relationship of hagfish with lampreys is disputed, morphologically they are simpler than lampreys and can be considered evolutionarily farther from gnathostomes. If intergenic enhancer like sequences are found in hagfish, then we may conclude that these sequences were present in the common ancestor of hagfish, lamprey and gnathostomes and were specifically lost in the lineage leading to lamprey. This hold true even if hagfish are considered to have diverged from lampreys after lampreys diverged from gnathostomes.

4.3. Activity of lamprey cis-acting regulatory elements in gnathostome genetic background.

Cross taxa transgenic analysis (lamprey *Dlx* regulatory elements in zebrafish and mice) like ours always carries the risk that the regulatory elements from one taxa of animals would not work as well in an animal from another. This risk is relatively low for *Dlx* genes as they are part of a conserved genetic cascade found in all metazoan species, components of which, and their interactions with each other wherever studied, have been found to be very well conserved (*FGFs*, *Aristaless* and *BMPs*).

It is possible that in agnathans there are cis-regulatory elements of *Dlx* which look very different from the intergenic enhancers in vertebrates but are functionally similar (bind to the same set of trans-acting factors). There are two very simplified possibilities of this situation pertaining to our cross taxa transgenic analysis. Either the gnathostome trans-acting factors identify and interact with the lamprey equivalents of the intergenic enhancers or they are unable to identify them (possibly due to simultaneous divergence of both the regulatory elements and the trans-acting factors involved in *Dlx* regulation in agnathans). The chances of the latter happening seem to be weak as some of these gnathostome intergenic enhancers display robust *Dlx* specific activity in *Drosophila* (which is evolutionarily farther away from gnathostomes than agnathans) (Zerucha 1999).

I have shown in this thesis that lamprey *DlxC/D* intergenic region can target expression of *Egfp* driven by betaglobin minimal promoter in the forebrain and branchial arches similar to its gnathostome counterparts. In addition, I have shown that DNA flanking the *DlxA* gene from lamprey can target robust reporter gene expression in zebrafish embryos.

4.3.1. Activity of lamprey *DlxC/D* intergenic region in gnathostomes

A 14 kb *Hin*dIII fragment containing the entire intergenic region was subcloned in reporter constructs (*egfp* and *lacZ* driven by betaglobin minimal promoter) and was tested in primary transgenic mice and zebrafish. This lamprey intergenic region can target expression of reporter gene *Egfp* in the forebrain, branchial arches and trunk somites when tested in primary transgenic zebrafish. None of the gnathostome *Dlx* genes are expressed in the trunk somites, whereas it has been reported by Neidert et. al. that lamprey *DlxA* and *D* both are expressed in trunk somites (Neidert, Virupannavar et al. 2001). We have had three mice embryos transgenic for *lacZ* with the intergenic construct but none showed any

expression. We need to make a larger number of transgenic mice before we can conclude anything about the activity of the intergenic region.

4.4. Regulatory evolution of lamprey *Dlx* genes.

The four lamprey *Dlx* genes do not segregate into the six orthologous gnathostome *Dlx* genes. Instead, they fall into the two subfamilies of gnathostome *Dlx* genes (*DlxD* in the gnathostome *Dlx1, 6, 7* clade and *Dlx A, B, C* in the *Dlx 2, 3, 5* clade). In gnathostomes each *Dlx* bigenic cluster contains one member from each of these subfamilies. Lamprey *DlxC* and *DlxD* form a bigenic cluster and are thought to have arose from a single bigene present in the common ancestor of gnathostomes and agnathans. However the duplication events that led to the *Dlx* complement in agnathans and gnathostomes seems independent because lamprey *Dlx* genes resemble the ancestral bigene more than they do individual gnathostome *Dlx* genes.

Even though we have so far found none of the vertebrate *Dlx* enhancers in lamprey, the lamprey *Dlx* intergenic region can target expression of reporter genes in patterns similar (forebrain and branchial arches) to that of the vertebrate intergenic enhancers when tested in zebrafish. These results along with others lets us propose a model for the regulatory evolution of *Dlx* gene family in vertebrates.

The common ancestor of gnathostomes and lampreys contained a *Dlx* bigenic cluster with cis-regulatory elements active in the forebrain and the branchial arches. This *Dlx* cluster was duplicated in both lineages independently (at least twice) and different cis-acting regulatory sequences were acquired and/or fixed in the two lineages after they diverged from each other. Assuming none of the vertebrate *Dlx* intergenic enhancers are present in lamprey.

5. Conclusions.

Here we have shown that lamprey have four *Dlx* genes, two of which are arranged in a convergently transcribed bigene cluster separated by an 11 kb intergenic region. To further study and manipulate the *DlxC/D* locus we have made a preliminary restriction map of this region. Further we have tested this intergenic region in transgenic zebrafish and mice in reporter gene assays. The intergenic region can target expression of reporter gene to the forebrain, branchial arches, and trunk somites in primary transgenic zebrafish embryo. We were unable to find the presence of the *Dlx* intergenic enhancers in the lamprey *Dlx* intergenic region by PCR and Southern hybridization experiments.

This data suggests that lamprey *Dlx* genes arose from a *Dlx* bigene cluster in the common ancestor of agnathans and gnathostomes. This ancestral intergenic region was active in the forebrain and branchial arch regions of the animal. The set of duplication events that gave rise to the present set of *Dlx* genes in lampreys and in gnathostomes occurred independently and different cis-acting regulatory sequences were acquired and/or fixed in the two lineages.

We have also shown that DNA flanking the lamprey *DlxA* gene in cosmid-A can target robust expression of *Egfp* in the apical ectodermal ridge of the pectoral fin, anal fin, and caudal fin of zebrafish embryos. The activity of this region in the pectoral fin of zebrafish suggests that the genetic changes underlying the absence of pectoral fins in lampreys are upstream of *Dlx* genes in the genetic cascade leading to limb development.

6. References.

Acampora, D., G. R. Merlo, et al. (1999). "Craniofacial, vestibular and bone defects in mice lacking the *Distal-less*-related gene *Dlx5*." Development **126**: 3795-3809.

Akimenko, M.-A., M. Ekker, et al. (1994). "Combinatorial expression of three zebrafish genes related to *distal-less*: part of a homeobox gene code for the head." J. Neurosci. **14**: 3475-3486.

Amores, A., A. Force, et al. (1998). "Zebrafish *hox* clusters and vertebrate genome evolution." Science **282**: 1711-1714.

Banerjee-Basu, S. and A. D. Baxevanis (2001). "Molecular evolution of the homeodomain family of transcription factors." Nucl. Acids Res. **29**(15): 3258-3269.

Bejerano, G., M. Pheasant, et al. (2004). "Ultraconserved elements in the human genome." Science **304**(5675): 1321-5.

Bulfone, A., H. J. Kim, et al. (1993). "The mouse *Dlx-2 (Tes-1)* gene is expressed in spatially restricted domains of the forebrain, face and limbs in midgestation mouse embryos." Mech. Dev. **40**: 129-140.

Cohen, S. M., G. Bronner, et al. (1989). "*Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*." Nature **338**: 432-434.

Datsenko, K. A. and B. L. Wanner (2000). "One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products." PNAS **97**(12): 6640-6645.

Depew, M. J., J. K. Liu, et al. (1999). "*Dlx5* regulates regional development of the branchial arches and sensory capsules." Development **126**: 3831-3846.

Depew, M. J., T. Lufkin, et al. (2002). "Specification of jaw subdivisions by *Dlx* genes." Science **298**: 381-385.

Diaz-Benjumea, F. J., B. Cohen, et al. (1994). "Cell interaction between compartments establishes the proximal-distal axis of *Drosophila* legs." Nature **372**: 175-179.

Dollé, P., M. Price, et al. (1992). "Expression of the murine *Dlx-1* homeobox gene during facial, ocular and limb development." Differentiation **49**: 93-99.

Eisenstat, D. D., J.-K. Liu, et al. (1999). "DLX-1, DLX-2 and DLX-5 expression define distinct stages of basal forebrain differentiation." J. Comp. Neurol. **414**: 217-237.

Ellies, D. L., D. W. Stock, et al. (1997). "Relationship between the genomic organization and the overlapping embryonic expression patterns of the zebrafish *dlx* genes." Genomics **45**: 580-590.

Gehring, W. J. (1987). The homeobox: structural and evolutionary aspects. Molecular approaches to developmental biology. R. A. Firtel and E. H. Davidson. New York, Liss: 115-129.

Ghanem, N., O. Jarinova, et al. (2003). "Regulatory roles of conserved intergenic domains in vertebrate *Dlx* bigene clusters." Genome Res. **13**: 533-543.

Gompel, N., B. Prud'homme, et al. (2005). "Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in Drosophila." Nature **433**(7025): 481.

Gorfinkiel, N., G. Morata, et al. (1997). "The homeobox gene Distal-less induces ventral appendage development in Drosophila." Genes Dev. **11**(17): 2259-2271.

Gregory, T. R. (2005). "Animal Genome Size Database."

Hogan, B., F. Constantini, et al. (1986). Manipulating the Mouse Embryo: A Laboratory Manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory Press.

Lecuit, T. and S. M. Cohen (1997). Nature **388(6638):** : 139-145.

Liu, J. K., I. Ghattas, et al. (1997). "*Dlx* genes encode DNA-binding proteins that are expressed in an overlapping and sequential pattern during basal ganglia differentiation." Dev. Dyn. **210**: 498-512.

McCauley, D. W. and M. Bronner-Fraser (2003). "Neural crest contributions to the lamprey head." Development **130**(11): 2317-2327.

McGinnis, W., M. Levine, et al. (1984). "A conserved DNA sequence in homeotic genes of the Drosophila antennapedia and bithorax complexes." Nature **308**: 478-483.

Nakamura, S., D. W. Stock, et al. (1996). "Genomic analysis of a new mammalian distal-less gene: Dlx7." Genomics **38**: 314-324.

Neidert, A. H., V. Virupannavar, et al. (2001). "Lamprey *Dlx* genes and early vertebrate evolution." Proc. Natl. Acad. Sci. USA **98**: 1665-1670.

Neumann, C. J. and S. M. Cohen (1997). "Long-range action of Wingless organizes the dorsal-ventral axis of the *Drosophila* wing." Development **124**: 871-880.

Panganiban, G. (2000). "*Distal-less* function during *Drosophila* appendage and sense organ development." Dev. Dyn. **218**: 554-562.

Panganiban, G. and J. L. R. Rubenstein (2002). "Developmental functions of the *Distal-less/Dlx* homeobox genes." Development **129**: 4371-4386.

Park, B. K., S. Sperber, et al. (2003). "Intergenic enhancers with distinct activities regulate Dlx gene expression in the mesenchyme of the branchial arches." submitted.

Price, M., M. Lemaistre, et al. (1991). "A mouse gene related to *distal-less* shows a restricted expression in the developing forebrain." Nature **351**: 748-751.

Qiu, M., A. Bulfone, et al. (1997). "Role of the Dlx homeobox genes in proximodistal patterning of the branchial arches: mutations of Dlx-1, Dlx-2, and Dlx-1 and -2 alter morphogenesis of proximal skeletal and soft tissue structures derived from the first and second arches." Dev. Biol. **185**: 165-184.

Qiu, M., A. Bulfone, et al. (1995). "Null mutations of *Dlx-2* results in abnormal morphogenesis of proximal first and second branchial arch derivatives and abnormal differentiation in the forebrain." Genes Dev. **9**: 2523-2538.

Quint, E., T. Zerucha, et al. (2000). "Differential expression of orthologous Dlx genes in zebrafish and mice: implications for the evolution of the Dlx homeobox gene family." J. Exp. Zool. (Mol. Dev. Evol.) **288**: 235-241.

Robinson, G. W. and K. A. Mahon (1994). "Differential and overlapping expression domains of *Dlx-2* and *Dlx-3* suggest distinct roles for *Distal-less* homeobox genes in craniofacial development." Mech. Devel. **48**: 199-215.

Robinson, G. W., S. Wray, et al. (1991). "Spatially restricted expression of a member of a new family of murine distal-less homeobox genes in the developing forebrain." New Biol. **3**: 1183-1194.

Robledo, R. F., L. Rajan, et al. (2002). "The *Dlx5* and *Dlx6* homeobox genes are essential for craniofacial, axial, and appendicular skeletal development." Genes Dev. **16**: 1089-1101.

Rossi, E., A. Faiella, et al. (1994). "order of six loci at 2q24-q31 and orientation of the *HOXD* locus." Genomics **24**: 34-40.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular cloning, a laboratory manual, 2nd Ed. New York, Cold Spring Harbor Laboratory Press.

Schwartz, S., Z. Zhang, et al. (2000). "PipMaker---A Web Server for Aligning Two Genomic DNA Sequences." Genome Res. **10**(4): 577-586.

Shigetani, Y., F. Sugahara, et al. (2002). "Heterotopic shift of epithelial-mesenchymal interactions in vertebrate jaw evolution." Science **296**: 1316-1319.

Simeone, A., D. Acampora, et al. (1994). "Cloning and characterization of two members of the vertebrate *Dlx* gene family." Proc. Natl. Acad. Sci. USA **91**: 2250-2254.

Stock, D. W., D. L. Ellies, et al. (1996). "The evolution of the vertebrate Dlx gene family." Proc. Natl. Acad. Sci. USA **93**: 10858-10863.

Sumiyama, K., S. Q. Irvine, et al. (2002). "Genomic structure and functional control of the Dlx3-7 bigene cluster." Proc. Natl. Acad. Sci. USA **99**: 780-785.

Sumiyama, K. and F. H. Ruddle (2003). "Regulation of Dlx3 gene expression in visceral arches by evolutionary conserved enhancer elements." Proc. Natl. Acad. Sci. USA **100**: 4030-4034.

Sunkel, C. E. and J. R. S. Whittle (1987). Roux's Archs. Dev. Biol. **196**: 124-132.

Thomas, B. L., A. S. Tucker, et al. (1997). "Role of *Dlx-1* and *Dlx-2* in patterning of the murine dentition." Development **124**: 4811-4818.

Vachon, G., B. Cohen, et al. (1992). "Homeotic genes of the *Bithorax* complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *distal-less*." Cell **71**: 437-450.

Westerfield, M. (1995). The zebrafish book. Eugene, University of Oregon Press.

Woolfe, A., M. Goodson, et al. (2005). "Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development." PLoS Biology **3**(1): e7.

Wu, J. and S. M. Cohen (1999). "Proximodistal axis formation in the Drosophila leg: subdivision into proximal and distal domains by Homothorax and Distal-less." Development **126**(1): 109-117.

Yang, L., H. Zhang, et al. (1998). "An early phase of embryonic *Dlx5* expression defines the rostral boundary of the neural tube." J. Neurosci. **18**: 8322-8330.

Yu, D., H. M. Ellis, et al. (2000). "An efficient recombination system for chromosome engineering in Escherichia coli." Proc Natl Acad Sci U S A **97**(11): 5978-83.

Zerucha, T. and M. Ekker (2000). "*Distal-less*-related homeobox genes of vertebrates: evolution, function, and regulation." Biochem. Cell Biol. **78**: 593-601.

Zerucha, T., T. Stuhmer, et al. (2000). "A highly conserved enhancer in the *Dlx5/Dlx6* intergenic region is the site of cross-regulatory interactions between *Dlx* genes in the embryonic forebrain." J. Neurosci. **20**: 709-721.

Zerucha, T. S. (1999). Evolution of auto- and cross-regulatory elements of members of the *distal-less*-related family of homeobox-containing genes. Cellular and Molecular Medicine. Ottawa, University of Ottawa: 192 pages.

Zhao, Z., D. W. Stock, et al. (2000). "Expression of *Dlx* genes during the development of the murine dentition." Dev. Genes Evol. **210**: 270-275.