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

Olga Jarinova

AUTEUR DE LA THÈSE / AUTHOR OF THESIS

Ph.D. (Cellular and Molecular Medicine)

GRADE / DEGREE

Department of Cellular and Molecular Medicine

FACULTÉ, ÉCOLE, DÉPARTEMENT / FACULTY, SCHOOL, DEPARTMENT

Molecular basis for functional diversification of duplicated *hoxb5* genes in zebrafish

TITRE DE LA THÈSE / TITLE OF THESIS

Marc 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

Ashley Bruce

David Lohnes

Guy Drouin

Michael McBurney

Gary W. Slater

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

Molecular basis for functional diversification of duplicated *hoxb5* genes in zebrafish

Olga Jarinova

This thesis is submitted as a partial fulfillment of the Doctor of Philosophy program in Cellular and Molecular Medicine

University of Ottawa

Department of Cellular and Molecular Medicine

Ottawa, Ontario, Canada

December, 2007

Your file Votre référence
ISBN: 978-0-494-49363-2
Our file Notre référence
ISBN: 978-0-494-49363-2

Statement of contribution

The study presented in the thesis "Molecular basis for functional diversification of duplicated *hoxb5* genes in zebrafish" was a collaborative effort between Dr. Marc Ekker and Dr. Lucie Jeannotte laboratories.

Radioactive in situ hybridization experiments on sections of mouse embryos were conducted by Dr. Josée Aubin, a research associate of Dr. Lucie Jeannotte. Analysis of transcription factor binding sites and generation of two primary transgenic embryos carrying MmJ2 and DrJ2a CNEs were performed by Dr. Christelle Prudhomme, a former postdoctoral fellow in the Dr. Jeannotte's laboratory.

In situ hybridization on zebrafish embryos, sequence analysis, generation of all the transgenic constructs, all zebrafish and a vast majority of mouse transgenic experiments were performed in the laboratory of Dr. Marc Ekker. The sequence and evolutionary analysis (phylogenetic footprinting, phylogenetic filtering, phylogenetic analysis etc.), in situ hybridization experiments on whole mount zebrafish embryos, quantitative RT-PCR, reporter construct design, mouse and zebrafish transgenic assays were performed by Olga Jarinova. Dr. Luc Poitras, a research associate in Dr. Ekker's laboratory has provided important technical advice and resources for the design and development of reporter constructs carrying large fragments of *hoxb5* loci via homologous recombination in bacteria. Lucille Joly, a previous research assistant in the laboratory, technically assisted with generation of reporter constructs containing β-globin minimal promoter and individual regulatory elements. Olga Jarinova has retrieved large fragments of *hoxb5a* and *hoxb5b* genes from the PAC clones and solely generated the *hoxb5alacZ*, the *hoxb5blacZ* and the *hoxb5a*J3Δ*lacZ* reporter constructs. Engineering of other large

reporter constructs (*hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*J3Δ*egfp*, *hoxb5b*insJ3*lacZ* and *hoxb5b*insJ3*egfp)* was peformed with the technical assistance of Gary Hatch, a research assistant, and Magdalena Grzyb, a former honour's student in the laboratory. All the transgenic mice were created by Adrianna Gamborotta. Zebrafish transgenic for β-globin carrying constructs were generated by Olga Jarinova. Generation of zebrafish transgenic for large reporter constructs was performed with technical assistance of Magdalena Grzyb, Vishal Saxena, a technical assistant, and Marion Alfero, a co-op student in Dr. Ekker's lab.

Abstract

The Duplication-Degeneration-Complementation (DDC) model predicts that division of the original function between duplicate genes (subfunctionalization) is a common mechanism for their preservation. The *Hox* genes constitute a good system to test this hypothesis as they underwent significant expansion during evolution. Mammals have four *Hox* complexes whereas teleosts have additional complexes resulting from duplication event(s) that took place after the separation of the ray-finned and lobe-finned lineages. Thus, zebrafish and *Takifugu* have two *hoxb* complexes, each containing a *hoxb5* gene. The zebrafish *hoxb5a* and *hoxb5b* genes are expressed in overlapping, yet distinct, domains during development with their combined expression patterns similar to that of the single mouse *Hoxb5* gene. To determine if changes in *cis*-acting regulatory elements account for *hoxb5a* and *hoxb5b* subfunctionalization, we compared the sequences of the *Hoxb5* loci of human, mouse, zebrafish and *Takifugu* and identified four conserved non-coding elements (CNEs). One of the CNEs, named J3, is only retained in the *hoxb5a* locus while three others are present in both teleost *hoxb5* loci. We examined individual and collective regulatory activity of the CNEs in transgenic assays. When tested individually, the enhancer activity of individual cognate CNEs from the *hoxb5a* and *hoxb5b* genes extensively overlapped. In contrast, large fragments encompassing multiple CNEs were able to target reporter gene expression to unique domains of *hoxb5a* and *hoxb5b* expression. The deletion of J3 from the *hoxb5a* locus and its insertion in the *hoxb5b* locus drastically altered the regulatory activity of the original loci. Our results suggest that patterns of regulatory evolution of *hoxb5* duplicates in teleosts may be more complex than ones taken in consideration by the DDC model.

Résumé

Le modèle DDC (Duplication-Dégénérescence-Complémentation) prédit que la division fonctionnelle entre gènes dupliqués (sous-fonctionnalisation) est un mécanisme fréquemment mis en œuvre pour expliquer leur rétention. Les gènes *Hox* constituent un bon système pour tester cette hypothèse puisque leur nombre a connu plusieurs augmentations au cours de l'évolution. Les mammifères possèdent quatre complexes *Hox* alors que l'on retrouve des complexes supplémentaires chez les poissons téléostéens suite à un ou des d'événement(s) de duplication qui se sont produits après la séparation des branches des actinoptérygiens et sarcoptérigiens. Ainsi, le génome du poisson-zèbre, *Danio rerio*, et celui du *Takifugu* contiennent chacun deux complexes *Hoxb*, incluant un gène *hoxb5*. Les gènes *hoxb5a* et *hoxb5b* du *Danio* sont exprimés selon des profils à la fois chevauchants et complémentaires au cours du développement. Leur expression combinée recrée celle du gène *Hoxb5* unique de la souris. Afin de déterminer si des changements survenus au niveau des éléments de régulation agissant en *cis* permettent d'expliquer la sous fonctionnalisation des gènes *hoxb5a* et *hoxb5b*, nous avons comparé les séquences des loci *hoxb5* de l'homme, de la souris, du *Danio*, et de *Takifugu*. Ceci nous a permis d'identifier 4 éléments non-codants (ENC) conservés. Un de ces ENCs, nommé J3, est présent seulement dans le locus *hoxb5a* alors que les trois autres se retrouvent à la fois dans les loci *hoxb5a* et *hoxb5b* des poissons téléostéens. Nous avons examiné l'activité de ces ENCs, de manière individuelle et combinée à l'aide d'animaux transgéniques. Lorsque testés individuellement, les activités intensificatrices des ENCs des gènes *hoxb5a* et *hoxb5b* se chevauchent de manière marquée. Par contre, de grands fragments d'ADN couvrant plusieurs ENCs parviennent à conférer à un gène rapporteur

l'expression spécifiques d'*hoxb5a* et *hoxb5b*. Le fait d'enlever la séquence J3 du locus *hoxb5a* ou de l'insérer dans le locus *hoxb5b* affecte de manière drastique l'activité intensificatrice globale du locus. Nos résultats montrent que l'évolution de la régulation des gènes *hoxb5* dupliqués des poissons téléostéens pourrait bien s'avérer plus complexe que celle prédite à la lumière du modèle DDC.

Contents

List of figures and tables

List of abbreviations

amp: ampicillin resistance gene

Ant: antennapedia

ASH: achaete-acute homolog, basic helix-loop-helix transcription factor

BCIP: 5-bromo-4chloro-3-indolyl phosphate

beta: recombineering protein that binds to single-strand DNA and stimulates annealling of complementary sequences

bp: base pair

BSA: bovine serum albumin

Bx: bithorax

°C: degrees Celsius

CD1: mouse strain

cDNA: complementary deoxyribonucleic acid

Cdx: caudal-type homeobox transcription factor

ChIP: chromatin immunoprecipitation

cm: chloramphenicol resistance gene

CNE: conserved non-coding element

CNS: central nervious system

Copia: a mobile element in the *Drosophila* genome

CpG: DNA site where cytosine (C) is connected to guanine (G) via phosphodiester (p) bond

cre: bacteriophage P1- derived recombinase

DDC: duplication–degeneration-complementation

DIG: digoxigenin

DMSO: dimethyl sulfoxide

DNA: deoxyribonucleic acid

Domain D: fragment of mouse *Hoxb5/Hoxb5* intergenic region encompassing the J2 CNE

Domain E: fragment of mouse *Hoxb4/Hoxb5* intergenic region encompassing the J3 CNE

Domain LPM/PV: fragment encompassing 5' flanking region of mouse *Hoxb5* and the J1 CNE

dpf: days post fertilization

Dr: *Danio rerio*, zebrafish

DRG: dorsal root ganglia

DY380: *E.coli* bacteria containing *gam*, *beta* and *exo* controlled by a temperature-sensitive repressor

E: embryonic day

EDTA: ethylenediamine tetraacetic acid

egfp: enhanced green fluorescent protein

EL350: *E.coli* bacteria containing *gam*, *beta* and *exo* controlled by the temperature-sensitive repressor and a *cre* gene under the control of an arabinose-inducible promoter

ET cloning: a cloning technology that relies on recombination between identical DNA sequences mediated by exonuclease (RecE/exo) and anealling proteins (RecE/beta)

Exd: extradenticle, a homeodomain transcription factor in *Drosophila*, a cofactor for Hom proteins

exo: exonuclease that acts on linear double-strand DNA to generate single-strand DNA

Fgf: fibroblast growth factor

Ficoll-400: polysucrose 400

FSH: follicle-stimulating hormone

gam: a recombineering protein that enhances recombination and inhibits degradation of linear DNA

GCR, global control region

Gypsy: mobile element in the *Drosophila* genome

Hom: homeotic

Hox: homeobox

*hoxb5a*PCR2.1Δ: a construct encompassing a large (18.495 bp) fragment from the *hoxb5a* zebrafish locus

*hoxb5b*PCR2.1Δ: a construct encompassing a large (12.169 bp) fragment from the *hoxb5b* zebrafish locus

hoxb5aegfp/lacZ: a reporter construct carrying the *hoxb5a* locus and the *egfp/lacZ* reporter gene inserted in frame with the coding sequence of *hoxb5a*

hoxb5begfp/lacZ: a reporter construct carrying the *hoxb5b* locus and the *egfp/lacZ* reporter gene inserted in frame with the coding sequence of *hoxb5b*

*hoxb5a*ΔJ3*egfp/lacZ:* reporter construct carrying the *hoxb5a* locus from which J3 was deleted and the *egfp/lacZ* reporter gene inserted in frame with the coding sequence of *hoxb5a*

*hoxb5b*insJ3*egfp/lacZ*: reporter construct carrying the *hoxb5b* locus in which J3 was inserted and the *egfp/lacZ* reporter gene inserted in frame with the coding sequence of *hoxb5b*

hpf: hours post fertilization

Hs: *Homo sapiens,* human

HTH: homothorax, homeodomain transcription factor, cofactor for Hox proteins

J1: *Hoxb5* enhancer residing in the 3'-flanking region of the *Hoxb5* gene

J2: *Hoxb5* enhancer residing in the 3'-flanking region of the *Hoxb5* gene

J3: regulatory element residing in the 3'-flanking region of the *Hoxb5* gene

kan: kanamycin resistance gene

kb: kilobase

lacZ: β-galactosidase

LH: luteinizing hormone

*lox*P*:* 34-base-pair recombination sites recognized by cre recombinase

Meis: homeodomain transcription factor, cofactor for Hox proteins

min: minutes

miRNA: microRNA

mir10a: miRNA gene residing in close proximity to mouse *Hoxb8*

mir196a: miRNA gene residing in close proximity to mouse *Hoxb4*

Mitf: microphthalmia-associated transcription factor

Mll: mammalian homologue of *Trx*

Mm: *Mus musculus,* mouse

N-: amino

NBT: 4-nitroblue tetrazolium chloride

NCBI: National Center of Biotechnology Information

NP-40: Nonidet-40, nonionic surfactant

Nudt*:* diphosphoinositol polyphosphate phosphohydrolase

P: a mobile element in the *Drosophila* genome

PAC: P1-derived artificial chromosome

PBS: phosphate buffered solution

PBST: phosphate buffered solution and Tween 20

Pbx: vertebrate homolog of *Exd*

PcG: polycomb group

PG: paralogous group

PCR: polymerase chain reaction

PFA: paraformaldehyde

Prep: prolyl endopeptidase enzyme, cofactor for Hox proteins

R: round of *Hox* gene duplication

RA: retinoic acid

RARE: retinoic acid response element

RNA: ribonucleic acid

rpm: rotations per minutes

RT-PCR: reverse transcriptase polymerase chain reaction

SDS: sodium dodecyl sulphate

Sox9: SRY (sex determining region Y)-box 9 gene

SSC: saline sodium citrate

TBS: tissue freezing medium

TE: Tris and EDTA

Tr: *Takifugu rubripes*, Fugu pufferfish

Trx: Trithorax, *Drosophila* transcription factor

UTR: untranslated region

Wnt: family of highly conserved secreted signaling molecules regulating cell-to-cell interactions during embryogenesis

WT: wild type

X-gal: 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

Acknowledgements

I owe a great deal of gratitude to many people whose support and encouragement made this thesis a reality. Foremost, I would like to thank my scientific advisor, Dr. Marc Ekker, for believing in this work, providing a large amount of technical resources and challenging me to push beyond what I thought my limits were. Thank you for pushing me! I would also like to thank my committee members, Dr. David Picketts, Dr. Kursad Turksen and Dr. Marie-Andree Akimenko for providing me with sound advice and encouragement throughout my graduate studies. I want to express my sincere gratitude to the National Sciences and Engineering Research Council of Canada and the Ministry of Education for awarding me with NSERC and OGS postgraduate scholarships that made this journey a lot easier and much more fulfilling for me.

I offer thanks to many former and present members of Dr. Ekker and Dr. Akimenko laboratories for their assistance and encouragement. I am especially grateful to Dr. Fabien Avaron for coping with my occasional scientific despairs and being abundantly helpful in finding simple solutions to technical problems. I have always been amazed with his ability to come up with life saving scientific ideas after a couple beers. I am grateful to Dr. Poitras for his useful advice on construct design, quantitative RT-PCR experiments and all of the time and effort he put into making the homologous recombination technology work in Marc's laboratory.

I would like to take this opportunity to acknowledge the critical role of Adrianna Gambarotta in generation of transgenic mice. Her dedicated care and willingness to help made the most difficult tasks much easier. Thanks go to Gary Hatch and Lucille Joly, talented research assistants, for their help in generation of several reporter constructs. To

Magdalena Grzyb, Marion Alfero, former undergraduate students and Vishal Saxena, a research assistant in the laboratory for their help with injections of *egfp* reporter constructs carrying large regions of the *hoxb5* loci and generation of transgenic zebrafish. I thank Amanda Smith, a PhD candidate and Danielle Guay, a research assistant in Dr. Akimenko's laboratory for providing a welcoming working envorinment, for good company, good advice and lots of good laughes. Thanks are also due to Ryan Macdonald, Gary Hatch and Nick Shipley for proofreading my thesis.

I am grateful to Donna Hooper, an academic assistant and Sylvie Deblois, an administrative secretary at the Faculty of Medicine for their kind assistance with a number of administrative issues I had during those arduous years of a Ph.D. It is with their valuable help I was able to nudge through the academic bureaucracy.

Outside of work, I would like to thank Elisa Rolon and Samir El Maach for all the emotional support, carrying and sound help they provided.

Lastly, and most importantly, I would like to thank my father, Yrii Jarinov, my mother, Julia Jarinova, and my brother, Sergey Jarinov, for their support and active interest in my career aspirations.

Thank you!

Olga Jarinova.

1. Introduction

Homeotic genes were identified by Lewis, Nüsslein-Volhard, and Wieschaus in 1975 following observation of homeotic transformations – the conversion of one body structure into the likeness of another - in *Drosophila melanogaster* (Gehring and Hiromi 1986; Scott et al. 1989). Genes responsible for these mutations shared a nearly identical motif known as an homeobox which codes for a DNA-binding homeodomain (Lewis 1978). Following studies confirmed that the homeodomain is a common feature of the homeotic genes which cluster together in a genome.

Two clusters of homeotic genes were identified in the *Drosophila* genome and termed *Antennapedia* and *Bithorax.* The *Antennapedia* cluster (*Ant-C*) comprises five genes responsible for specification of anterior thorax identity. The *Bithorax* complex (*Bx-C*) cluster harbors three genes that control the identity of posterior thorax and abdominal body segments of *Drosophila.* Combined, *Ant-C* and *Bx-C* homeotic clusters in *Drosophila* are often referred to as the *Hom-C* cluster (Lappin et al. 2006).

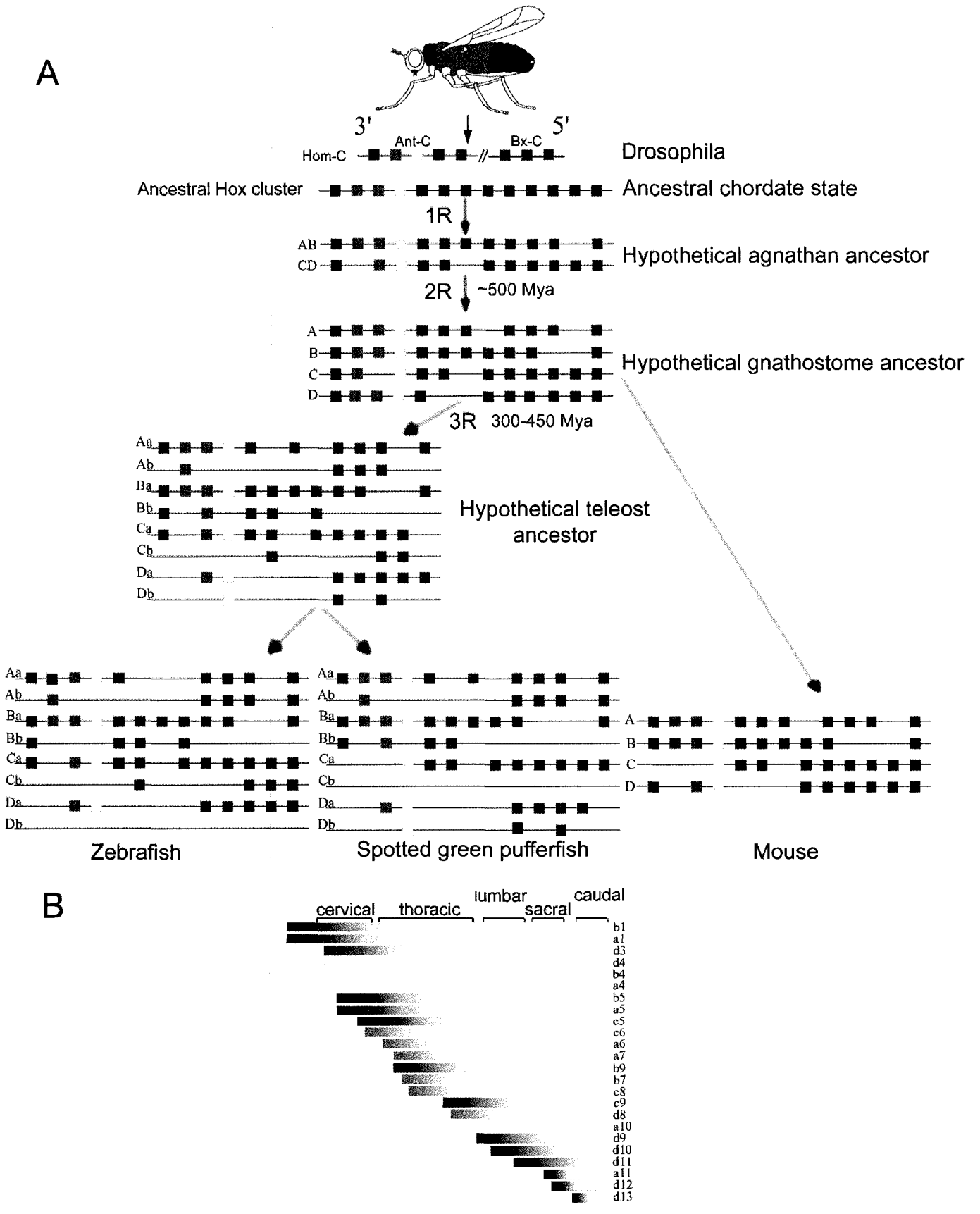
To date, at least 39 genes functionally and structurally homologous to *Drosophila* homeotic genes have been identified in vertebrates (reviewed in: Krumlauf 1994; Akin and Nazarali 2005). They are referred to as *Hox* genes and their function is to assign specific identity to individual segments along the anterior-posterior axis of the body (reviewed in: McGinnis and Krumlauf 1992; Favier and Dolle 1997; Prince 2002. *Hox* genes are organized in multiple clusters that are thought to have resulted from the sequential duplications of a proto*Hox* cluster early in metazoan evolution (Patel and Prince 2000). Based on the criteria of sequence conservation, *Hox* genes have been divided into several paralogous groups - closely related groups of genes on different

Figure 1.1 The *Hox* gene family underwent significant expansion during evolution; yet their function in axial patterning remained unchanged.

A. This simplified scheme reflects our current view on *Hox* gene evolution. Homologous genes of a common origin are presented as squares of the same color. The phylogenetic timing of *Hox* cluster duplications (1R, 2R and 3R), gene losses and appearances of pseudogenes (black squares) are shown.

B. Co-linear distribution of *Hox* gene transcripts along the AP axis in the mouse mesoderm suggests conservation of *Hox* functions throughout evolution. The expression gradient of each *Hox* gene along the anterior-posterior axis in the mouse mesoderm is shown by rectangles of the corresponding colours. Darker colours reflect abundance of *Hox* gene transcripts at the anterior border of expression. Expression usually fades off towards the posterior end of the embryo.

This figure has been adapted from (Lappin et al. 2006) and (Hoegg and Meyer 2005).


A
3'
5'
Ant-C
Bx-C
Hom-C
Drosophila
Ancestral Hox cluster
Ancestral chordate state
1R
AB
CD
Hypothetical agnathan ancestor
2R
~500 Mya
A
B
C
D
Hypothetical gnathostome ancestor
3R
300-450 Mya
Aa
Ab
Ba
Bb
Ca
Cb
Da
Db
Hypothetical teleost ancestor
Zebrafish
Spotted green pufferfish
Mouse
B
cervical
thoracic
lumbar
sacral
caudal
b1
a1
d3
d4
b4
a4
b5
a5
c5
c6
a6
a7
b9
b7
c8
c9
d8
a10
d9
d10
d11
a11
d12
d13

clusters. Nine such groups (numbered 1 through 9) are thought to be presented in the ancestral cluster while four additional paralogous groups have been identified in vertebrate *Hox* clusters. There is a high degree of sequence similarity between homeotic genes from the *Hom-C* cluster in *Drosophila* and genes residing in multiple *Hox* clusters in vertebrates. *Hox* paralogs from PG 1-8 are closely related to *Ant-C* genes and genes located at the 5' end of the cluster (PG 9-13) are more closely related to *Bx-C* (Figure 1.1).

Striking structural and functional conservation of *Hox* gene family members have been shown for many metazoan species. This likely reflects their indispensable ability to confer positional information to the axial and paraxial tissues during embryogenesis and hence ensures the correct patterning of the embryo (Deschamps and van Nes 2005). A unique combination of *Hox* gene expression in a specific anterior-posterior region of an embryo is referred to as "the *Hox* code". Although morphologically identical, regions with differential combinatorial *Hox* expression will give rise to distinct mesodermal tissues as specified by differential "*Hox* code" established in each segment (Figure 1.1).

1.1. The Hox transcription factors

1.1.1. Functional evidence for *Hox* complexes being truly homeotic clusters

To clearly define function of *Hox* genes during embryogenesis, one should examine phenotypes of *Hox* mutations that abolish gene expression and/or analyze consequences of expression of individual *Hox* genes in abnormal locations.

The function of homeotic genes in *Drosophila melanogaster* has been inferred from the observation of phenotypes of several mutations, of which *Antennopedia* and *Bithorax* are the most striking. In *antennopedia* mutants the antennae are changed into legs while *bithorax* mutation causes transformation of the haltere (balancing organ) into a part of the wing (Lewis 1978; Figure 1.2, panels: A, B). These and other observations helped determine that homeotic genes in *Drosophila* act as master switches activating the genetic program that leads to the correct patterning of each segment of the body along the anterior-posterior axis.

Similarly to that in *Drosophila*, loss of *Hox* gene function in vertebrates leads to changes in regional identity of mesodermal, ectodermal and endodermal derivatives (reviewed in Krumlauf, 1994). For example, knock-out of mouse *Hoxc8* that is normally expressed in the thoracic and more posterior regions of the developing mouse embryo, caused development of an extra pair of ribs on the first lumbar vertebrae (Le Mouellic et al. 1992; Figure 1.2C). In addition, *Hoxd13* null-mutant mice have been reported to display anterior transformation of the axial skeleton (Dolle et al. 1993).

However, it has to be emphasized that, in general, disruption of individual *Hox* genes in vertebrates results in much less severe consequences than can be anticipated from their expression patterns. As a result, phenotypes of *Hox* null mutants are much less predictable than that in *Drosophila* (Krumlauf 1994; Wolpert 1998). For example, disruption of the *Hoxa3* gene in the mouse causes significant structural defects in the head and thoracic region of the body where this gene is normally expressed. However, no abnormalities have been detected in posterior axial structures where *Hoxa3* expression has also been detected (Chisaka and Capecchi 1991). Similarly, although *Hoxa1* null

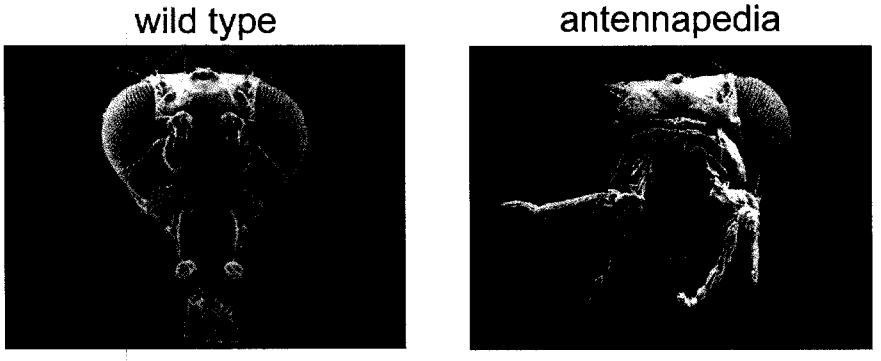
Figure 1.2. Loss of function of individual *Hox* genes results in homeotic transformations in mouse and Drosophila.

A. *antennapedia* mutation in *Drosophila melanogaster* causes transformation of the antennae into legs. Adapted from (Turner and Mahowald 1979).

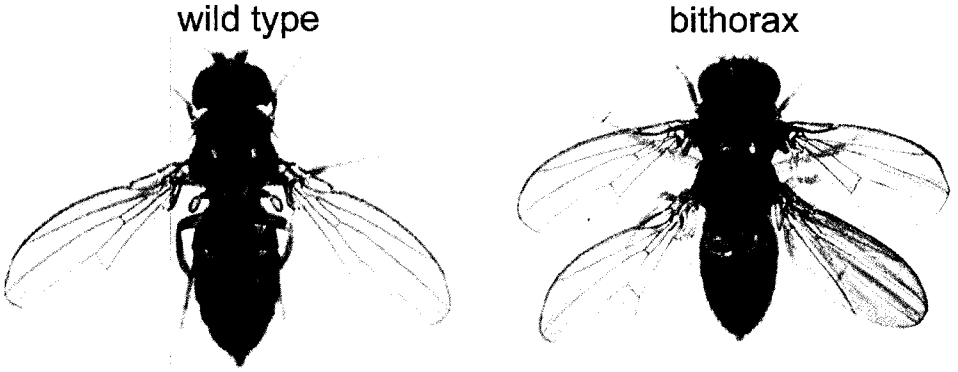
B. In *bithorax* mutant flies an additional pair of wings developed at the place of halteres. Adapted from (Lewis 1978).

C. *Hoxc8* null mice develop rib-bearing thoracic vertebra at the place of the fist lumbar vertebra. Adapted from (Le Mouellic et al. 1992) and (Gilbert 2000).

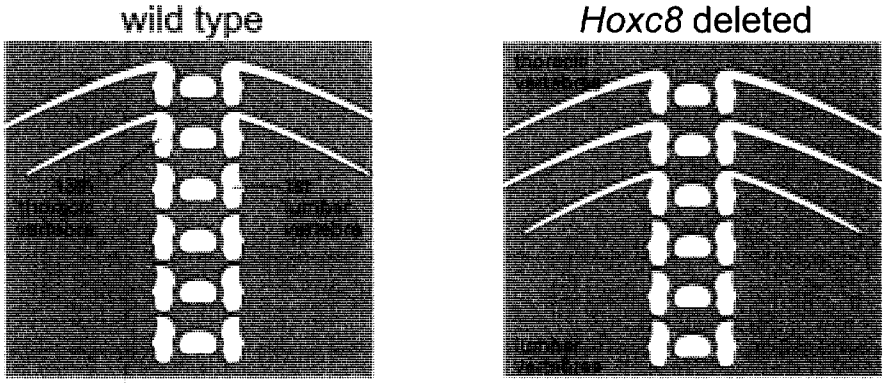
A


wild type
antennapedia

B


wild type
bithorax

C


wild type
Hoxc8 deleted

mutants exhibited cranial and mesenchemal neural crest defects, no homeotic transformations were observed in these mice (reviewed in Krumlauf 1994). This is likely due to the fact that paralogous *Hox* genes in vertebrates (genes from the same paralogous group located in different clusters) often share overlapping expression domains (Krumlauf 1994). Moreover, there are instances when neighboring genes, residing in the same *Hox* cluster, share a common anterior border of expression (Prince et al. 1998). These data suggest that a single set of *Hox* genes may be sufficient to establish the embryonic body plan and imply possible functional redundancy between paralogous *Hox* genes residing in different *Hox* clusters. Indeed, mouse embryos lacking all nine of the *Hoxc* genes were able to survive until the time of birth with only minor developmental transformations (Suemori and Noguchi 2000). In contrast, mice deficient for more than one *Hox* co-paralog showed severe developmental abnormalities. Hence mice in which all paralogous group 8 *Hox* genes were simultaneously inactivated exhibit more severe phenotypes with more extensive homeotic transformations than found in a single and double mutants for these genes (van den Akker et al. 2001).

Taken together, these findings suggest that the vertebrate "*Hox* code" is more complex than that in *Drosophila* owning to synergistic relationship of co-paralogs in regional patterning (Krumlauf 1994).

1.1.2. Molecular characteristics of *Hox* genes

Typically, vertebrate *Hox* genes comprise two exons separated by an intron of variable length (Castronovo et al. 1994). The highly conserved homeobox sequence that encodes the homeodomain usually resides in the second exon of the gene (Figure 1.3A). Although

the vast majority of characterized *Hox* genes share this simple gene structure, a recent study by Hadrys et al indicates that there may be exceptions to that rule. Hence, *hoxb4a* and *hoxb3a* zebrafish genes were reported to possess three and six exons respectively (Hadrys et al. 2006).

The helix-turn-helix tertiary structure of the homeodomain plays an important role in sequence-specific interaction (Wright et al. 1989). These homeotic proteins trigger transcriptional activation of their target genes in a segment-restricted manner (Hoey and Levine 1988; Hoey et al. 1988). Traditionally, amino acid residues congregating to assemble the homeodomain structure are numbered 1-60. The first 10 residues form the N-terminal arm of the homeodomain that falls into the minor DNA grove. Three helical regions follow and occupy the 10-22, 28-37 and 42-58 positions (Lewin 2000). When bound to a target sequence, the third helix contacts the major groove of DNA while the first two helical regions lie above the DNA (Figure 1.3B; Lewin 2000). The homeodomain structure is estimated to be over a billion years old and is found in fungi, plants and animals (Scott 1992).

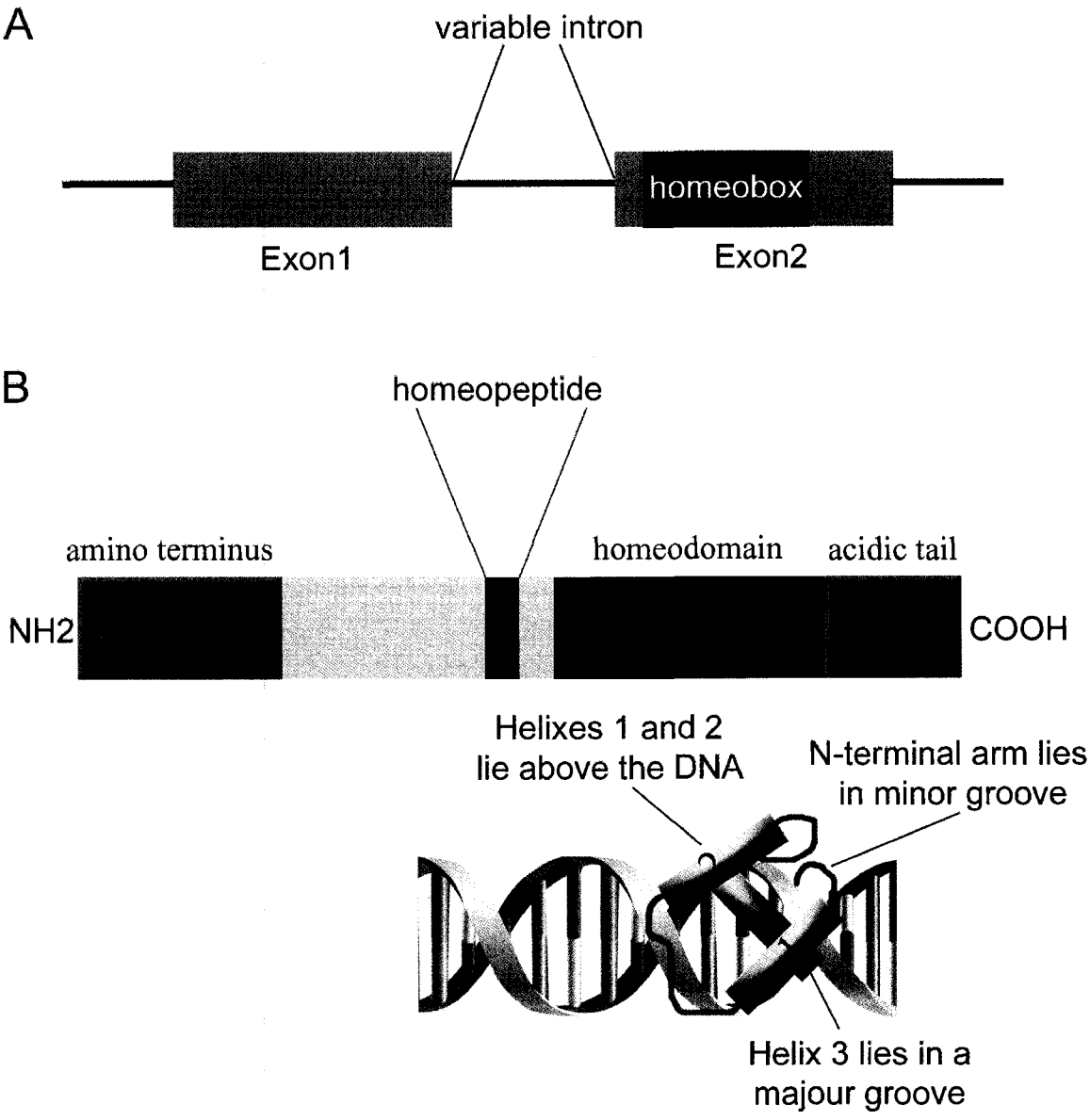
Besides homeodomains, there are other conserved regions in homeotic proteins. These include an N-terminal region that in some proteins extend up to 45 amino acid residues and a pentapeptide motif (IYPWM) located just upstream of the homeodomain (Wright et al. 1989).

Figure 1.3. Molecular characteristics of *Hox* genes.

A. A typical *Hox* gene consists of one intron and two exons. The homeobox resides in the second exon of the gene.

B. Protein structure of a homeodomain protein. The homeodomain consists of three helices and has a helix-turn-helix tertiary structure. One helix binds to a target sequence, while the other two lie above the DNA. Other conserved protein domains include amino terminus, acidic tail and pentapeptide.

This figure has been adapted from (Castronovo et al. 1994) and (Lewin 2000).


A
variable intron
homeobox
Exon1
Exon2
B
homeopeptide
amino terminus
homeodomain
acidic tail
NH2
COOH
Helixes 1 and 2
lie above the DNA
N-terminal arm lies
in minor groove
Helix 3 lies in a
majour groove

1.1.3. Cofactors of Hox proteins increase specificity to their downstream targets

As mentioned above, members of the *Hox* gene family share a conserved DNA binding domain termed the homeodomain that mediates interactions of Hox proteins with the regulatory elements of their target genes. Previous studies demonstrated that different Hox proteins bind to highly similar DNA sites *in vitro* (Catron et al. 1993; Pellerin et al. 1994). In 1994, Pellerin et al. reported that five members of the *Hox* family: *Hoxa5*, *Hoxb4*, *Hoxa7*, *Hoxc8*, and *Hoxb1* were able to bind a common consensus DNA site {(C/G)TAATTG} and activate transcription of a target gene as determined in transfection assays (Pellerin et al. 1994). At first glance, such poor target selectivity may seem puzzling as Hox proteins are implicated in imposing the "*Hox* code" during embryogenesis and thus are anticipated to act differentially on a given target. However, it has to be considered that, despite the fact that different Hox proteins are able to interact with the same consensus site, they exhibit subtle preferences for distinct nucleotide variations within the consensus motif (Pellerin et al. 1994). Furthermore, Hox proteins differ in their relative affinity for a particular DNA site. This property is further influenced by cofactors known to provide additional specificity during target gene selection (reviewed in Akin and Nazarali 2005). Interestingly, many of the cofactors that function in multiprotein complexes with *Hox* genes are also homeodomain proteins. Thus *Exd/Pbx* family members have been shown to alter activity of Hox proteins *in vitro* and to increase specificity of homeodomain proteins for their downstream targets *in vivo* (Rauskolb and Wieschaus 1994). In turn, Meis, Prep and HTH proteins have been shown to directly bind to Exd/Pdx proteins *in vitro* and are thought to control Exd/Pdx

availability to interact with Hox proteins in the nucleus (Rieckhof et al. 1997; Mann and Affolter 1998). In addition to the cofactors' ability to increase Hox protein specificity for a particular recognition site, they may also regulate Hox activity by inducing conformational changes in certain protein domains (Mann and Affolter 1998).

1.1.4. Fundamental role of *Hox* genes in early vertebrate development

Hox genes are master developmental genes that reside within the genetic hierarchy and regulate crucial developmental processes such as morpho- and embryogenesis (reviewed in Krumlauf 1994). They are sequentially activated in time and space according to their position on a chromosome: "anterior" genes on the 3' end of *Hox* clusters are expressed first providing embryonic segments with more rostral identity whereas "posterior" 5' *Hox* genes are expressed later in embryogenesis endowing a specific segment with more caudal positional information (van der Hoeven et al. 1996; Meyer and Malaga-Trillo 1999). This phenomenon is known as "spaciotemporal co-linearity" and is found in all metozoans examined thus far suggesting the conservation of developmental mechanisms.

Combinatorial expression of *Hox* genes establish intersegmental boundaries and facilitate the formation of local signaling centers which function to inform adjacent cell lineages of their position and fate (Kiecker and Lumsden 2005). Indeed, analysis of *Hox* genes in the chick and mouse suggests association of individual *Hox* genes with specific anatomical boundaries (Burke et al. 1995). Later in development, cell lineages arising from individual functional modules will develop structures in accordance with their position along the body axis.

The role of *Hox* genes in somitogenesis has been best studied in mouse (reviewed in Krumlauf 1994). As in all vertebrates, the very first wave of *Hox* gene expression is detected during gastrulation where they confer positional information to the axial and paraxial tissues emerging from the posterior part of an embryo (Mark et al. 1997). In mouse, *Hox* gene expression is first detected in the primitive streak and later spreads through the node region (Deschamps 2005).

Typically, *Hox* genes have a sharp anterior border of expression while their posterior border is not well defined (Fienberg et al. 1987; Figure 1.1B). Despite the overlap in expression, anterior segments of the developing embryo possess a unique set of expressed *Hox* genes. For example, only *Hoxb1* and *Hoxa1* are expressed in the most anterior somites of the mouse embryo, while anterior borders of expression for *Hoxa11* and *Hoxd12* fall into the sacral region (most posterior part of the caudal body axis) (Small and Potter 1993; Herault et al. 1998; Studer et al. 1998). In addition to spatial coordination of the *Hox* genes, there is a correlation between embryonic stage and time when expression of individual *Hox* genes first appears. For example, the most anterior genes such as *Hoxa1* and *Hoxb1* are expressed in the posterior head mesoderm early in mouse development, while most posterior genes such as *Hoxd12* are expressed in the anal tail which develops much later (Small and Potter 1993; Herault et al. 1998; Studer et al. 1998). Through overexpression of *Hoxb* genes in chick embryo, Imura and Pourque recently demonstrated that cells expressing more anterior *Hox* genes contribute to anterior portions of somatic columns; whereas cells expressing more posterior *Hox* genes - to more caudal somatic domains (Iimura and Pourquie 2006). Hence the colinear

spaciotemporal distribution of *Hox* genes predetermines the fate of presomic mesoderm prior to somite formation.

Somite formation occurs in the mesodermal region anterior to the regressing Hensen's node with each somite pair appearing concurrently on each side of the body. In all vertebrate embryos, somite formation proceeds rhythmically in the anterior-posterior direction. Later in development, the somites will undergo further differentiation. The dorsal and lateral cells of newly formed somites will form myotome and dermatome; the ventral part of the somite will give rise to sclerotome cells (Dockter and Ordahl 1998). These cells will then develop into vertebrae while myotome and dermatome will form muscle and dermis cells respectively (Kato and Aoyama 1998; Ordahl and Williams 1998; Dockter and Ordahl 2000). Unlike simultaneous segmentation seen in the *Drosophila* embryo, vertebrates form segments sequentially in a process synchronized with the posterior extension of the embryo (Palmeirim 1999; Andrade et al. 2007).

The role of *Hox* genes in embryogenesis is not confined to the patterning of paraxial mesoderm. Coordinated patterns of spaciotemporal *Hox* expression are also known to initiate patterning of the vertebrate hindbrain by supplying anterior-posterior identity to the neuroectoderm (Kiecker and Lumsden 2005). Borders of *Hox* gene expression correlate with hindbrain segmentation pattern: 3' *Hox* genes begin to express earlier and in more rostral positions of the hindbrain (rhombomeres 1-4), while expression of the *Hox* genes at the 5' end of the cluster is detected more caudally (rhombomeres 5-8) and later in development.

In addition to patterning of the hindbrain and presomitic mesoderm, *Hox* genes play an important role in structural development of many organs. For example, at least 23 *Hox*

genes are involved in limb development of the chick (Nelson et al. 1996). In addition, high expression levels of 3' genes from the *Hoxa* and *Hoxb* clusters have been detected in fetal human and rodent lungs (Lappin et al. 2006). As a general rule, *Hox* gene expression levels significantly decrease once organ morphogenesis is complete. However, some *Hox* genes continue to be expressed throughout organ development such as *Hoxa5* in murine lungs (Kim and Nielsen 2000). CpG methylation of non-coding regulatory regions of these genes has been associated with restricted patterns of *Hox* expression in postnatal tissues (Hershko et al. 2003). Methylation of CpG sites enforces gene repression by either recruitment of histone deacetylase complexes or by preventing binding of transcription factors to *cis*-regulatory regions via direct interference (Razin and Cedar 1991; Levine et al. 1992).

1.2. Mechanisms orchestrating positional and temporal co-linearity of *Hox* gene expression

1.2.1. Initiation and establishment phases of *Hox* gene expression

Clustered organization is fundamental for the precise spaciotemporal regulation and the function of each *Hox* gene (van der Hoeven et al. 1996) (Gould 1997); (Mann 1997) (Kondo et al. 1998). Although the mechanisms effectuating the 3' to 5' colinear expression of the *Hox* genes remain under investigation, recent findings have lifted a corner of the veil.

Coordinated spaciotemporal expression of *Hox* transcription factors likely relies on molecular polarity of the *Hox* clusters that directs initiation of transcription at the 3' end

and its subsequent progression towards the 5' end of the complex. It has been suggested that such progression occurs by means of changes in chromatin structure of corresponding genes and is linked to the sequential release of corresponding genes from silencing factors-associated repression (Duboule and Deschamps 2004).

A functional gradient of endogenous retinoids along the anterior-posterior axis is known to impose differential spacio-temporal activation of individual *Hox* genes. "Anterior" *Hox* genes located at the 3' end of the cluster appear to be more susceptible to retinoid acid (RA) treatment than "posterior" *Hox* genes in cell lines and embryonic tissues (Papalopulu et al. 1991; Prince 1998). Retinoic acid mediates its effects via RA receptors that in turn bind to RAREs (Retinoid Acid Response Elements) abundantly present throughout the *Hox* clusters (Marshall et al. 1994).

A recent study devoted to regulation of the *Hoxb* cluster in mice has clearly demonstrated that RA treatment results in approximately nine-fold chromatin de-condensation and is accompanied by histone acetylation (Chambeyron and Bickmore 2004). However, the coordinated chromatin de-condensation, although believed to establish a transcriptionaly competent state, is not sufficient to provide temporal coordination of *Hox* gene expression (Chambeyron and Bickmore 2004). Chambeyron and Bickmore proposed that progressive looping of the corresponding region away from chromosomal territory and towards the center of the nucleus following the initial de-condensation event. "Looping out" coincides with high rates of individual gene expression and thus represent an important mechanism for providing temporal sequential accessibility of *Hox* genes (Chambeyron and Bickmore 2004). These experiments reveal that RA is an endogenous morphogen that plays a physiological role in specifying anterior-posterior patterning. The colinear response of genes to RA is suggested to be an

important property of the *Hox* clusters necessary for the establishment of coordinatively restricted gene patterns during embryogenesis (Papalopulu et al. 1991).

Recent findings demonstrated that RA signaling is not only essential for establishment of expression domains, but also controls the late phases of *Hox* gene expression. For example, the expression domains of the 5' genes from mouse *Hoxb* (*Hoxb5-Hoxb8*) do not reach their definitive rostral boundaries until rhombomeric segmentation is complete while domains of their 3' counterparts extend to their anterior boundaries before hindbrain segmentation (Oosterveen et al. 2003). The rostral extension of 5' *Hoxb* genes in the Central Nervous System (CNS) occurs later in development, depends on RA signalling, and is mediated by a functional RARE element located 3'upstream of *Hoxb5* (Oosterveen et al. 2003).

Fgf and Wnt signaling pathways, that are known to contribute to formation and expansion of the primitive streak, also play a pivotal role in *Hox* gene regulation during early embryogenesis (Ciruna and Rossant 2001; Forlani et al. 2003). Another group of homeodomain proteins, *Cdx*, is known to modulate *Hox* gene expression in a dose-specific way (reviewed in Deschamps and van Nes 2005). *Cdx* may serve to convey positional information from RA, Fgf and Wnt signaling to the *Hox* genes in vertebrates (Lohnes 2003).

Finally, regulatory DNA elements residing in non-coding regions of the *Hox* complexes contribute to the establishment of nested *Hox* expression patterns. *Cis*-acting regulatory elements dispersed throughout *Hox* clusters control tissue specific expression of neighboring genes. Spitz and colleagues have recently reported the existence of a global control region (GCR) located over 240 kb upstream of *Hoxd* that coordinatively regulates expression of 5' *Hoxd* genes in the limb bud (Spitz et al. 2003). It is possible

that similar global control elements are involved in transcriptional regulation of all clusters of the *Hox* genes enabling harmonious activation of the corresponding genes (Deschamps and van Nes 2005).

1.2.2. Trx and Polycomb regulate maintenance of *Hox* gene expression

Recent studies in flies and mammals suggest that Polycomb and Trithorax proteins are also involved in regulation of the spatially restricted active and repressed states of *Hox* expression (reviewed in Schwartz and Pirrotta 2007). Polycomb group proteins (PcG) were first discovered in the fruit fly where they act to repress inappropriate spacial expression of the *bithorax* genes (Busturia and Morata 1988). In addition to their role in regulation of *Hox* expression, PcG have been reported to be involved in mammalian X-inactivation and genomic imprinting (Delaval and Feil 2004; Heard and Disteche 2006). Their function as global enforcers of the transcriptionally repressed state is widely conserved from plants through to mammals (Pirrotta 1997). Polycomb proteins are known to operate in association with other proteins: at least three kinds of multiprotein PcG complexes referred to as PRC1, PRC2 and PhoRC, have been identified to date (Schwartz and Pirrotta 2007). Purified and reconstituted PCR1 complexes from both flies and humans were shown to inhibit chromatin de-condensation and repress transcription *in vitro*, likely acting at the level of primary chromatin fibre (reviewed in Schwartz and Pirrotta 2007). Recent chip-on-ChIP studies performed on *Drosophila* and mammalian tissues have revealed a large number of genes, which includes but are not limited to morphogens, receptors and signaling proteins, as PcG targets. PcG repression mechanisms repress most differentiation and developmental pathways with the exception

of the ones required for a particular stage at any given point (Schwartz and Pirrotta 2007).

Interestingly, all the identified targets of PcG silencing are positively regulated by Trithorax (Trx), ASH1 and ASH2 proteins. These proteins set the mark for an active transcriptional state in segments and are referred to as Polycomb antagonists (Schwartz and Pirrotta 2007). *Trx* and its mammalian homologues, *Mll*, are associated with promoters of transcriptionally active genes and are crucial for embryonic development (Bajusz et al. 2001). In *Mll* null mutant mice, *Hox* gene expression has been reported to be correctly initiated but was not sustained (Terranova et al. 2006). Hence, *Trx* products are required to maintain *Hox* gene expression in the appropriate segments of the body while Polycomb proteins are involved in its repression (Pirrotta 1997). Although *Trx* and *PcG* genes play a pivotal role in maintaining of proper expression levels of *Hox* genes during development, they are likely not to be required at the initiation stage of *Hox* transcription.

1.2.3. Posttranscriptional regulation of *Hox* genes

In situ hybridization and immunostaining experiments demonstrate that *Hox* gene expression in vertebrates are at the highest levels at its anterior border and gradually diminish towards the posterior end of an embryo (Prince et al. 1998; Brend et al. 2003), (Figure 1.1B). However, most *Hox* reporter constructs encompassing different combination of enhancer sequences fail to recapitulate the gradual regression of *Hox* expression levels (Sharpe et al. 1998; Brend et al. 2003). This suggests that elements

other than enhancers are present in endogenous *Hox* environments and are involved in late stages of *Hox* regulation.

Brend and colleagues demonstrated that inappropriate expression of a mouse *Hoxb4* transgene, in the posterior embryonic regions, is the result of ectopic transcription that is normally suppressed by mechanisms of posttranscriptional regulation. They found that the 3' untranslated region of the *Hoxb4* gene is able to confer transcript instability to the transgene transcripts (Brend et al. 2003).

Furthermore, recent findings suggest that the presence of microRNA (miRNA) genes in multiple locations inside the *Hox* clusters is not coincidental. miRNAs are a class of short non-coding RNAs that are involved in the posttranscriptional regulation of gene expression by inhibiting translation or facilitating degradation of their target mRNA (Hartmann et al. 2004). Using the transgenic sensor approach, Mansfield et al. have shown that two miRNA genes, *mir-196a* and *mir10a*, residing in close proximity to mouse *Hoxb4* and *Hoxb8* respectively, are expressed in *Hox*-like patterns. Interestingly, *mir-196* genes are found in corresponding locations of *Hoxc8* and *Hoxd8* mouse loci suggesting commonalities in mechanisms posttranscriptional regulation for all the genes from paralogous group 8 (Mansfield et al. 2004).

Taken together, these data suggest that microRNA genes and regulatory elements residing in untranslated regions of the genes might fine tune colinear *Hox* gene expression. This is done by limiting maximum extension of *Hox* expression domains along anterior-posterior axis of the body (Deschamps and van Nes 2005).

The described above aspects of *Hox* regulation contribute to achieving perfectly orchestrated spacio-temporal expression of Hox proteins that ensures harmonious embryogenesis.

1.3 Duplication as a source of evolutionary novelty

1.3.1. Expansion of the *Hox* gene family during evolution correlates with increased complexity of body plans and teleost radiation

The *Hox* gene family underwent significant expansion during evolution (Pendleton et al. 1993; Meyer and Malaga-Trillo 1999). It is believed that the amplification of *Hox* genes proceeded in two steps which encompass events of both *cis*- and *trans*- duplication (Meyer and Malaga-Trillo 1999). Ancestral *Hox* gene first underwent tandem duplications predating the divergence of arthropods and chordates which produced a cluster of at least 13 genes (Kappen and Ruddle 1993). The following *trans*-duplications of *Hox* clusters (1R, 2R and 3R) occurred through either a duplication of an individual chromosome or as a result of whole genome duplications during the transition from cephalochordates to vertebrates (Meyer and Malaga-Trillo 1999). Consistent with the proposed "two steps" model of *Hox* genes evolution, *trans*-paralogous genes resulting from more recent duplication events have a higher degree of sequence conservation than *cis*-paralogous genes residing in the same cluster. Hence all vertebrates contain several *Hox* clusters that are thought to have resulted from the sequential duplication of a pro*Hox* cluster early in metazoan evolution. Thus, mammals have four clusters of *Hox* genes: *Hoxa*, *Hoxb*, *Hoxc* and *Hoxd* whereas zebrafish possess seven: *hoxaa*, *hoxab*, *hoxba*, *hoxbb*, *hoxca*, *hoxcb* and *hoxd*. The additional clusters in teleost likely resulted from a "recent" duplication event that happened after the divergence of the fish and tetrapod lineages somewhere between 300 and 450 million years ago (Amores et al. 1998; Taylor et al. 2001). The presence of additional *hox* clusters is common, but not ubiquitous, in

teleosts, including those species with a small genome such as *Takifugu rubripes* and *Spheroides nephelus* (Amores et al. 2004; Chiu et al. 2004).

It has been suggested that duplications of *Hox* gene clusters may have facilitated the evolution of the vertebrate body plan (Holland et al. 1994; Ruddle et al. 1994). That is thought because an increase in overall complexity of the body plan correlates with the increasing numbers of *Hox* clusters. Indeed, while invertebrates possess only one *Hox* cluster, higher vertebrates such as mammals and human have four. This suggests that the ancestral *Hox* cluster was duplicated at least twice during the evolution of chordates (Meyer and Malaga-Trillo 1999). Aburomia et al have provided quantitative evidence for a high rate of increase in morphological complexity during early chordate evolution by estimating number of morphological characters present in different classes of vertebrates (Aburomia et al. 2003).

The view of increasing phenotypic complexity in the vertebrate lineage which is tightly associated with *Hox* cluster duplications has long been favored in comparative developmental biology (Meyer and Malaga-Trillo 1999; Wagner et al. 2003). Intriguingly, the third duplication event that produced additional *Hox* clusters in teleost fish did not induce further increases in complexity of the teleost body plan. Instead, it appears to have contributed to adaptive radiation and to an increase in morphological diversity observed in fish lineages which originated after the presumed duplication event (Hoegg et al. 2004). Indeed, Hoegg et al. have estimated that the fish-specific duplication event took place between 335 to 404 million years ago and observed a striking increase in the number of species in the lineages that separated after the duplication (Hoegg et al. 2004). Therefore, the most recent duplication event is thought to enable fish lineages with genetic potential to adapt quickly and flexibly to changing ecological environments

(Meyer and Malaga-Trillo 1999). Elucidating patterns of molecular evolution that allow duplicated genes to acquire a specific function during evolution remains an area of increased interest in the field of comparative developmental biology.

1.3.2. Possible mechanisms of functional diversification of the duplicated genes

Following a duplication event, the evolutionary constraint may be temporarily lifted giving paralogous genes the opportunity to diverge and acquire specific new functions. These changes may occur in coding sequences or regulatory elements, with a single mutation sometimes being sufficient to have profound consequences (Cooper et al. 1995); (Cooper and Youssoufian 1988).

Mutations that lead to the partial or full loss of a subfunction or acquisition of a new one may occur by several mechanisms, including nucleotide substitution, deletion, inversions, insertions of transposable elements etc. (Force et al. 1999). Transposable elements (e.g. transposons and retrotransposons) can alter the original gene expression or even create novel gene function by adding and distributing specific control elements throughout the host genome. Indeed, there are plenty of examples when insertion of mobile elements influences the regulation of gene expression (Kidwell and Lisch 1997; Manna et al. 2004; Sanogo et al. 2007). For example, P, copia and gypsy elements are known to be mutagenic when inserted into the 5' regions of *Drosophila* genes (reviewed in Kidwell and Lisch 1997).

Under the classic model, the most likely fate for the duplicate genes was predicted to be nonfuctionalization (Ohno 1970). Nonfunctionalization is a process in which one

member of the duplicated pair degenerates, accumulating deleterious mutations, while the other gene retains the original function (Li 1980; Watterson 1983). Indeed, this fate is permissible because a single gene copy is sufficient to convey the original function.

This model has also postulated, that in rare cases, one of the duplicated genes may acquire a new function while the other preserves the original gene function (Force et al. 1999). If such change is beneficial, it will be quickly fixed by positive selection. This fate of a gene duplicate is referred to as neofunctionalization.

1.3.3. DDC model

The classic model of evolution of duplicated genes proposed in late 1930s have perceived neo- and nonfunctionalization as the only possible mechanisms for resolution of inter-gene redundancies (Haldane 1933; Fisher 1935). It has fallen short to explain extensive retention of duplicated genes at the whole genome scale which is revealed in comparative studies of several extant genomes (Prince and Pickett 2002). The DDC model (duplication-degeneration-complementation) was proposed by Force and colleagues in 1999 in an attempt to explain the presence of large numbers of duplicated genes that remained preserved after duplication (Force et al. 1999). According to the DDC model, in the absence of positive selection for beneficial mutations, gene duplicates are often preserved in evolution as the result of splitting the original gene function between the duplicates. That process is called subfunctionalization and is predicted to be the most common mechanism of preservation of duplicated genes (Force et al. 1999).

Indeed, the splitting of a single functional unit into separate subunits makes the complete loss of the original function by nucleotide substitutions, deletions etc. an

unlikely event. Moreover, the subfunctionalized duplicates are under relaxed evolutionary constraints compared to the ancestral gene. The DDC model estimates that once the process of subfuctionalization has been initiated, resolution of redundant subfunctions occurs within 4-12.5 million years after the duplication event (Lynch and Force 2000). That grants organisms with silent genetic potential to adaptively respond to different types of selective pressures if they apply before the intergene redundancy is resolved. Therefore, subfunctionalization of the duplicates increases an organism's chances of survival and may be a transitional state leading to neofunctionalization (Lynch and Force 2000).

1.3.4. Possible mechanisms of adaptations of developmental pathways following functional diversification of duplicated genes

Organisms consist of units, or interacting modules, that are coherent within themselves and yet part of a larger unit. Examples of these modules are morphogenetic fields, signaling/developmental pathways, imaginal discs, insect parasegments and vertebrate organ rudiments (Gilbert 2000). This modular structure of organisms allows one part of the embryo to be changed in evolution while other parts are left unchanged. Modifications occurring in one part of the module have to be accommodated by changes in the rest of this module. In other words, only those developmental changes that are followed by a compensatory change in the rest of organism will persist (Nijhout and Emlen 1998; Bull et al. 2007). If a new configuration divergent from the original pathway via mechanisms of sub- or neofunctionalization is beneficial for an organism, it will remain preserved in evolution.

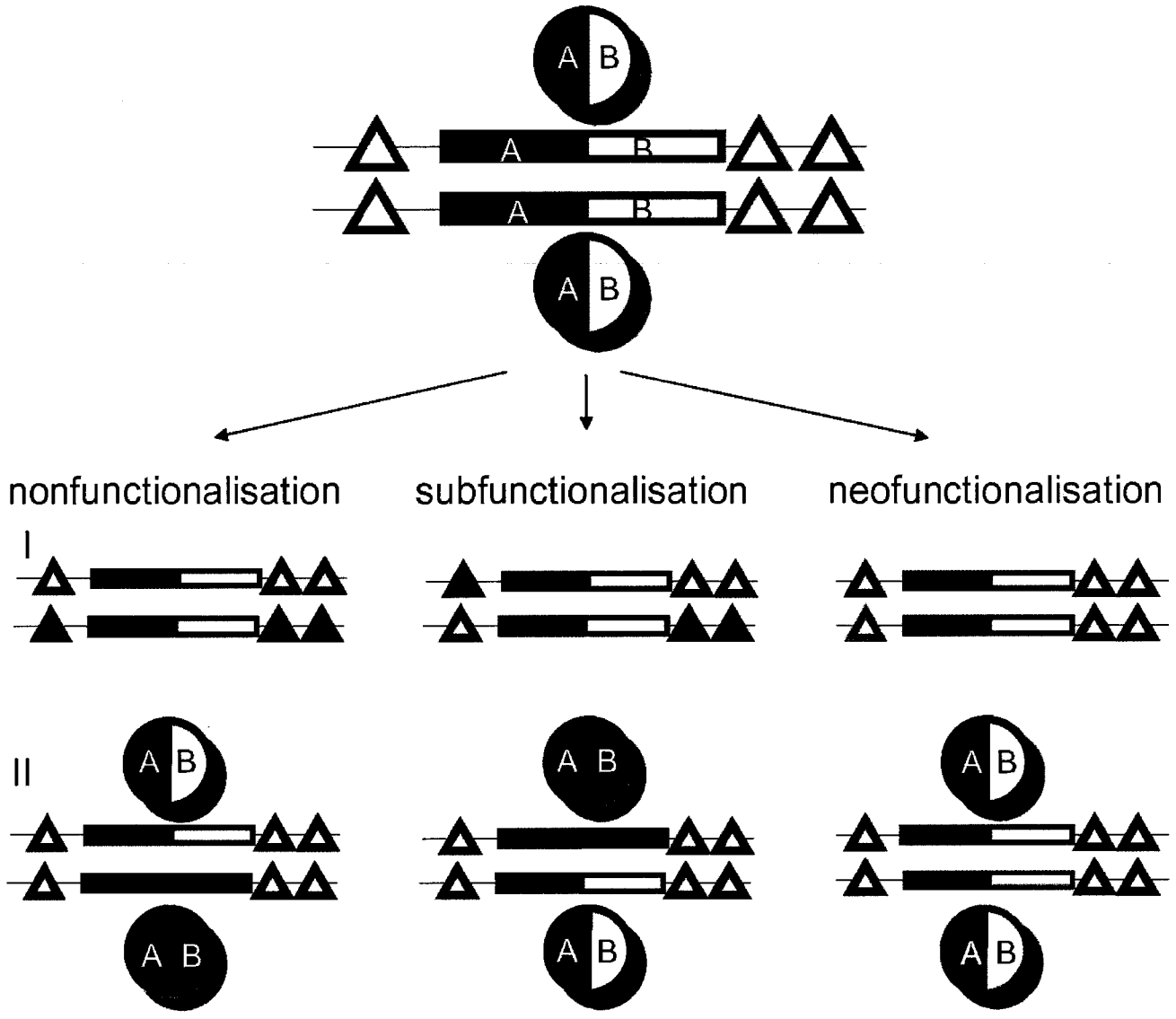
Let us consider an example: a duplication event produces two copies of the genes coding for both a receptor and its ligand. Freed of selective pressure, at least one of the gene copies encoding the receptor may mutate. If the mutation of the receptor affects its binding specificity to the corresponding ligand, complementary change in the ligand should follow. Support for this theory called ligand-receptor coevolution was described for LH (luteinizing hormone) and FSH (follicle-stimulating hormone). If the receptors and ligand are proteins on the sperm and egg, complimentary change in these proteins can cause reproductive isolation (Moyle et al. 1994).

One can predict that functional mutation of duplicated gene/genes at the top of the hierarchy of a signal transduction pathway may give a rise to a completely new developmental pathway, provided that compensatory changes also occur downstream. The Dorsal-Cactus pathway that is involved in the establishment of dorsal-ventral polarity of *Drosophila* and is used to induce production of macrophage inflammatory proteins in mammals (Whalen and Steward 1993) provide an evidence supporting this possibility. It can be hypothesized that mutation in the main genetic trigger of the developmental pathway enabled a product of the modified gene to bind to a different set of targets.

Interestingly, it has been demonstrated that genes located at the top of the hierarchy in the sex determination developmental pathway often diverge fast, while their downstream targets fall into the slowly evolving fraction of the pathway (Marin and Baker 1998). If this observation reflects a general tendency for evolution of developmental pathways, one can anticipate that expansion of *Hox* genes, which are primary regulators of many crucial developmental processes, could have generated tremendous genetic potential for evolutionary innovation.

Figure 1.4. Possible mechanisms of functional resolution of duplicated genes.

A hypothetical protein coding gene with two independently mutable functions (A and B) and three regulatory elements (grey, green and red triangles) is considered. (I): resolution at the level of non-coding regulatory elements; (II): resolution at the level of protein coding sequences. A circle reflects the functionality of two protein domains A and B before and after functional diversification at the level of coding sequences. A black fill indicates a loss of a regulatory or protein subfunction while an orange color illustrates the acquisition of a new function.


A B
A
B
A
B
A B
nonfunctionalisation
subfunctionalisation
neofunctionalisation
I
II
A B
A B
A B
A B
A B
A B

1.3.5. Functional diversification of duplicated genes at the level of coding sequences and non-coding elements

A vertebrate gene often has several functions which are dependent on the configuration of its regulatory elements and coding sequence. Regulatory elements target gene expression to restricted regions specifying its "area of operation". Genes often possess multiple regulatory elements which are responsible for individual regulatory subfunctions such as the expression of a gene in a specific tissue, cell lineage, or developmental stage (Piatigorsky and Wistow 1991; Gogarten and Olendzenski 1999). Mechanisms of functional diversification of duplicated genes can operate through changes in either coding or non-coding gene regions (Figure 1.4). For example, accumulation of deleterious mutations in the coding sequence of one of the duplicates will abolish production of a functional gene product. Alternatively, randomization of functional modules in regulatory elements will eliminate tissue specific expression of the gene. No matter what type of changes first resulted in nonfuctionalization, once the gene function is lost, all the functional modules of this gene are expected to be quickly eliminated in evolution as their presence is no longer required.

Depending on what part of the gene was affected by the change, there are two possible outcomes of neofuctionalization. Beneficial changes occurring in the coding sequence of a gene can enable the gene product to interact with a new set of targets (Figure 1.4II). Alternatively, changes in non-coding regulatory regions may facilitate acquisition of a new domain of expression (Figure 1.4I). Consistently, if complementary degenerative mutations alter the regions of coding sequences responsible for important protein domains, that would result in splitting of original subset of targets between the

duplicates. Degenerative mutations in regulatory elements can facilitate the preservation of duplicated gene copies by the dividing ancestral functions between the two gene copies (Force et al. 1999).

Subfunctionalization can occur by two different routes: qualitative or quantitative. Qualitative subfunctionalization occurs when one of the duplicates completely loses one of the gene's subfunction, whereas the other duplicate loses another subfunction. Quantitative subfunctionalization results from accumulation of mutations that cause a reduction in efficiency of different subfunctions in both duplicates. However, full or partial loss of subfunctions must complement each other in order to retain the full set of functions of the original gene (Force et al. 1999).

1.3.6. Role of *cis*-regulatory changes in diversification of the duplicated genes

Changes in regulatory regions of duplicated genes exemplify an important evolutionary mechanism that allows function diversification after gene duplication (Wittkopp 2006). Indeed, an increase in number of genes in higher organisms compared to lower organisms is not sufficient to explain a significant increase in morphological complexity (Aburomia et al. 2003). Therefore, changes in the coding sequences of the duplicated genes that are anticipated to mediate evolution of protein structures are far from being the most common mechanism responsible for functional diversity of duplicated paralogs. Tvrdik and Capecchi have recetly examined interchangeability of coding sequences of the two *Hox* genes, *Hoxa1* and *Hoxb1* by engineering mice where the *Hoxb1* gene was substituted for *Hoxa1* and vise versa. Remarkably, these genetic alterations did not interfere significantly

with normal mouse development (Tvrdik and Capecchi 2006). This study further reinforced the long standing view that changes in regulatory elements and not in the coding sequences of the duplicated genes have a principle role in diversification of duplicated paralogous *Hox* genes (Tvrdik and Capecchi 2006). Consistently, comparisons of *Hox* cluster sequences obtained from numerous genomic loci revealed the presence of a large number of putative regulatory elements (Hoegg and Meyer 2005). In addition, it has been recently demonstrated that non-coding regulatory DNA regions show accelerated substitution rate and are the fastest evolving elements in the human genome (Pollard et al. 2006). By extrapolation, the drastic increase in morphological complexity might have been achieved, at least in part, through the changes in non-coding regions of the duplicated genes. Indeed, the changes in *cis*-acting regulatory elements of the duplicated genes are thought to be easily tolerated as an individual regulatory unit is only responsible for a subset of overall gene expression. Therefore subtle and subsequent changes in functional subunits of regulatory *elements* can be viewed as a "safe" way for evolution to operate.

In the last five years, several research groups identified a number of duplicated genes that were presumably retained via changes in their regulatory elements as inferred from their differential expression domains. Hua and others have reported that two paralogous mammalian genes, *Nudt10* and *Nudt11* show differential pattern of expression but code for identical proteins (diphosphoinositol polyphosphate phosphohydrolases) (Wing et al. 2003). Other examples of duplicated genes which expression patterns diverged while their coding sequences remain unchanged include *sox9a/sox9b* genes and *mitf-m/mitf-b* in teleosts (Alschmied et al. 2002; Kluver et al. 2005). Both of these gene pairs resulted from teleost-specific duplication event.

Although the importance of changes in regulatory elements in functional diversification of duplicated genes is well accepted, data demonstrating firm correlation between the expression domains of the duplicated genes and molecular changes in their regulatory elements remain scarce and incomplete.

1.3.6. The *Hox* gene family provides an exceptional model to study molecular changes involved in evolution of gene duplicates

The *Hox* gene family is ideally suited to address the question of how genes acquire new functions as *Hox* genes underwent significant expansion during vertebrate evolution (Pendleton et al. 1993; Meyer and Malaga-Trillo 1999). Their clustered organization and structure remained unchanged defining precise locations of paralogs within the duplicated clusters and providing a useful model to study mechanisms of functional diversification for duplicated genes that lead to evolutionary innovations.

Comparison of the structure of the teleost *hox* clusters with those of mammals reveals the loss of individual genes within some of the duplicated clusters (Meyer and Malaga-Trillo 1999). Indeed genomic content appears only slightly increased after the whole genome duplication. For instance, mice and humans possess 39 *Hox* genes, while 47 *hox* genes are present in zebrafish (Meyer and Malaga-Trillo 1999; Hoegg et al. 2004).

Semon and Wolf have estimated that after the duplication event, a large proportion of the descendants of ancestral loci (~6.1% – 8.2%) underwent reciprocal gene loss in the *Takifugu* and zebrafish genomes. Reciprocal gene loss is one form of resolution of gene redundancy that is thought to contribute to speciation events (Semon and Wolfe 2007). Another possible fate of divergent resolution of gene duplicates is the retention of one

copy of gene duplicates in one species and retention of both copies in the other. Such is the case for *hoxb2* genes. The pufferfish *Takifugu rubripes* contains two *hoxb2* genes, *hoxb2a* and *hoxb2b* whereas only one, *hoxb2a*, is found in the zebrafish genome. This suggests the loss of one of the *hoxb2* duplicates following the divergence of zebrafish and *Takifugu* lineages (Amores et al. 1998). Finally, in some cases, paralogous genes were retained after the duplication event in all the teleost species examined thus far. Such is the case for the *hoxb5* genes in both zebrafish and *Takifugu* (Amores et al. 1998; Amores et al. 2004).

In their recent article, Krumlauf's team investigated regulatory basis for functional diversification of *hoxb2* duplicated genes in pufferfish (Tumpel et al. 2006). *hox2a* expression is observed in the anterior hindbrain (rhombomeres 1-2), whereas *hoxb2b* was reported to be expressed more posteriorly, in rhombomeres 2 through 7. By comparing regulatory activities of duplicated *cis*-regulatory elements in primary transgenic chicken and mouse embryos, Tumpel et al. were able to conclude that drift in functional modules of *hox2a* and *hox2b* regulatory elements is responsible for differences in expression pattern of the coparalogous genes (Tumpel et al. 2006).

Although this study was the first to experimentally address the significance of changes in *cis*-acting regulatory elements of duplicated genes, there were a few weaknesses in the experimental set up which, in my opinion, significantly limited the impact of the study. For example, regulatory elements were placed in reporter constructs outside of their endogenous environment. The constantly growing amount of data on the subject points towards high complexity of pathways involved in *Hox* gene regulation. Indeed, *Hox* clusters appear to operate as complex genetic units and regulation of corresponding *Hox* genes require regional integrity as it relies on competitive interactions

between multiple regulatory elements (Sharpe et al. 1998). Therefore, some of the intrinsic regulatory activities carried by regulatory subunits are bound to be missed when the regulatory elements are separated from their endogenous location. Furthermore, the activity of the regulatory elements from *hoxb2a* and *hoxb2b* genes was tested in mosaic transgenic animals that have further limited resolution of this study.

It is clear that given its importance in functional diversification of duplicated genes and its potential implications in acquisition of evolutionary novelties in vertebrate lineages, the role of regulatory evolution of *Hox* genes requires further investigation. My PhD project was devoted to provide a comprehensive study of changes in the regulatory elements that allow functional diversification of a specific pair of genes resulted from teleost specific duplication event, *hoxb5a* and *hoxb5b*.

While many individual genes were lost in one or more teleost lineages after the duplication, both *hoxb5a* and *hoxb5b* co-paralogs are found in all the teleost fish examined to date (Amores et al. 1998; Amores et al. 2004). Moreover, these loci occupy central position in the *Hoxb* clusters bridging "anterior" and "posterior" *hox* genes and were reported to harbor important regulatory elements, such as RARE that are necessary for proper functioning of other genes within the same cluster (Oosterveen et al. 2003). We reasoned that exploring the fate of *hoxb5* genes after duplication may provide us with significant insights into dynamics of molecular changes in duplicated gene loci.

If duplicated *hoxb5* genes were retained as a result of subfunctionalization at the level of regulatory elements, their combined expression patterns should correspond to that of a single ancestral *Hoxb5* gene. If the division of the original function occurred at the level of the coding sequences, Hoxb5a and Hoxb5b proteins would be expected to possess differential biochemical activities and would each have unique sets of targets that

complement the original gene function. To distinguish between the two possibilities, one should compare gene function and regulatory mechanisms in species that possess two copies of duplicated *hoxb5* genes to a species that have not undergone the *Hoxb* cluster duplication. While species with duplicated *Hoxb* clusters are present in abundance (most teleost fish) and are relatively easy to study, finding of a perfect representative of a "pre-duplicated" ancestral state appears to be a challenging task (Bruce et al. 2001). The fish-specific duplication event was estimated to occur in the actinopterygian fish after the divergence of sarcopterygians (Hoegg and Meyer 2005). In that case, actinopterygian fish, such as sturgeon, would be good candidates to represent an ancestral "preduplicated" state (Bruce et al. 2001). However, since gene expression data and genomic organization of *hox* genes in actinopterygians remain largely unknown, our comparisons should rely on well-studied sarcopterygians, such as the mouse.

1.4. Expression pattern and regulation of a single *Hoxb5* gene in mammals and duplicated *hoxb5* genes in teleost

1.4.1. Expression patterns of *Hoxb5* genes

Bruce and colleagues (2001) have previously examined expression patterns of *hoxb5a* and *hoxb5b* genes in zebrafish by in situ hybridization analysis. They determined that two zebrafish *hoxb5* genes, *hoxb5a* and *hoxb5b*, are expressed in overlapping yet distinct domains in the notochord, neural tube, somites and branchial arches (Figure 1.5II). Only *hoxb5a* was shown to be expressed in brachial arches, somites and notochord (Figure 1.5II, panels: A, C, D, F) while regions of exclusive *hoxb5b* expression included posterior

hindbrain and lateral line ganglia (Figure1.5II, panels: B, E). Combined expression domains of zebrafish *hoxb5a* and *hoxb5b* are strikingly similar to that of the single *Hoxb5* gene in the mouse that was shown to be expressed in somites, neural tube, dorsal root ganglia, lung, kidney, gut and nodose ganglion of the Xth cranial nerve (Hogan et al. 1988). Furthermore, Bruce and colleagues reported the biochemical equivalence of the Hoxb5a and Hoxb5b proteins (Bruce et al. 2001). Combined, these results suggest that the *hoxb5a* and *hoxb5b* genes underwent subfunctionalization through loss of *cis*-acting regulatory elements.

1.4.2. The interplay between regulatory elements and shared regulatory mechanisms of *Hox* gene expression

Regulatory properties of *Hoxb4/Hoxb5* intergenic region were first determined by Sharpe and colleagues (Sharpe et al. 1998). They described three regulatory regions from *Hoxb5* loci named LPM/PV, E, and D that were able to drive *lacZ* expression in mouse embryos with patterns that recapitulate aspects of endogenous *Hoxb5* and *Hoxb4* expression. The LPM/PV region spans the 5' flanking region of *Hoxb5* and is likely to include the gene promoter. In addition, LPM/PV was able to target reporter gene expression to lateral plate mesoderm derivatives and prevertebrae. Regions D and E were described as somite and neural enhancers respectively (Sharpe et al. 1998). Furthermore, Sharpe et al. tested enhancer activity of the three large fragments with respect to their differential interactions with the *Hoxb5* and *Hoxb4* promoters. Although regulatory elements responsible for the activity of the tested fragments were still unknown, Sharpe and colleagues demonstrated

Figure 1.5. The combined expression patterns of zebrafish *hoxb5a* and *hoxb5b* are similar to that of the single *Hoxb5* gene in the mouse.

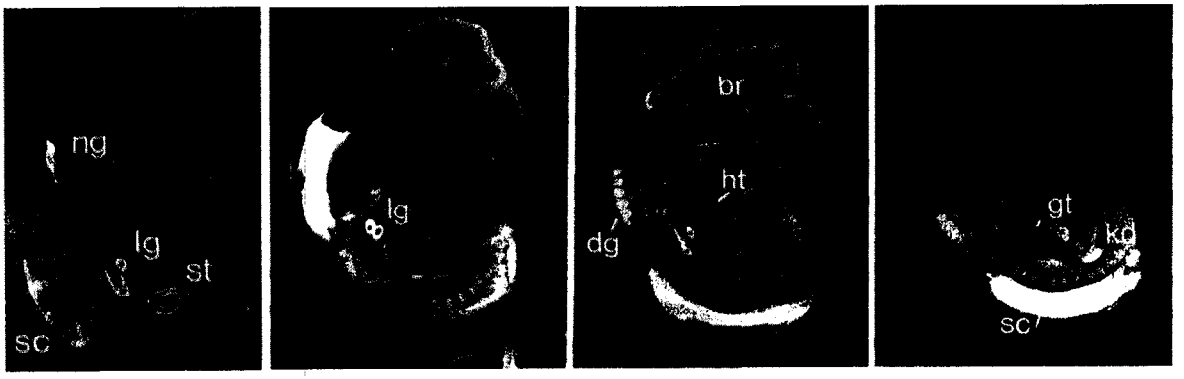
I. Expression of mouse *Hoxb5* as described in Hogan et al, 1988.

A hybridization signal was detected in the spinal cord (sc), dorsal root ganglia (Papalopulu et al.), lungs (Woolfe and Elgar), stomach (st), kidney (kd) and gut (gt) and nodose ganglion (ng) of the Xth cranial nerve.

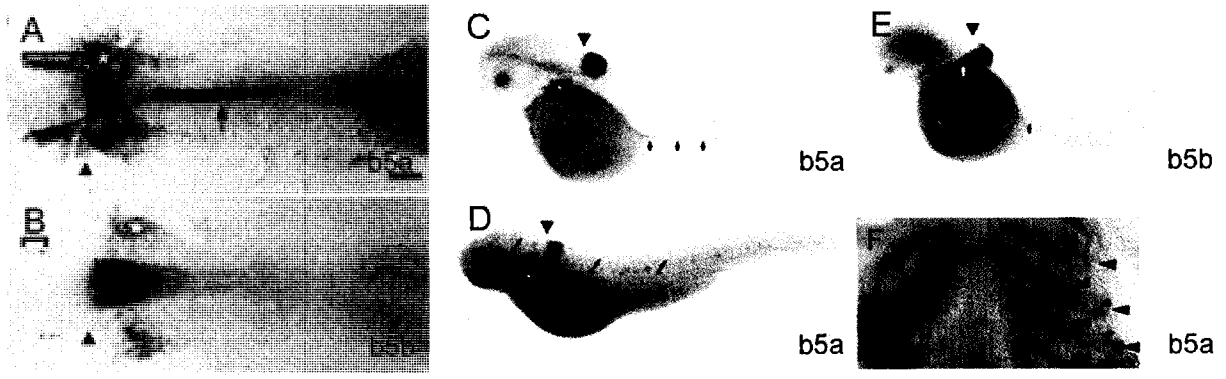
II. Expression of zebrafish *hoxb5a* and *hoxb5b* as described in Bruce et al., 2001.

Only *hoxb5a* was shown to be expressed in branchial arches, somites and notochord while *hoxb5b* is expressed in posterior hindbrain and lateral line. (A), (B) – 10 somites strage, (C), (E) – 2dpf, (D) – 3 dpf, (F) – 5dpf. Arrowheads indicate the anterior limit of expression in the CNE in all the panels. In A and B, the bracket indicates expression flanking the hindbrain whereas the asterisk highlights the expression flanking the spinal cord. The arrow in (A) points at expression in the notochord. *hoxb5a* expression in the branchial arches is marked with an asterisk in (C) and (D). Expression of *hoxb5b* in posterior lateral line ganglia is highlighted with a white arrow in (E). In (D), slanted arrows point at the neuromasts of the lateral line system where *hoxb5a* is expressed at 3dpf. Arrowheads in (F) indicate expression of *hoxb5a* in the mesenchyme between the ceratobranchials (cb1–4).

I



II



that transcriptional regulation of *Hoxb5* and *Hoxb4* genes relies on competitive interactions between shared regulatory regions residing in the intergenic locus.

A recent study by Hadrys et al. suggested that in addition to its role as a prevertebrae enhancer, the LPM/PM region also harbors a "master promoter" responsible for expression of both *hoxb3a* and *hoxb4a* genes in zebrafish (Hadrys et al. 2006). Hadrys and colleagues concluded that the upstream promoter controls transcription of pre-mRNA that contains exon sequences of both *hoxb4a* and *hoxb3a* genes. They speculated that this rather unconventional mechanism of regulation is used to ensure overlapping and tissue specific expression of the neighboring genes (Hadrys et al. 2006).

1.4.3. A RARE in 3' flanking region of *Hoxb5* is essential for rostral extension of the gene after rombomeric segmentation

While expression patterns of 3' *hox* genes (*Hox1-Hox4*) reach their anterior boundaries before hindbrain segmentation (Gould et al. 1998), 5' *hox* genes (*Hox5*, *Hox6* and *Hox8*) do so only after rhombomeric segmentation is complete (reviewed in Oosterveen et al. 2003). Thus, at E9.5, the *Hoxb5* anterior boundary of expression in the neural tube is at the level of the fifth somite that extends until the levels of rhombomeres 6/7 at E11.5 (Oosterveen et al. 2003). Oosterveen et al. have identified a functional RARE located downstream of *Hoxb5* and have demonstrated that this element plays a pivotal role in establishment of definite anterior boundaries of expression for at least two 5'*Hoxb* genes: *Hoxb5* and *Hoxb8* (Oosterveen et al. 2003). This study further reinforced the view that retinoic signaling is crucially important in *Hox* gene regulation.

Statement of inquiry

Gene duplication is an important mechanism leading to an increase in the amounts of genetic material and is a potential source of evolutionary novelty. The *Hox* genes constitute one of the best models for studying mechanisms of functional divergence for duplicated genes since they underwent significant expansion during evolution. The main goal of my PhD project was to identify and examine molecular changes that allowed the retention of *hoxb5a* and *hoxb5b* genes, duplicated in teleosts. Since Hoxb5a and Hoxb5b proteins were shown to be functionally equivalent in zebrafish, it is parsimonious to assume that functional diversification of *hoxb5* duplicates was mediated by changes in non-coding regulatory regions. To test this, we first carried out phylogenetic footprinting and identified three Conserved Non-coding elements (CNEs), J1, J2 and J3, residing in the 3'-flanking region of *Hoxb5* loci. We have generated reporter constructs carrying large fragments from the *hoxb5a* and *hoxb5b* zebrafish loci that encompass all the identified CNEs. When tested in transgenic animals, these constructs recapitulated both overlapping and unique aspects of *hoxb5a* and *hoxb5b* expression of the genes suggesting that changes at the level of regulatory elements may be responsible for the functional diversification of the two paralogs.

Our comparative sequence analysis indicated that one of the putative regulatory elements, J3, is only present in one of the two duplicated loci, *hoxb5a.* In constrast, J1 and J2 are retained in both duplicated *hoxb5* loci in *Takifugu* and zebrafish. In an attempt to understand the consequences of a loss of the J3 element from the *hoxb5b* locus, we generated two modified *hoxb5* loci: the *hoxb5a* locus from which J3 was deleted and the *hoxb5b* locus with an added J3. The functional tests of the reporter constructs carrying

the modified *hoxb5* loci have demonstrated that the loss of the J3 element from *hoxb5b* may indeed be responsible for at least some aspects of the functional diversification of the duplicated *hoxb5* paralogs in zebrafish.

To further clarify the individual contribution of each of the putative regulatory elements, identified by phylogenetic footprinting, to the overall expressional divergence of the two zebrafish paralogs, we separately tested the enhancer activity of J1, J2 and J3 from mouse and zebrafish in transgenic animals. The results of our functional studies suggested that molecular changes that occurred in the individual CNEs from the duplicated teleost loci account for some of the differences in expression patterns of *hoxb5a* and *hoxb5b* genes (Bruce et al., 2001). However, in general, the individual CNEs from the cognate zebrafish loci targeted the reporter gene expression to overlapping domains of the *hoxb5a* and *hoxb5b* expression. That contrasted with regulatory activity of constructs encompassing all the CNEs from zebrafish *hoxb5* loci that consistently recapitulated both overlapping and exclusive domains of expression of the corresponding genes. These data indicate that discrepancies between expression patterns of the two paralogous genes may require interaction between multiple regulatory elements from *hoxb5a* and *hoxb5b* loci identified in this study.

2. Materials and methods

2.1. List of oligonucleotides

Oligo name	Oligo sequence	Primary use
ampFOR	AACCGCGGTTGGTAACTGTCAGACC	PCR, cmPCRΔ vector
kanREV	AACCGCGGCCCTGCGCCATCAGATCC	PCR, cmPCRΔ vector
1AFOR	AAGGTACCGGTAGAGTTTGCTAATTGCG	PCR, 1A, *hoxb5a*
1AREV	AAGAGCTCGGACTAGCAGGTTTACCACC	PCR, 1A, *hoxb5a*
1BFOR	AAGGTACCGGTCCTGTATTCAGTCCACG	PCR, 1B, *hoxb5b*
1BREV	AAGAGCTCGGCTCATTGTACCAGAGAAGACG	PCR, 1B, *hoxb5b*
2AFOR	AAGGATCCCGGTGCACAAACCCTCTGTTC	PCR, 2A, *hoxb5a*
2AREV	AAACTATGGTACCGTCAGGGTTTGAGCTATGGCACC	PCR, 2A, *hoxb5a*
2BFOR	AAGGATCCCCCATAACAAAGCTACCCTTC	PCR, 2B, *hoxb5b*
2BREV	AACTAGTGGTACCCGCTAAATTCTATTAGGGCAGC	PCR, 2B, *hoxb5b*
3AFOR	AAAAGCTTCGTGATGCCCTAACGCTCTC	PCR, 3A, *hoxb5aegfp*
3AREV	AAGGTACCCCTGAGAACGAGTTTACAAAG	PCR, 3A, *hoxb5aegfp*
3AFOR	AAGGTACCCGTGATGCCCTAACGCTCTC	PCR, 3A, *hoxb5alacZ*
3AREV	AAGGATCCCCTGAGAACGAGTTTACAAAG	PCR, 3A, *hoxb5alacZ*
3BFOR	AAAAGCTTGGTCAACAAAGGCACGTGAT	PCR, 3B, *hoxb5begfp*
3BREV	AAGGTACCCCTGAGAACGAGTTTAGAAA	PCR, 3B, *hoxb5begfp*
3BFOR	AAGGTACCGGTCAACAAAGGCACGTGAT	PCR, 3B, *hoxb5blacZ*
3BREV	AAGGATCCCCTGAGAACGAGTTTAGAAA	PCR, 3B, *hoxb5blacZ*
4AFOR	AACTCGAGTCTTCCAGTCTCCAGCTCCG	PCR, 4A, *hoxb5alacZ/egfp*
4AREV	AACTCGAGGTCTCCGATGAAGCACTGGC	PCR, 4A, *hoxb5alacZ/egfp*
4BFOR	AACTCGAGAATCCGAGTCAGGAGACCC	PCR, 4B, *hoxb5blacZ/egfp*
4BREV	AACTCGAGGTGTGCTATTGCTGCTGCTGGTA	PCR, 4B, *hoxb5blacZ/egfp*
5AFOR	GGTACCCTATTCTTGAACGGGTTTTGG	PCR, 5A, *hoxb5a*ΔJ3
5AREV	GAATCCCTAAAGATGAGCAGCGTGCA	PCR, 5A, *hoxb5a*ΔJ3Δ
5BFOR	GGTACCCCGCGGTTGCTGGCACCA	PCR, 5B, *hoxb5b*insJ3
5BREV	GAATTCCAGACCAGTAAAACAGACCAC	PCR, 5B, *hoxb5b*insJ3
6AFOR	GGATCCCCCTCTTCTAGCCAGCAAGAT	PCR, 6A, *hoxb5a*ΔJ3
6AREV	CCGCGGCCTTTGTGTGTTCAACTCCC	PCR, 6A, *hoxb5a*ΔJ3
6BFOR	GGATCCTGCAGCAGCAGGATGCC	PCR, 6B, *hoxb5b*insJ3
6BREV	CCGCGGTTCCTTTGCCTCCAGTAAGC	PCR, 6B, *hoxb5b*insJ3
DrJ1AFOR	AAGCTTTCTAAGTTTTGAGAGCC	PCR, J1 CNE, *hoxb5a*
DrJ1AREV	GGTACCGCTTTCAAATGGGTAGAACCG	PCR, J1 CNE, *hoxb5a*
DrJ1BFOR	AAGCTTCAACCTTTCAATTTACACGG	PCR, J1 CNE, *hoxb5b*
DRJ1REV	GGTACCGTTGCCTTTTCACATCCAGC	PCR, J1 CNE, *hoxb5b*
MmJ1FOR	AAGCTTATGGTGAGGGTCTTGGAGGCTGTTT	PCR, J1 CNE, *Hoxb5*

MmJ1REV	GGTACCGACTTACTGTGGAGTTCTGG	PCR, J1 CNE, *Hoxb5*
DrJ2AFOR	AAGCTTACATACGACTAGTTGCAGG	PCR, J2 CNE, *hoxb5a*
DrJ2AREV	GGTACCCTTACACATAAATGTAGG	PCR, J2 CNE, *hoxb5a*
DrJ2BFOR	AAGCTTGGAACAATATTTTGAGCATAAG	PCR, J2 CNE, *hoxb5b*
DrJ2BREV	GGTACCAGGAGAATTACAGTCCTGAC	PCR, J2 CNE, *hoxb5b*
MmJ2FOR	AAGCTTTTTGTTGTCCGTGCTGTGGCC	PCR, J2 CNE, *Hoxb5*
MmJ2REV	GGTACCGCGTGACCTTCAAGTGGCTG	PCR, J2 CNE, *Hoxb5*
DrJ3AFOR	AAGCTTTGTTCTGAATACAACCAGGG	PCR, J3 CNE, *hoxb5a*
DrJ3AREV	GGTACCAACGCAGCCTTCTCTGTCCC	PCR, J3 CNE, *hoxb5a*
MmJ3FOR	AAGCTTTTGGAATGGAGACTTTTGG	PCR, J3 CNE, *Hoxb5*
MmJ3REV	GGTACCTGTGTATCCGGAAAACCG	PCR, J3 CNE, *Hoxb5*
lacZREV	CGCTCATCCGCCACATATCC	PCR, mouse screening
ßFOR	AGGGCAGAGCCATCTATTGC	PCR, mouse screening
hoxb5aFOR	CATAAATCGTGAAGCACAGCG	PCR, mouse screening
hoxb5bFOR	CATAAATTGTGCGGCGCTGG	PCR, mouse screening

2.2. Sequence alignments and phylogenetic analysis

Sequences of the *Danio rerio*, *Takifugu rubripes*, *Homo sapiens*, and *Mus musculus Hoxb* clusters, surrounding the *Hoxb5* gene, were obtained from Ensembl Genome Data Resources available at http://www.sanger.ac.uk/ by comparisons with the cDNA sequences of the corresponding genes. The zebrafish, human and mouse *Hoxb5* coding sequences were already available in GenBank. The coding sequences of *Takifugu rubripes hoxb5a* and *hoxb5b* genes were first identified by comparison with the zebrafish cDNAs and then confirmed by alignment with the genomic sequences of the *Takifugu rubripes hox* clusters (Amores et al. 2004).

Genomic sequences used in the analysis were obtained from the following contigs:

Danio rerio hoxaa, contig no. zv4_scaffold 335.4 (v.4), position 244122-265494;

Danio rerio hoxaa contig no. 10694.3 (v.3), position 66826-110453;

Danio rerio hoxba, contig no. 11157.4 (v.3), position 40334-67614;

Danio rerio hoxbb, contig no. BX001014.6 (v.4), position 25612-57312;

Takifugu rubripes hoxba, contig no. Scaffold_706, position 25001-50001;

Takifugu rubripes hoxbb, contig no. Scaffold_1245, position 50965-70027;

Mus musculus Hoxb, contig no. AL645478.15.1.121560, position 108131-130740;

Mus musculus Hoxa, contig no. CAAA01010313.1.1.18630, position 8656-31371;

Mus musculus Hoxc, contig no. AC124345.4.1.184532, position 6205-30704;

Homo sapiens HOXB, GenBank accession no. AC103702.3.1.187386, position 84742-109241.

Multiple sequence alignments were computed with MULTIPIPMAKER (available at http://bio.cse.psu.edu/pipmaker/) or with the Pretty Results program of the GCG Wisconsin package. The boundary of each CNE was determined as the longest region of high sequence similarity between mammalian and teleost sequences. Phylogenetic and molecular evolutionary analyses were conducted with the Kimura two-parameter model using neighbour-joining method and bootstrap as a test of phylogeny (MEGA version 3); (Kumar et al. 2001). Search for potential transcription binding sites in the identified CNEs was performed with Patch Search available at http://www.gene-regulation.com/cgi-bin/pub/programs/patch/bin/patch.cgi.

2.3. Design and engineering of transgenic constructs

2.3.1. Generation of reporter constructs that carry large regions of *hoxb5a* and *hoxb5b* loci via ET cloning

To generate reporter constructs encompassing large regions of zebrafish *hoxb5a* and *hoxb5b* loci, we used ET cloning technology that allows for the manipulation of large DNA fragments by homologous recombination in bacteria (Copeland et al. 2001). Molecular recombineering was conducted in DY380 and EL350 strains of *E. coli* provided by Dr. Neal Copeland, US National Cancer Institute, Frederick, USA. Both DY380 and EL350 cells contain defective λ prophage with recombineering genes *gam*, *beta* and *exo* controlled by a temperature-sensitive repressor. In addition, *E. coli* EL350 also contains a *cre* gene under the control of an arabinose-inducible promoter.

First, PACs containing *hoxb5a* and *hoxb5b* zebrafish loci were obtained from RZPD (plate positions are 254O17 and 227H9 respectively). Since the ET cloning technique relies on antibiotic resistance genes to select for recombined DNA molecules, the reporter genes could not be directly inserted into the original PACs as their backbone, pCYPAC6, contains *kan* and *amp* genes. In addition, the pCYPAC6 vector lacks convenient restriction sites for the isolation of the reporter cassette that is necessary in the procedure of mouse transgenesis. For these reasons, generation of reporter constructs was done in two steps.

Step 1. In the first recombineering step, we inserted fragments that contain all the conserved non-coding elements from the duplicated *hoxb5* loci (18,495 bp region from *hoxb5a* and 12,169 bp region from the *hoxb5b* locus) from corresponding PACs into a

modified chloramphenicol resistance vector (PCR2.1Δ)* (Figure 2.1). Homology arms, named 1A and 2A (~200 bp each) that contained sequences flanking the fragment from the *hoxb5a* locus were amplified from zebrafish genomic DNA and then inserted into PCR2.1Δ vector (Figure 2.1). Forward primers used to amplify homology arms were designed to incorporate *Kpn*I adapters to allow for easy extraction of the transgene cassettes from the final transgenic constructs, a step that is necessary for mouse transgenesis experiments.

Arm 1A matches a 200 bp DNA region that is located 1230 bp upstream of *hoxb5a* and arm 2A corresponds to the region located 1300 bp downstream of *hoxb4a* (Figure 2.1). The resulting construct is referred to as 1A2APCR2.1Δ (Figure 2.1).

To insert a large portion of the *hoxb5b* locus, homology arms 1B and 2B that correspond to ~200 bp DNA regions located ~1 kb upstream and ~9600 bp downstream of *hoxb5b* (directly upstream of *hoxb1b* gene) respectively were amplified from zebrafish genomic DNA and then subcloned into PCR2.1Δ vector (Figure 2.2). The resulting construct carrying the large region from the *hoxb5b* locus is referred to as 1B2BPCR2.1Δ.

Linear targeting cassettes, 1A2APCR2.1Δ and 1B2BPCR2.1Δ, were electroporated into DY380 cells containing the corresponding PACs (PAC 254O17 for *hoxb5a* and PAC 227H9 for the *hoxb5b* insert). The cells were incubated 42 °C for 15 minutes to induce the production of recombineering proteins, RecE and RecT. Homologous recombination between the linear targeting cassettes and the PACs generated recombinants that contained corresponding fragments from the *hoxb5a* and

*PCR2.1Δ vector was previously generated by amplification of PCR2.1 (Invitrogen) vector carrying cm resistance gene (Dr. Luc Poitras) by using primers that anneal directly downstream of *amp* and upstream of *kan* genes respectively and contained SstII adaptors. Introduced SstII restriction sites were then used to ligate the ends of the amplified vector.

Figure 2.1. Generation of the *hoxb5a*PCR2.1Δ construct: retrieving a large region of the *hoxb5a* locus from a PAC clone into a chloramphenicol resistance vector.

The *hoxb5a* gene from zebrafish is present in the 254O17 PAC clone. 1A and 2A homology arms that contained sequences flanking a 18,495 bp fragment from the *hoxb5a* locus were inserted into the PCR2.1Δ vector to generate the targeting construct. To allow recombination to occur (cross), the linear targeting construct was electroporated into DY380 electrocompetent cells containing the 254O17 PAC. Rectangles represent exon sequences of the corresponding genes, 1A and 2A – homology arms, *cm* – chloramphenicol resistance gene. See text for more details.

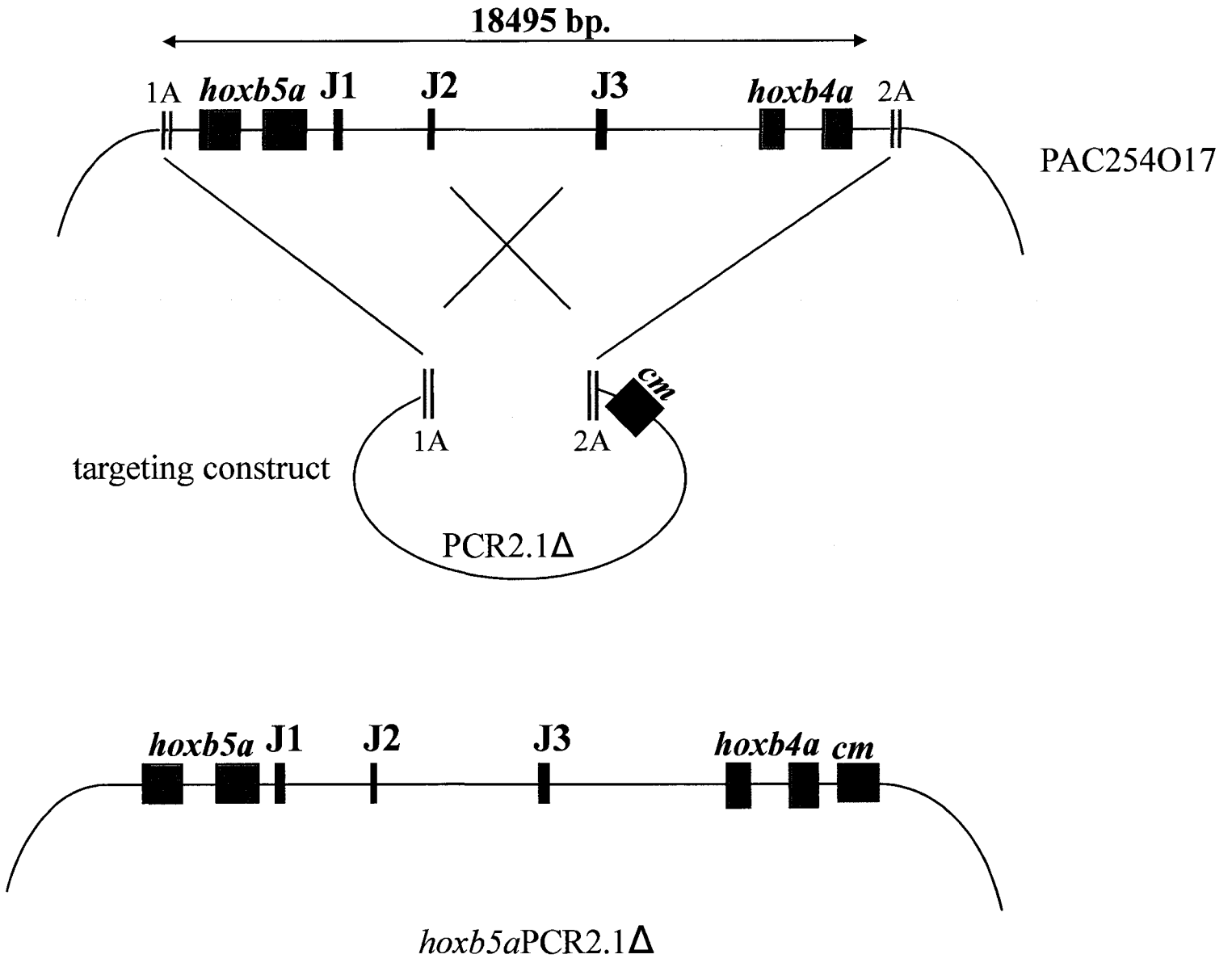
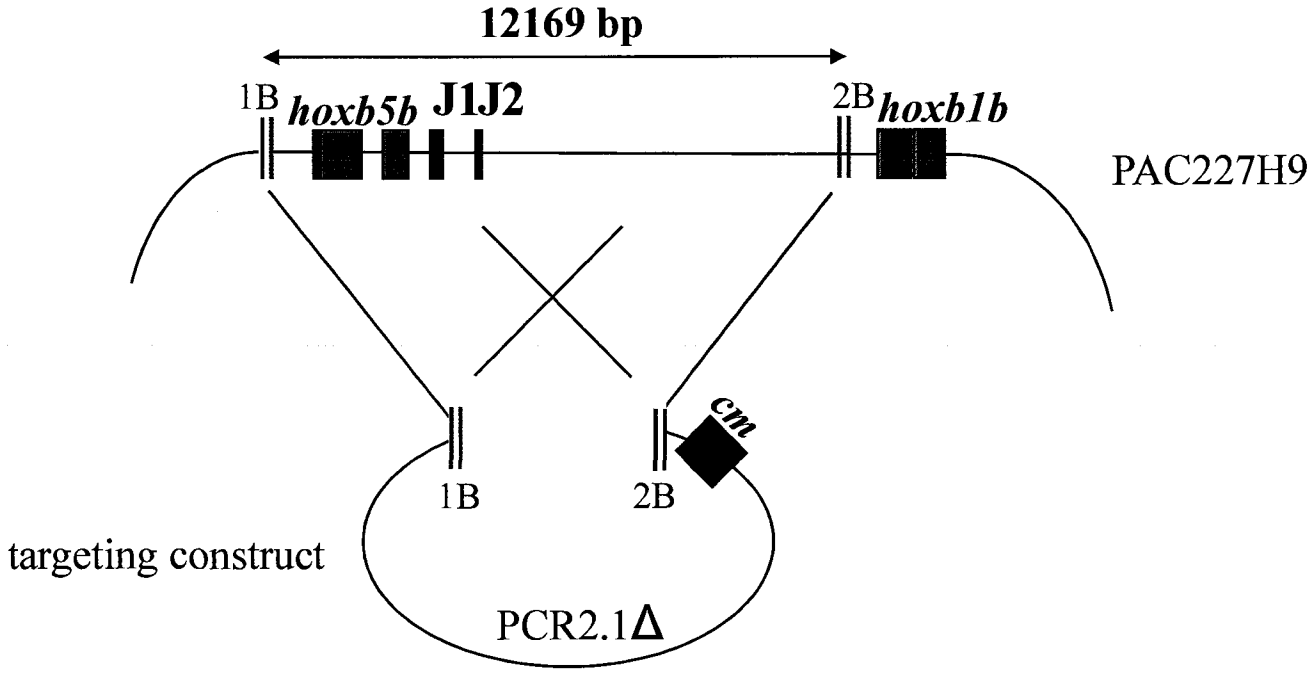
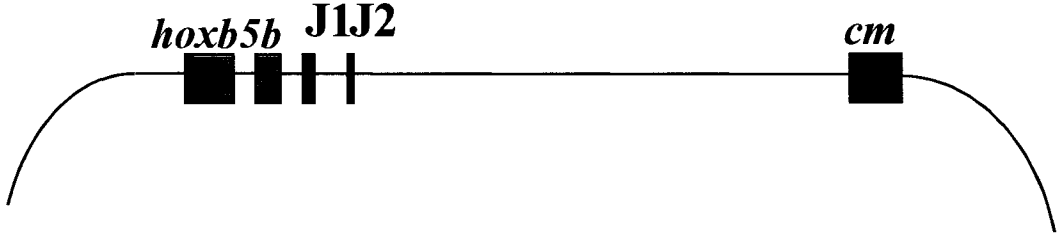

18495 bp.
1A
hoxb5a
J1
J2
J3
hoxb4a
2A
PAC254O17
cm
1A
2A
targeting construct
PCR2.1Δ
hoxb5a
J1
J2
J3
hoxb4a
cm
hoxb5aPCR2.1Δ
1kb
1"

Figure 2.2. Generation of the *hoxb5b*PCR2.1Δ construct: retrieving a large region of the *hoxb5b* locus from a PAC clone into a chloramphenicol resistance vector.

The *hoxb5b* gene from zebrafish is present in the 227H9 PAC clone. 1B and 2B homology arms containing sequences flanking a 12,169bp fragment from the *hoxb5b* locus were inserted into the PCR2.1Δ vector to generate a targetting construct. To allow recombination to occur (cross), the linear targeting construct was electroporated into DY380 electrocompetent cells containing the 227H19 PAC. Rectangles represent exon sequences of the corresponding genes, 1B and 2B – homology arms, *cm* – chloramphenicol resistance gene. See text for more details.



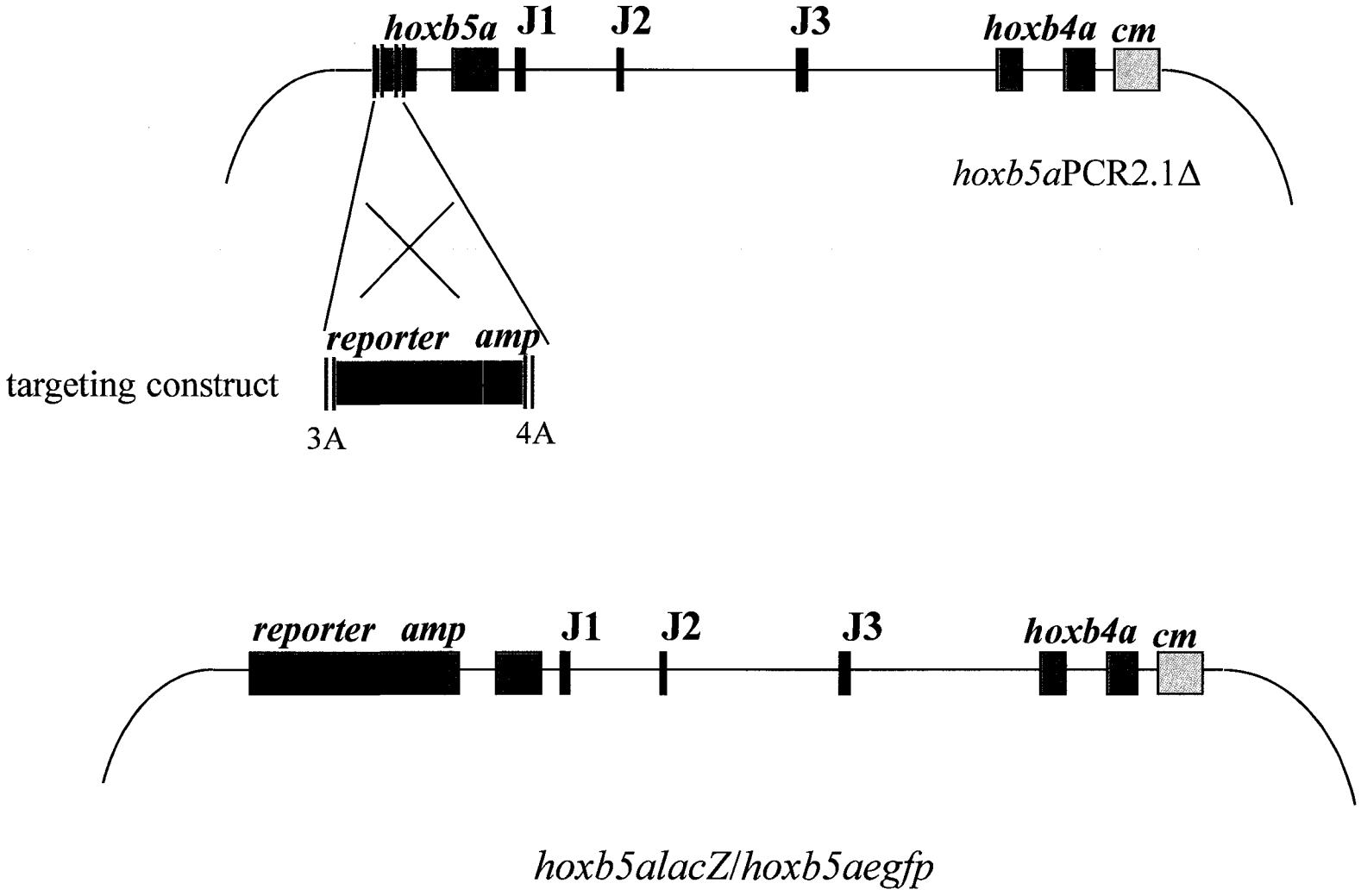


*hoxb5b*PCR2.1Δ

1kb
1”

Figure 2.3. Generation of the *hoxb5aegfp* and *hoxb5alacZ* reporter constructs: Insertion of reporter genes in frame with the coding sequence of the *hoxb5a* gene.

The targeting contructs contained two cloned segments of the *hoxb5a* gene (3A and 4A homology arms) with the reporter gene (*lacZ/egfp*) inserted between them. The *lacZ* and *egfp* targeting cassettes were separately electroporated into DY380 electrocompetent cells containing the *hoxb5a*PCR2.1Δ construct resulting from a previous round of ET cloning. Rectangles represent exon sequences of the corresponding genes, 3A and 4A – homology arms, reporter – *egfp/lacZ* genes, *cm* – chloramphenicol resistance gene, *amp* – ampicillin resistance gene. See text for more details.


hoxb5a
J1
J2
J3
hoxb4a
cm
hoxb5aPCR2.1Δ
reporter
amp
targeting construct
3A
4A
reporter
amp
J1
J2
J3
hoxb4a
cm
hoxb5alacZ/hoxb5aegfp
1kb
1"

hoxb5b loci in the PCR2.1Δ vector backbone (Figure 2.1 and Figure 2.2). The occurrence of a proper recombination event was confirmed by sequencing. DY380/*hoxb5a*PCR2.1Δ and DY380/*hoxb5b*PCR2.1Δ cells were made electrocompetent and used to generate reporter constructs for the second round of ET cloning.

Step 2. In the second round of ET cloning, reporter genes were inserted in frame with coding sequences of the *hoxb5a* and *hoxb5b* genes (Figure 2.3 and Figure 2.4).

Two ~200-bp sequences homologous to the two adjacent regions of exon 1 from each of *hoxb5a* and *hoxb5b* (named 3A, 4A and 3B, 4B respectively) were first PCR-amplified and then inserted into an *egfpamp*PCR2.1 vector*. Homologous recombination between linear targeting cassettes (3A*egfpamp*4A) and (3B*egfpamp*4B) and the recombined plasmids from the first round of ET cloning generated amp^R recombinants that carry large fragments of *hoxb5a* and *hoxb5* loci in which the *egfp* reporter gene was inserted in frame with the coding sequence of *hoxb5a* and *hoxb5b* genes (Figure 2.3 and Figure 2.4). Proper recombination was confirmed by sequencing. The resulting transgenic constructs are further referred to as *hoxb5aegfp* and *hoxb5begfp*.

A similar approach was taken for the generation of *lacZ* reporter constructs. The homology arms 3A, 3B, 4A and 4B were used to generate *lacZ* reporter constructs which were identical to those used for the *egfp* constructs (described above) and were inserted into a *lacZamp*PCR2.1 vector** (Figure 2.3 and Figure 2.4). Oligonucleotides designed

---

* *egfpamp*PCR2.1 vector that was used as a backbone to generate *egfp* targeting constructs was previously constructed in collaboration with Dr. Luc Poitras by inserting *amp* gene into *EcoR*I site and egfp into SstI site into PCR2.1 vector provided by Invitrogen.

** *lacZamp* PCR2.1 construct that contains original PCR2.1 (Invitrogen) vector, in which *amp* gene was inserted into *EcoR*I site and *lacZ* was inserted into BamHI site, was previously constructed in collaboration with Dr. Luc Poitras.

for the *lacZ* targeting constructs (3AFOR, 3AREV and 3BFOR, 3BREV) anneal to the same sequences of the *hoxb5* exons (and thus amplified DNA regions identical to those used to generate *egfp* targeting constructs) but contain different adaptors. Linear targeting cassettes were electroporated into electrocompetent DY380 cells containing plasmids generated at the first step of ET cloning: *hoxb5a*PCR2.1 and *hoxb5b*PCR2.1. Recombinants were selected for amp resistance and the occurrence of a proper recombination event was confirmed by sequencing. The resulting reporter constructs are referred to as *hoxb5alacZ* and *hoxb5blacZ*.

Cells containing plasmids engineered in the second round of ET cloning (DY380/*hoxb5aegfp*, DY380/*hoxb5begfp* and DY380/*hoxb5alacZ*, DY380/*hoxb5blacZ*) were made electrocompetent and used in the third round of ET cloning.

2.3.2. Generation of a "deletion construct": construct containing the zebrafish *hoxb5a* locus from which the J3 element was deleted

Amplified homology regions (~200 bp long) located upstream and downstream of J3, named 5A and 6A were inserted into the PL452 vector in front of a neomycin resistance gene (Copeland et al. 2001). The targeting cassette was extracted and electroporated into electrocompetent DY380 cells containing *hoxb5alacZ* and *hoxb5aegfp* constructs. Electroporated cells were plated on kanamycin plates to select for bacteria containing circular plasmids carrying the neomycin resistance gene (Figure 2.5). The occurrence of proper recombination event (substitution of J3 for the *neo* gene) was further verified by sequencing. The deletion constructs are referred to as *hoxb5a*ΔJ3*lacZ* and *hoxb5a*ΔJ3*egfp*.

Figure 2.4. Generation of the *hoxb5begfp* and *hoxb5blacZ* reporter constructs: Insertion of reporter genes in frame with the coding sequence of the *hoxb5b* gene.

The targeting contructs contained two cloned segments of the *hoxb5b* gene (3B and 4AB homology arms) with the reporter gene (*lacZ/egfp*) inserted between them. The *lacZ* and *egfp* targeting cassettes were separately electroporated into DY380 electrocompetent cells containing the *hoxb5b*PCR2.1Δ construct resulting from a previous round of ET cloning. Rectangles represent exon sequences of the corresponding genes, 3B and 4B – homology arms, reporter – *egfp*/*lacZ* genes, *cm* – chloramphenicol resistance gene, *amp* – ampicillin resistance gene. See text for more details.

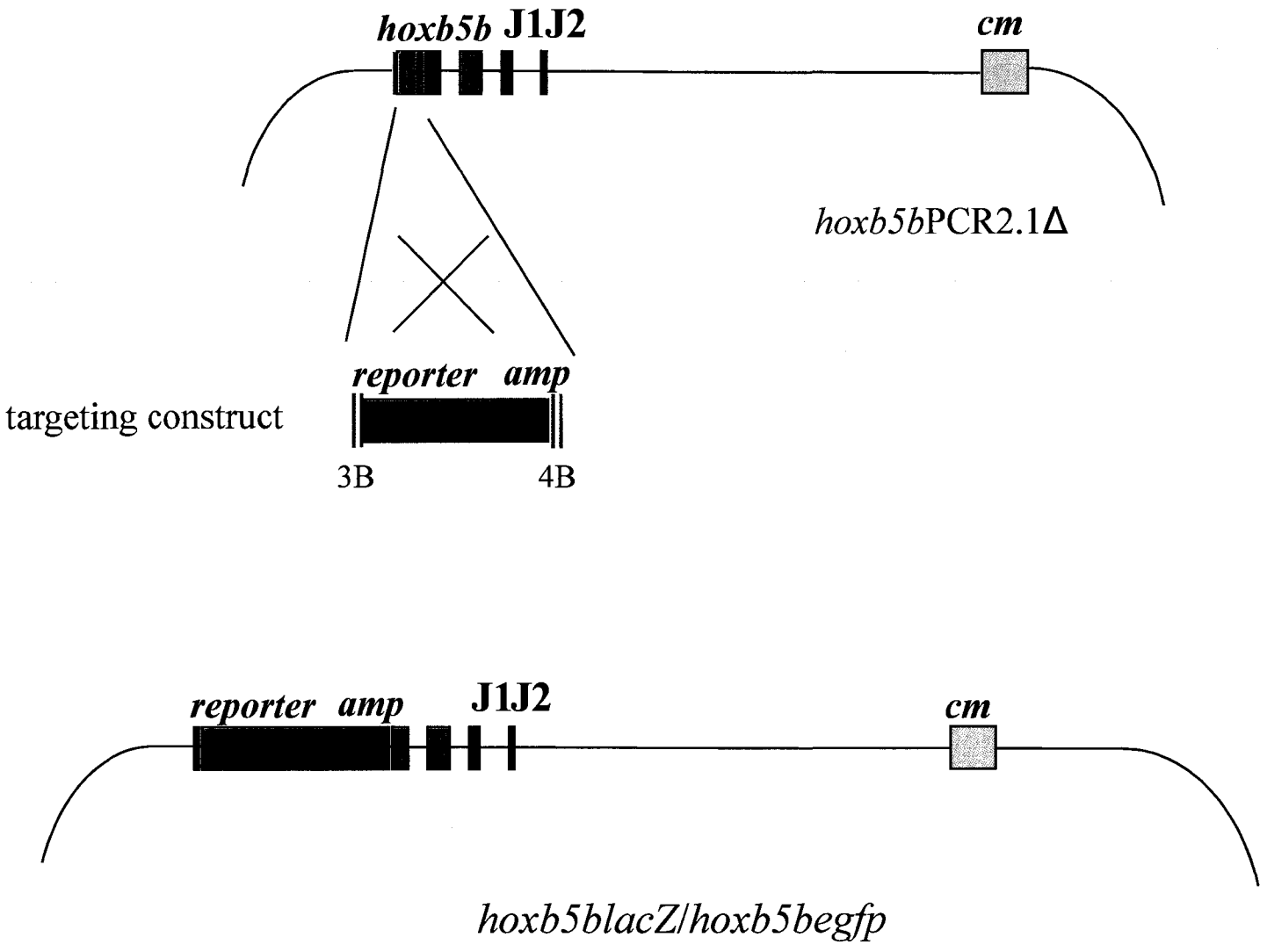
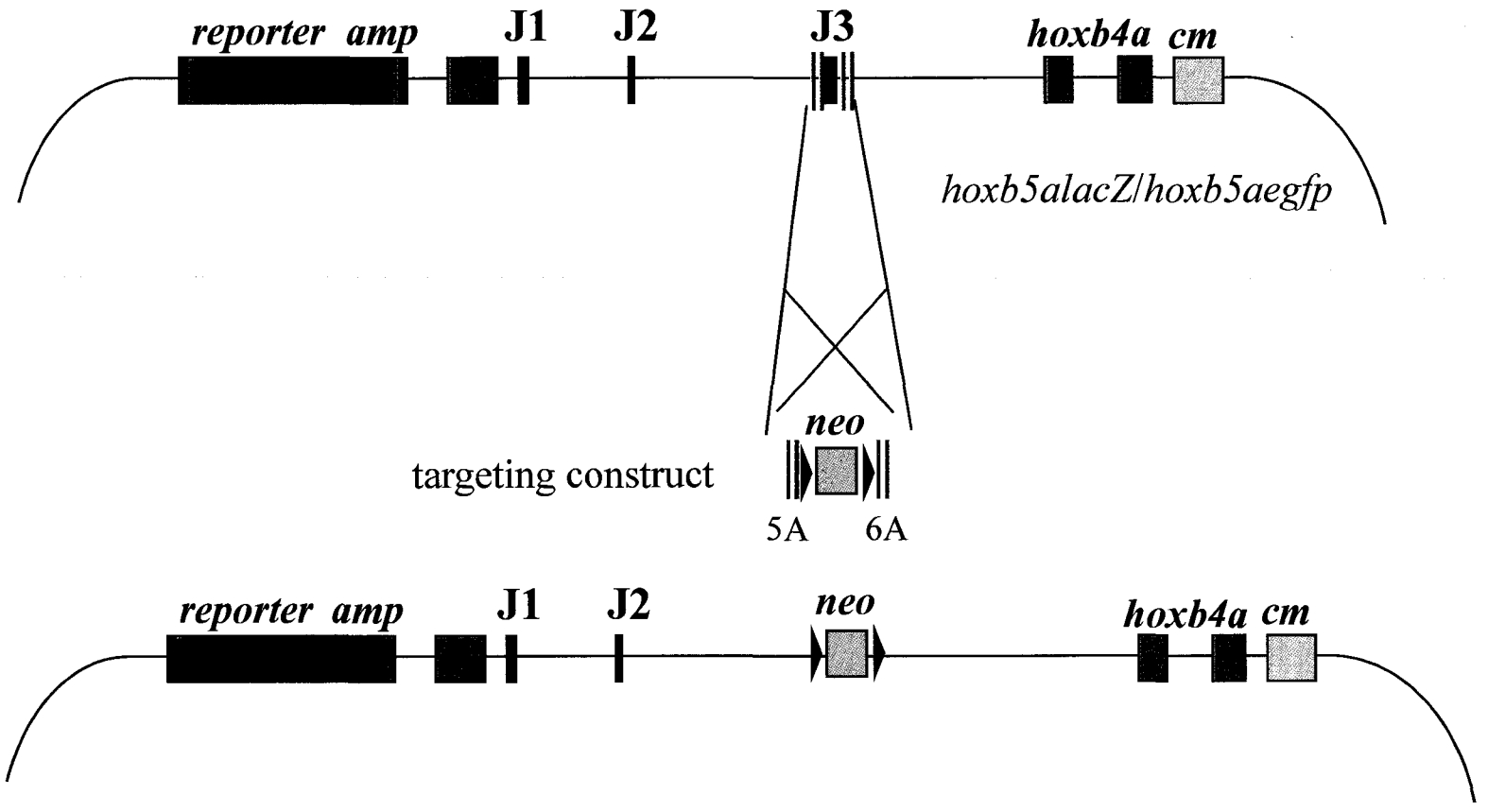

hoxb5b J1J2
cm
hoxb5bPCR2.1Δ
reporter amp
targeting construct
3B
4B
reporter amp
J1J2
cm
hoxb5blacZ/hoxb5begfp
1kb
1"

Figure 2.5. Generation of the *hoxb5a*ΔJ3 reporter constructs: deletion of the J3 element from the *hoxb5aegfp* and *hoxb5alacZ* constructs.

To produce a targeting construct for this step of ET cloning, two DNA segments flanking the J3 element (5A and 6A homology arms) were subcloned into the PL452 vector containing two *lox*P sites and the neomycin resistance gene. The linear targeting cassette was then separately electroporated into DY380 electrocompetent cells containing the *hoxb5aegfp* and *hoxb5alacZ* constructs resulting from the previous round of ET cloning. Rectangles represent exon sequences of the corresponding genes, 5A and 6A – homology arms, reporter – *egfp*/*lacZ* genes, *cm* – chloramphenicol resistance gene, *neo* – neomycine resistant gene, *amp* – ampicillin resistance gene, black triangles – *lox*P sites. See text for more details.



hoxb5aΔJ3lacZ/hoxb5aΔJ3egfp

1kb
1”

for the *lacZ* targeting constructs (3AFOR, 3AREV and 3BFOR, 3BREV) anneal to the same sequences of the *hoxb5* exons (and thus amplified DNA regions identical to those used to generate *egfp* targeting constructs) but contain different adaptors. Linear targeting cassettes were electroporated into electrocompetent DY380 cells containing plasmids generated at the first step of ET cloning: *hoxb5a*PCR2.1 and *hoxb5b*PCR2.1. Recombinants were selected for amp resistance and the occurrence of a proper recombination event was confirmed by sequencing. The resulting reporter constructs are referred to as *hoxb5alacZ* and *hoxb5blacZ*.

Cells containing plasmids engineered in the second round of ET cloning (DY380/*hoxb5aegfp*, DY380/*hoxb5begfp* and DY380/*hoxb5alacZ*, DY380/*hoxb5blacZ*) were made electrocompetent and used in the third round of ET cloning.

2.3.2. Generation of a "deletion construct": construct containing the zebrafish *hoxb5a* locus from which the J3 element was deleted

Amplified homology regions (~200 bp long) located upstream and downstream of J3, named 5A and 6A were inserted into the PL452 vector in front of a neomycin resistance gene (Copeland et al. 2001). The targeting cassette was extracted and electroporated into electrocompetent DY380 cells containing *hoxb5alacZ* and *hoxb5aegfp* constructs. Electroporated cells were plated on kanamycin plates to select for bacteria containing circular plasmids carrying the neomycin resistance gene (Figure 2.5). The occurrence of proper recombination event (substitution of J3 for the *neo* gene) was further verified by sequencing. The deletion constructs are referred to as *hoxb5a*ΔJ3*lacZ* and *hoxb5a*ΔJ3*egfp*.

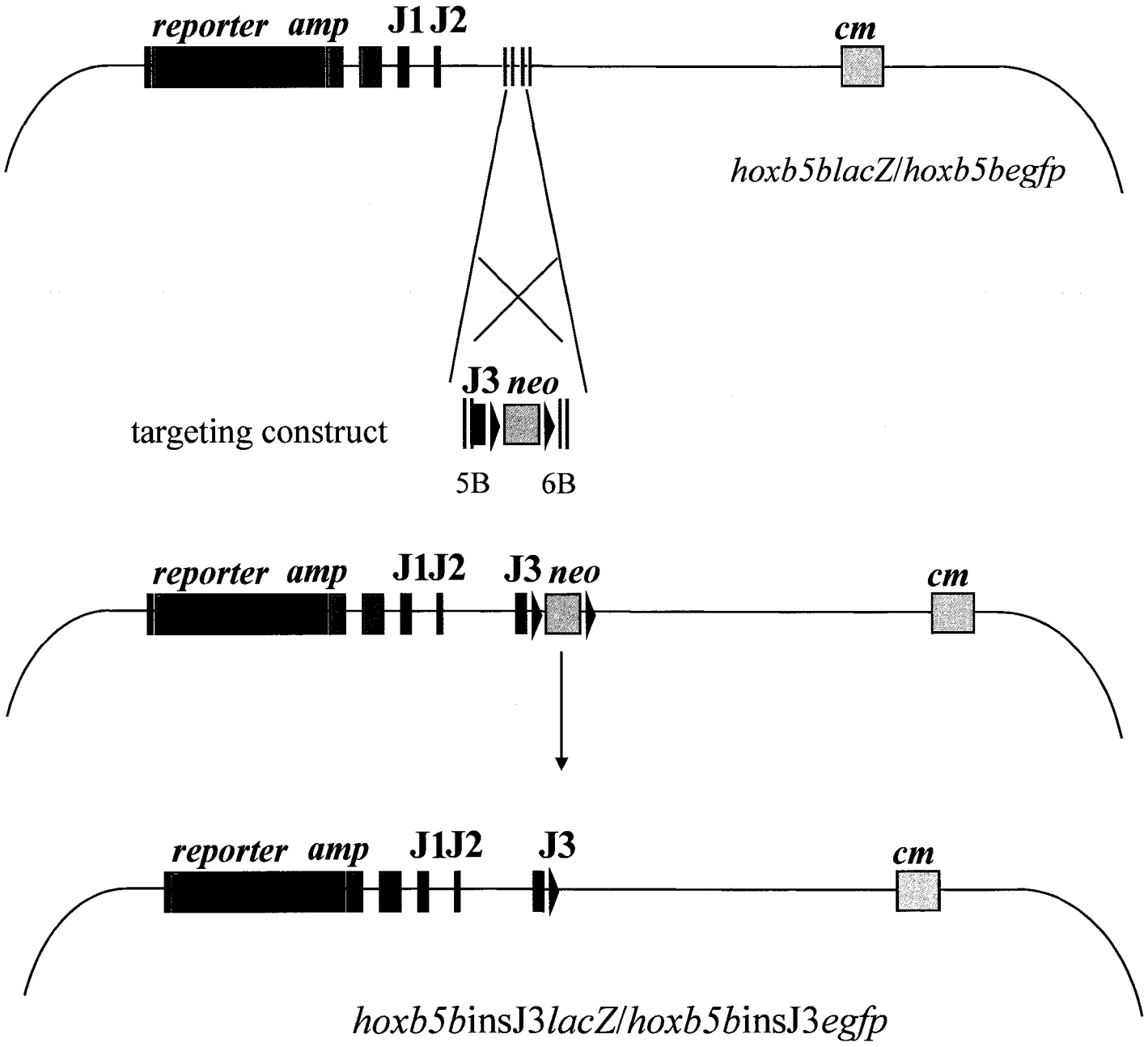
2.3.3. Generation of an "insertion construct": construct encompassing zebrafish hoxb5b locus in which the J3 element was inserted

We first estimated where the J3 element would be located if present in the *hoxbb* cluster. Although the *hoxb5b* locus appears to have shrunk after duplication, elements existing in both loci (J1, J2 and mir10a) are present in the same order and the relative distances between them remained the same. By extrapolation, if present, the J3 element would be located ~3.1 kb downstream of the *hoxb5b* gene.

We amplified ~200 bp homology regions named 5B and 6B directly located upstream and downstream of the hypothetical location of J3 and separately inserted them into the PL452 vector (Copeland et al. 2001) that contains a floxed neomycin resistance gene. The J3 element was amplified from the *hoxb5a* locus and inserted between the homology arms. Next, the targeting cassette was extracted and electroporated into electrocompetent EL350 cells carrying *hoxb5blacZ* and *hoxb5begfp* constructs. Electroporated cells were plated on kanamycin plates to select for bacteria in which a recombination event occurred (Figure 2.6). To excise the neo gene adjacent to the introduced J3 element, EL350 cells hosting the *hoxb5b*insJ3*lacZ* and *hoxb5b*insJ3*egfp* plasmids were grown in arabinose-containing medium to induce recombination between loxP sites flanking the *neo* gene.

Figure 2.6. Generation of the *hoxb5b*insJ3 reporter constructs: insertion of the J3 element in the *hoxb5begfp* and *hoxb5blacZ* constructs.

To produce a targeting construct, two DNA regions from the *hoxb5b* locus (5B and 6B) and the J3 element from the *hoxb5a* locus were inserted into the PL452 vector containing *lox*P sites and the neomycin resistance gene. The linear targeting cassette was then separately electroporated into EL350 electrocompetent cells containing the *hoxb5begfp* and *hoxb5blacZ* constructs resulting from the previous round of ET cloning. To remove the neomycine resistance gene directly adjacent to the J3 element, cells containing the resultant constructs were grown in arabinose-containing medium to induce cre that catalyzes recombination between the *lox*P sites. Rectangles represent exon sequences of the corresponding genes, 5B and 6B – homology arms, reporter – *egfp*/*lacZ* genes, *cm* – chloramphenicol resistance gene, *neo* – neomycine resistant gene, *amp* – ampicillin resistance gene, black triangles – *lox*P sites. See text for more details.


reporter amp J1 J2
cm
hoxb5blacZ/hoxb5begfp
J3 neo
targeting construct
5B 6B
reporter amp J1J2 J3 neo
cm
reporter amp J1J2 J3
cm
hoxb5binsJ3lacZ/hoxb5binsJ3egfp
1kb
1”

2.3.4. Generation of *lacZ* and *egfp* reporter constructs carrying individual CNEs from mouse and zebrafish loci

To determine the activity of the identified CNEs in mouse and in zebrafish, individual CNEs were PCR-amplified with the corresponding oligonucleotides (see list of oligonucleotides) designed from zebrafish and mouse genomic DNA sequences. *Hin*dIII and *Kpn*I restriction sites were incorporated in the forward and reverse oligonucleotide sequences respectively to allow for directional subcloning of individual elements in p1229 downstream of the *lacZ* reporter gene. The CNEs were then separately subcloned into the p1229 vector containing a human ß-globin minimal promoter coupled to the *lacZ* reporter gene (Yee and Rigby 1993).

To generate primary transgenic zebrafish, the PCR fragments were inserted in a ßpegfp vector that includes the minimal promoter from the p1229 plasmid and the coding sequence of the enhanced green florescent protein (egfp) (Cormack et al. 1996; Zerucha et al. 2000).

Forward and reverse oligonulcleotides used to amplify individual CNEs were identical in sequence to those designed for the *lacZ* reporter but contained *Kpn*I adaptors. Individual elements were inserted downstream of a reporter gene in *Kpn*I site of ßpegfp and p1229 vectors in a way that mimics their position and orientation in the original genomic context.

2.4. Zebrafish and mouse transgenesis

2.4.1. Production of transgenic mice

All experiments were performed according to the guidelines of the Canadian Council on Animal Care and approved by the institutional animal care committees. For the production of transgenic mice, the transgenic cassettes were freed from the vector sequences, purified using GeneCAPSULE^{TM} nucleic acids extraction kit and then injected into fertilized oocytes at a concentration of 2.5- 5ng/μl according to standard procedures (Hogan 1986). Reporter activities of transgenic constructs were analyzed in either founder embryos or from established transgenic lines.

2.4.2 Extraction of genomic DNA from mouse tissues

Placental or ear tissue was collected from each mouse embryo/pup and placed in individual Eppendorf tube 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, 350 μg Proteinase K) was added to the each tissue sample and incubated overnight at 55°C with intermittent mixing. Next, 250 μl of 6M NaCl solution was added to each sample and the tubes were centrifuged for 10 min at 13000 rpm. The supernatant containing the DNA was transferred to a fresh tube and 500 μl of isopropanol was added. DNA was precipitated by centrifuging for 15 minutes at 13,000 rpm at 4°C and washed once with 70% ethanol. Depending on the amount of DNA, the pellet was resuspended in 200-500 μl of TE buffer (pH 8.0).

2.4.3. Screening of mouse embryos or pups for the presence of the *lacZ* transgene using PCR

The presence of a transgene in mice carrying the *hoxb5alacZ*, *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* constructs was detected by PCR using genomic DNA extracted from extraembryonic tissue. One of the oligonucleotides recognized the *lacZ* sequence (LACZREV, AGGGCAGAGCCATCTATTGC) which is present in all four types of transgenics while another oligonucleotide annealed to the *hoxb5a* or the *hoxb5b* promoter region (hoxb5aFOR, CATAAATCGTGAAGCACAGCG and hoxb5bREV, CATAAATTGTGCGGCGCTGG respectively). For the genotyping of primary transgenic embryos which were injected with β-globin carrying constructs, oligonucleotides annealing to the ß-globin promoter and *lacZ* reporter gene were used (ß FOR, AGGGCAGAGCCATCTATTGC and lacZREV, CGCTCATCCGCCACATATCC respectively). Part of the fetal hemoglobin gene was amplified with the FHFOR, GATCATGACCGCCGTAGG and FHREV, CATGAACTTGTCCCAGGCTT oligonucleotides as an internal positive control for each PCR reaction.

2.4.4. Whole mount β-galactosidase staining of transgenic mouse embryos

Mouse embryos were harvested from foster mothers and analyzed for *lacZ* expression by ß-galactosidase staining as previously described (Fraidenraich et al. 1998) (Larochelle et al. 1999). Briefly, the harvested embryos were dissected from extraembryonic tissues, washed 3 x 15 minutes in 1 x Phosphate Buffered Saline (PBS: 140 mM NaCl, 3 mM

KCl, 10 mM sodium phosphate, pH 7.2), transferred in FIX solution (1% formaldehyde, 0.2% gluteraldehyde, 0.02% NP-40, 1 X PBS) and placed at +4°C for 40 minutes. The fixed embryos were washed again in 1xPBS and then incubated in X-gal staining solution (1mg/ml X-gal, 5 mM $K_3Fe(CN)_6$, 5 mM K_4Fe(Gavalas et al.)$_6$, 2 mM $MgCl_2$, 0.02% NP-40, 1XPBS) overnight at 30°C. The next morning, stained embryos were removed from the staining solution, washed in 1xPBS and stored in 1xPBS with 10 mM EDTA.

2.4.5. β-galactosidase staining of parasaggital sections of transgenic mouse embryos

Mouse embryos were collected at E12.5 (12.5 days post fertilization) from a cross between wild type female and transgenic male. They were fixed and washed as described above. Prior to sectioning, embryos were cryoequilibrated in 30% sucrose, 1xPBS solution overnight at +4°C. Next, embryos were embedded in TBS (Tissue Freezing Medium) on dry ice and sectioned parasagitally at a thickness of 20 μm or 40 μm using a LEICA CM1850 cryostat. The sections were spread onto Fisher Superfrost/Plus slides and incubated overnight in X-gal staining solution at 30°C. Stained sections were washed several times with 1xPBS and mounted in Aquatex medium.

2.4.6. Production of transgenic zebrafish

Male and female zebrafish were separated for one day before the microinjection. On the morning of injection, fish were placed in the same tank for breeding. Once embryos were produced and fertilized, they were collected and arranged in several continuous rows on a

1% agarose plate as previously described (Westerfield 1995). The *egfp* reporter constructs were microinjected at concentration of 100 ng/µl in the presence of phenol red (0.05%) and KCl (100 mM) into single-cell wild-type zebrafish using a Narashigi IM300 microinjector. The microinjection needles were pulled using a P90 micropipette puller. The amount of injected solution was kept constant and corresponded to approximately one tenth of the volume of an embryo as visualized by red color. Embryos were raised in Petri dishes at 28.5°C and *egfp* expression in developing zebrafish embryos was monitored by direct observation under a fluorescent microscope (SMZ 1500, Nikon; Camera: Retiga 1300, Q-Imaging). Zebrafish embryos with at least two *egfp*-expressing cells were referred to as primary transgenic embryos.

2.5. In situ hybridization experiments

2.5.1. Whole-mount in situ hybridization for zebrafish *hoxb5a* and *hoxb5b*

Antisense riboprobes to detect *hoxb5a* and *hoxb5b* transcripts were made as described from plasmids provided by Ashley Bruce (Bruce et al. 2001). An antisense probe for *hoxb5a* was synthesized from pGEM-T plasmid containing the 3'UTR of *hoxb5a* (~550 bp long) using the SP6 RNA polymerase following digestion with SstII. A *hoxb5b* probe was made from pBlueScript plasmid containing a 600bp cDNA encompassing the first exon of *hoxb5b*. Whole mount in situ hybridizations were performed as described (Thisse et al. 1993). Briefly, embryos were dechorionated, fixed in 4% paraformaldehyde, 1x PBS overnight at +4°C and dehydrated in methanol at -20 °C. Embryos were then

gradually rehydrated in a series of methanol/PBS dilutions with increasing concentration of PBS and were then transferred in PBST (1xPBS, 0.1% Tween 20). Rehydrated embryos were treated with Proteinase K (10 μg/ml) for several minutes depending on a stage of development. Embryos were postfixed at room temperature in 4% paraformaldehyde, 1xPBS for 20 min, rinsed in PBT 5 x 5 minutes and then prehybridized for 1-3 hours at 65°C in PreHyb solution (50% formamide, 5x SSC, 0.1% Tween20, 5 mg/ml salmon sperm DNA, 50 μg/ml heparin). Hybridization was done in the same buffer (PreHyb solution) containing 50 ng of DIG-labeled riboprobe overnight at 65°C. The next morning, embryos were successively washed at 65°C for 10 minutes in the following solutions: 100 Hyb solution; 75% Hyb: 25% 2xSSC; 50% Hyb: 50% 2xSSC; 25% Hyb: 75% SSC; 100% 2xSSC. Embryos were then washed twice for 30 minutes in 0.2xSSC at 60 °C. Further washes were performed at room temperature for 5 minutes in the following solutions: 75% 0.2xSSC: 25% PBT; 50% 0.2xSSC: 50% PBT; 25% 0.2xSSC: 75% PBT; 100% PBT.

The embryos were blocked for one hour in 2 mg/ml BSA, 2% Calf Serum, 1% DMSO and incubated overnight at 4°C with the preabsorbed anti-DIG at 1/5000 dilution. Finally, embryos were washed 6 x15 minutes in PBT at room temperature. Colourimetric reaction was performed with staining buffer (100 mM Tris, pH 9.5, 50 mM $MgCl_2$, 100 mM NaCl, 0.1% Tween 20) in the presence of BCIP and NBT (at final concentrations of 175μg/ml and 507μg/ml respectively). Once the color developed, the reaction was stopped in 1x PBS, embryos were postfixed in 4% paraformaldehyde, 1x PBS overnight at +4 °C, washed in PBS and stored in PBS containing 10 mM EDTA.

2.5.2. Radioactive in situ hybridization for *Hoxb5* on cryosections of mouse embryos

An antisense riboprobe for mouse *Hoxb5* was made as previously described from PGEM2-based construct that contains a ~600bp fragment of *Hoxb5* cDNA which corresponds to the N-terminal part of Hoxb5 protein (Krumlauf et al. 1987). This construct was linearized by *EcoR*I digestion and an antisense probe was synthesized with T7 RNA polymerase.

The in situ hybridization experiments were done as previously described (Jaffe et al. 1990). Embryos were obtained from wild type mice at E12.5 and were fixed overnight in 4% paraformaldehyde at 4°C. The next morning, embryos were transferred in 1xPBS and then washed through successive changes of 80% saline, 0.85% saline/ethanol (1:1) and 70% ethanol. Embryos were then dehydrated by treatment with graded alcohol and toluene mixtures, embedded in paraffin wax and sectioned at a 8 μm thickness. Immediately before hybridization, sections were dewaxed, slide sections were rehydrated, transferred through graded alcohols, washed in (0.85% saline and 1xPBS) and postfixed in 4% paraformaldehyde. The slides were then treated with proteinase K (20μg/ml), refixed and treated with triethanolamine and 0.25% acetic anhydride. Once more, sections were washed in (0.85% saline and 1xPBS) and dehydrated.

Slides were prehybridised at room temperature for more than 2 hours in the prehybridization solution (4xSET, 0.024XSET, 0.02% Ficoll-400, 0.02% polyvinylpyrrolidone, 0.1% BSA fraction V, 500 pg/ml sheared denatured salmon sperm DNA, 600 pg/ml yeast total RNA, and 50% deionized formamide). Slides were then hybridized in hybridization solution containing 4XSET, 0.02% Ficoll-400. 0.02%

polyvinylpyrrolidone. 0.1% BSA fraction V. 100 pgjml sheared denatured, salmon sperm DNA, 10% dextran sulfate. 0.1% SDS, 10 mM dithiothreitol, 50% deionized formamide and [35S]UTP labelled probe at a concentration of 50 ng/ml. Hybridization was done overnight at 50°C. Next morning, slides were washed at 50°C in 50% formamide, 1x SSC, 10 mM dithiothreitol for 30 min and were then washed at room temperature in 0.5 x SSC for 30 minutes. Finally, slides were treated with RNAseA (20 μg/ml) for 30 minutes at room temperature, washed in 0.1x SSC for 2 hours at 65°C and dehydrated by passing through a graded series of alcohol. The slides were then exposed to X-ray film overnight, dipped in Kodak NTB-2 emulsion and autoradiaographed for 3 to 6 days at 4°C.

2.6. Quantitative RT-PCR

One hundred transient transgenic zebrafish embryos were generated for each of the *hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp* and *hoxb5b*insJ3*egfp* constructs. Transgenic embryos from each group were then randomly divided in three pools of 33 embryos. Total RNA was extracted from each pool of collected embryos (3x4=12 pools total) using RNAeasy kit (Qiagen) and cDNA synthesis was carried out using oligodT primers (Invitrogen). Relative amounts of the *egfp*, *hoxb5a* and *hoxb5b* transcripts within each group of transient transgenics were estimated by PCR amplification of ~150 bp regions of the corresponding genes using QuantiTect SYBR Green PCR kit (Qiagen) on Mx4000 system. Relative quantification of the *egfp*, *hoxb5a* and *hoxb5b* cDNA targets were performed using standard curves inferred from quantitative PCR results on five different concentrations of the standard (1/1000 μl, 1/100 μl, 1/10 μl, 1 μl and 4 μl of *hoxb5aegfp*

cDNA sample) for each amplicon. Relative copy numbers of *egfp* transcripts were then quantified in each group of transgenic embryos as normalised for endogenous *hoxb5a* gene expression.

3. Results

3.1. Comparative expression analysis of the duplicated *hoxb5a* and *hoxb5b* genes in zebrafish and of the single *Hoxb5* gene in the mouse

In order to facilitate comparison of the expression patterns of the duplicated *hoxb5* genes in zebrafish to that of the single mouse gene, we conducted in situ hybridizations for zebrafish *hoxb5a* and *hoxb5b* and mouse *Hoxb5* at developmental stages at which these genes have been previously detected by Bruce et al and Hogan et al respectively (Hogan et al. 1988; Bruce et al. 2001). Expression of *hoxb5a* and *hoxb5b* was analyzed on whole mount zebrafish embryos at 24, 48 and 72 hpf using DIG-labeled probes and mouse *Hoxb5* expression was examined on sagittal sections of E12.5 mouse embryos using a radioactive probe.

3.1.1 *hoxb5a* and *hoxb5b* in situ hybridization on whole mount zebrafish embryos

Bruce and colleagues previously reported that the duplicated *hoxb5* genes in zebrafish are expressed in overlapping yet distinct domains (Bruce et al. 2001). This study mainly described the expression of *hoxb5a* and *hoxb5b* at early developmental stages (3 somites – 24 hpf). In an attempt to obtain a more complete understanding of the functional

diversification of duplicated *hoxb5* genes, we examined the expression of these genes at later stages (24hpf, 48 hpf and 72 hpf).

Our experiments confirmed the more posterior boundary of *hoxb5b* expression when compared to *hoxb5a* at 24 hpf, as previously described by Bruce et al (Bruce et al. 2001; Figure 3.1, compare panels A and B, arrows). Later in development (48 hpf and 72 hpf), the anterior expression boundaries are indistinguishable for the two zebrafish paralogs (Figure 3.1, compare panels: C, D to E, F, arrows). Consistent with the previous data, both *hoxb5* genes in zebrafish were highly expressed in the embryonic neural tube immediately posterior to the hindbrain, similar to that reported for the mouse *Hoxb5* gene (Hogan et al. 1988). At 48 hpf, only *hoxb5a* is expressed in the somites and branchial arches (Figure 3.1I) while expression of *hoxb5b* is observed in the anterior portion of the pronephric duct and the cranial nerves at the level of the trigeminal ganglion associated with the rostral lateral line (Figure 3.1J). Interestingly, Bruce et al. have demonstrated that only *hoxb5b* is expressed in the posterior lateral ganglia of the developing zebrafish embryo (Bruce et al. 2001). Combined, these data suggest that at early embryonic stages, only one of the two zebrafish *hoxb5* paralogs, *hoxb5b*, is expressed in cells of developing peripheral nervous system in zebrafish.

At 72 hpf, expression of *hoxb5a* persisted in the neural tube, somites and branchial arches; a hybridization signal of lower intensity was also detected in the blood cells and the heart atrium (Figure 3.1O; data not shown). At the same stage, *hoxb5b* was expressed in the neural tube and posterior part of the heart atrium (Figure 3.1P). Interestingly, low levels of *hoxb5b* expression were detected in the branchial arches and somites of zebrafish embryos at 48 and 72 hpf. In general, the staining pattern of the riboprobe used

to detect *hoxb5a* expression was more consistent and extensive than the one obtained with *hoxb5b* specific riboprobe.

Our experiments confirmed that zebrafish *hoxb5* paralogs possess domains of exclusive expression. For example, only *hoxb5b* was found to be expressed in the anterior portion of the pronephric duct and the cranial nerves at the level of trigeminal ganglion, a structure associated with the rostral lateral line system in zebrafish. In addition, *hoxb5b* expression has also been detected in the posterior lateral line by Bruce et al (Bruce et al. 2001). Taken together, these data indicate that the expression of the *hoxb5b* gene is likely to be associated with some structures of the peripheral nervous system. Furthermore, our results indicate variability in the levels of expression in overlapping domains of *hoxb5a* and *hoxb5b* expression. For instance, while strong levels of *hoxb5a* expression were detected in the branchial arches and somites, only weak *hoxb5b* expression was detected in these tissues.

3.1.2. *Hoxb5* in situ hybridization on mouse cryosections

Previous in situ hybridization studies in mouse embryos showed *Hoxb5* expression in the spinal cord and posterior hindbrain. *Hoxb5* expression has also been recorded in the somites, the foregut, the kidney and the lungs (Hogan 1986; Krumlauf et al. 1987). However, to date, data on *Hoxb5* expression in mouse embryos has remained scarce and incomplete. We have conducted in situ hybridization experiments using a radioactive riboprobe on saggital sections of mouse embryos at E12.5 when *Hoxb5* is known to be abundantly expressed (Hogan et al. 1988).

Figure 3.1. Spatiotemporal expression patterns of duplicated *hoxb5a* and *hoxb5b* genes in zebrafish.

Detection of *hoxb5a* and *hoxb5b* gene transcripts in whole mount zebrafish embryos by in situ hybridization. "b5a" or "b5b" at the lower-right corner of each panel indicate whether a *hoxb5a* or a *hoxb5b* antisense riboprobe was used. Panels (A), (B), (G), (H) show zebrafish embryos at 24 hpf, (C), (D), (I), (J), (K), (L) – at 48 hpf and panels (E), (F), (M) – (P) at 72 hpf. Panels (A), (B), (I) - (L), (O), (P) are lateral views of zebrafish embryos, anterior to the left; panels (C) – (F), (G), (H), (M), (N) are dorsal views. Scale bars are 100 μm in (G) – (J) and 250 μm in (A) – (F) and (K) – (P). A black arrow points at the neural tube. Panels (I) and (J) show branchial arches ("ba") and nerves derived from trigeminal ganglion ("gn") seen in panels (K) and (L) under higher magnification. "pd" indicates hybridization signal in the pronephric duct in (L), "ha" – in the heart atrium in (O) and (P), "s" marks somites in (I).

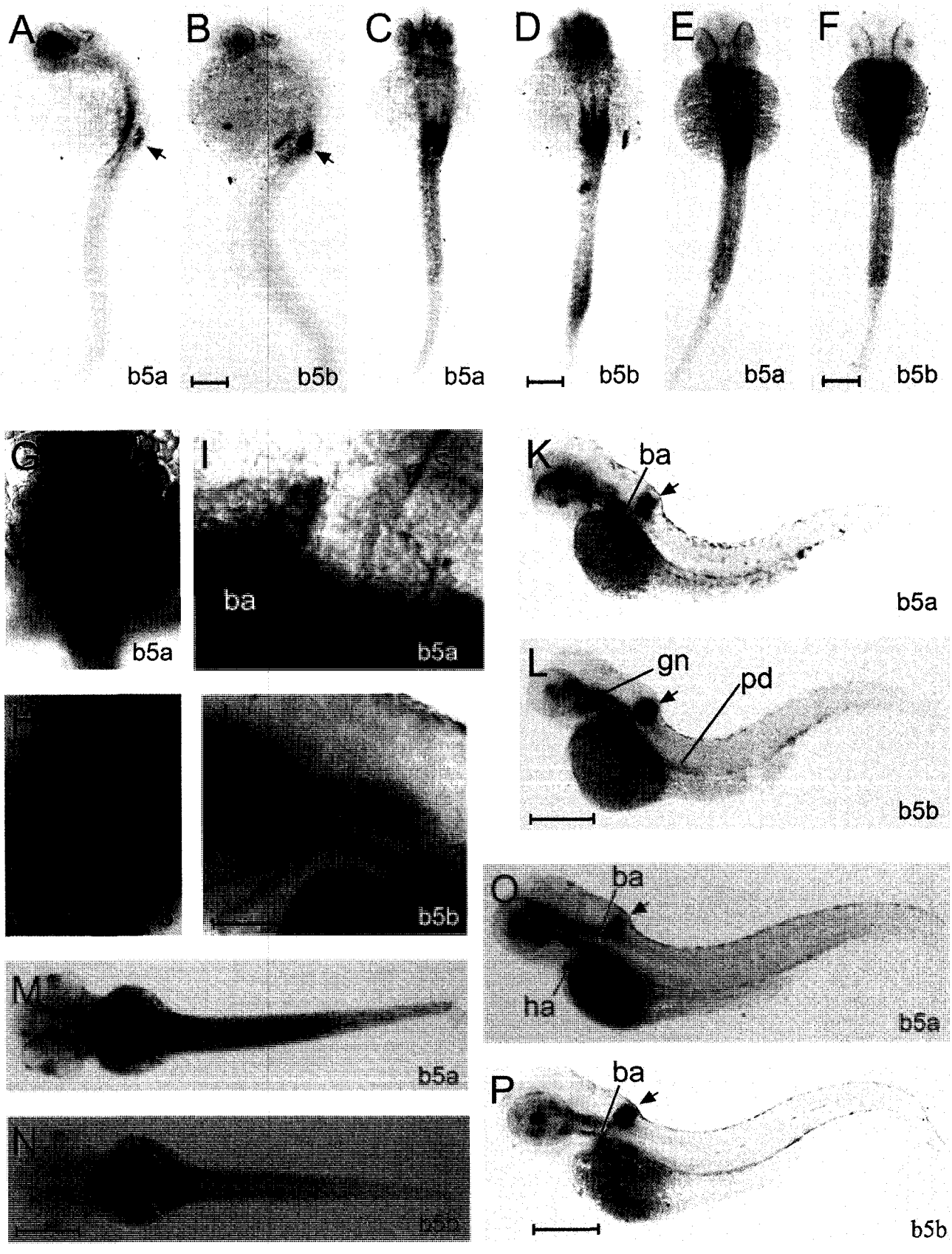
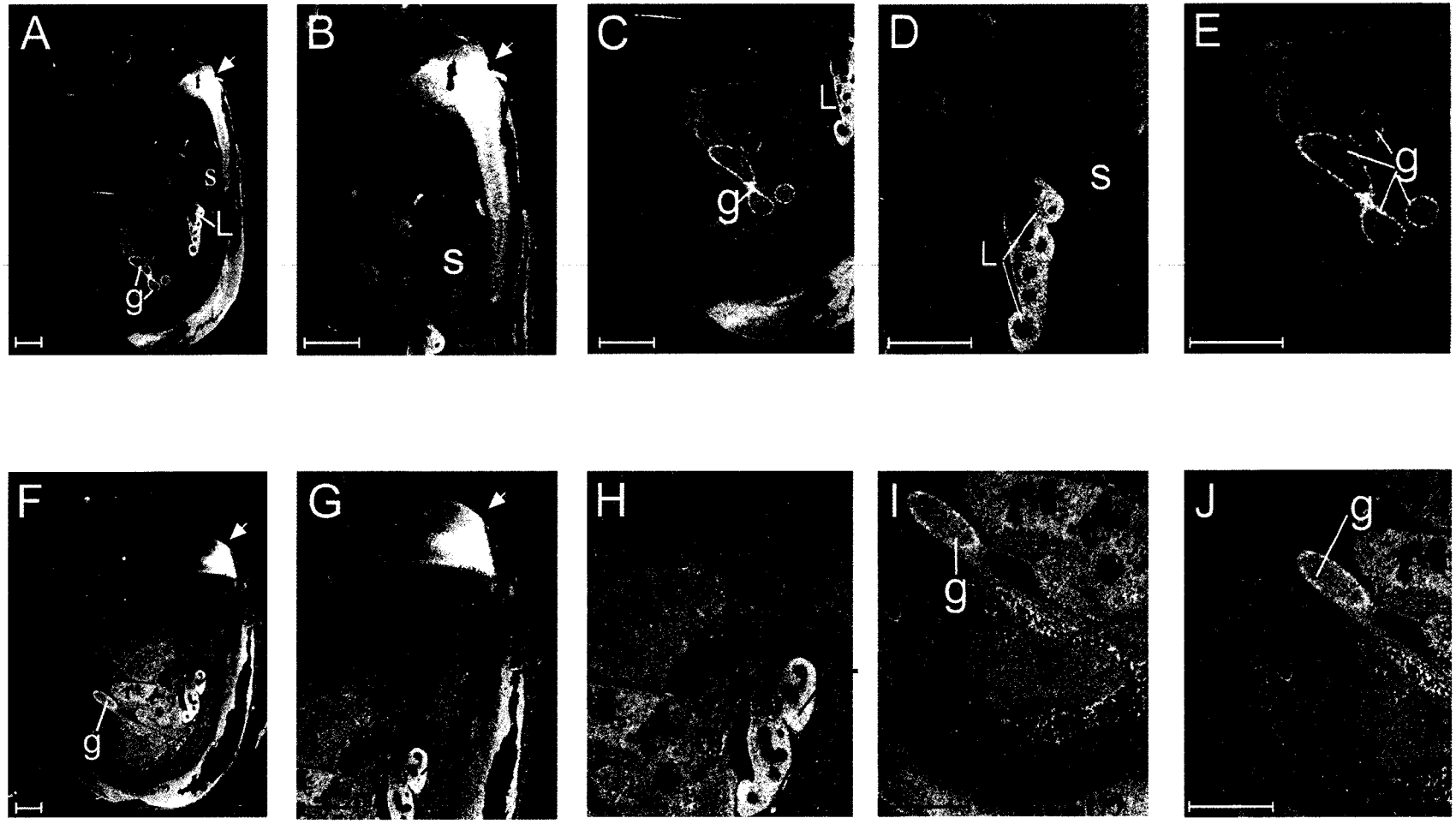

A
B
C
D
E
F
b5a
b5b
b5a
b5b
b5a
b5b
G
b5a
I
ba
b5a
K
ba
b5a
L
gn
pd
b5b
b5b
O
ba
ha
b5a
M
b5a
P
ba
b5b
N
b5b

Figure 3.2. Expression patterns of the *Hoxb5* gene in early murine embryogenesis.

Localization of *Hoxb5* expression by in situ hybridization on sections using an antisense *Hoxb5* riboprobe. All panels show sagittal sections of E12.5 mouse embryos. Panels (A) and (F) illustrate full section views of two stained embryos. Panels (B) – (E) and (G) – (J) display structures seen in the panels (A) and (F) under higher magnification. Scale bars are 1/2 mm for all the panels except 1/4 mm in (H). A strong hybridization signal is detected in developing lungs ("L"), somites ("s"), embryonic neural tube (arrow) and gut ("g").


A
s
L
g
B
s
C
L
g
D
s
L
E
g
F
g
G
H
I
g
J
g

Our results demonstrated the presence of *Hoxb5* transcripts in the somites and along the neural tube, with a particularly high hybridization signal at the level of the spinal cord/hindbrain junction with levels of expression progressively decreasing towards the posterior end (Figure 3.2A). As mentioned above, this domain of *Hoxb5* expression is shared by both zebrafish *hoxb5* paralogs. Along with its regulatory roles in anteroposterior positional specification, *Hoxb5* is speculated to play an important role in organogenesis (Hogan et al. 1988). Consistently, we have detected a radioactive signal in the gut primordium and embryonic lungs (Figure 3.2; data not shown). Noteworthy, mouse *Hoxb5* gene expression was detected in domains that are exclusive for one of the two *hoxb5* paralogs in zebrafish. As an example, a signal was detected in the somites – an expression domain mostly associated with *hoxb5a* expression in zebrafish - and the nodose ganglion of X cranial nerve, similar to the exclusive expression of *hoxb5b* in the peripheral nervous system in zebrafish (data not shown).

Overall, our results confirmed the findings of Bruce et al. that, although *hoxb5a* and *hoxb5b* genes exhibit highly similar expression patterns, there are differences between expression domains of the two *hoxb5* paralogs in zebrafish. Furthermore, the sum of the *hoxb5a* and *hoxb5b* expression domains in zebrafish corresponds to the single *Hoxb5* gene in mouse. Combined with evidence of the biochemical equivalency of the Hoxb5a and Hoxb5b proteins (Bruce et al. 2001), complimentary domains of *hoxb5* duplicates in zebrafish suggest that the *hoxb5a* and *hoxb5b* genes underwent subfunctionalization after duplication through changes in *cis*-acting regulatory elements.

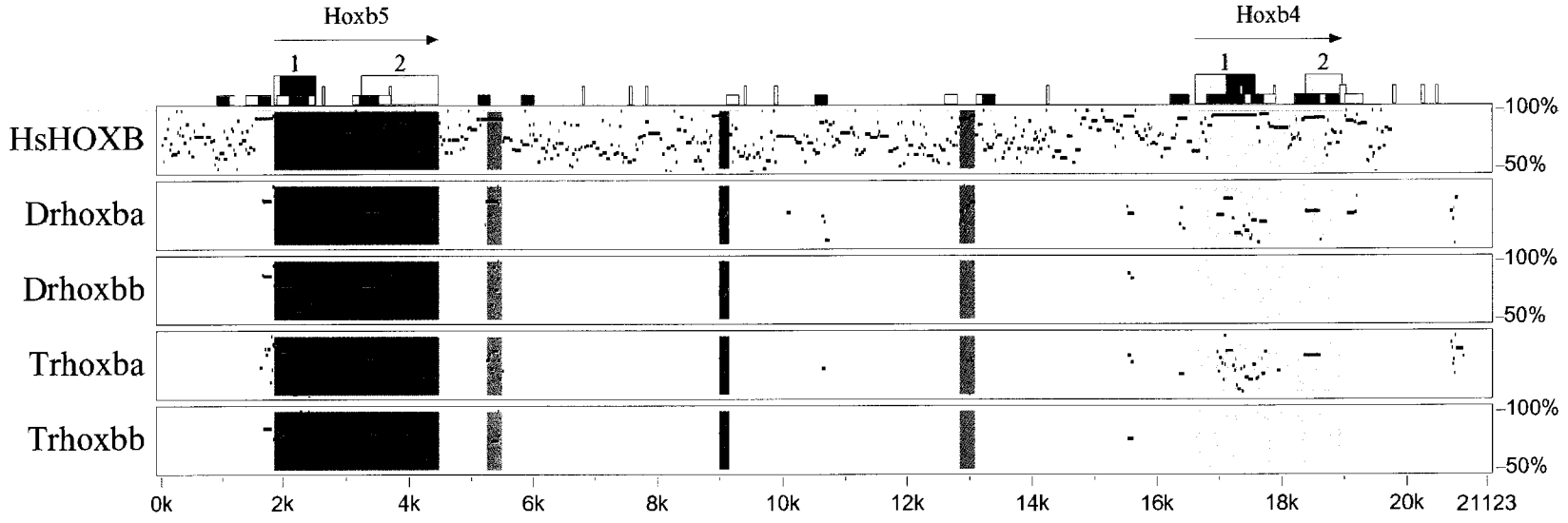
3.2. Identification of putative regulatory elements in the non-coding regions of paralogous group 5 *Hox* genes by phylogenetic footprinting

In order to determine how the regulatory mechanisms that control the expression of the two *hoxb5* genes in zebrafish and *Takifugu* compare with the regulation of the single *Hoxb5* gene of mammals, we carried out a phylogenetic footprinting analysis of the sequences flanking the *hoxb5* transcription units from two mammalian species, mouse and human, and from the two teleosts, zebrafish and *Takifugu*. As mammals and teleosts are separated by approximately 450 million years of evolution, comparisons of orthologous non-coding sequences should allow the identification of conserved regulatory elements since non-functional DNA is randomized.

We identified three conserved sequences, named J1, J2 and J3, located in the 3'-flanking region of *Hoxb5* loci (Figure 3.3). The length of the sequences varied from approximately 120 to 260 bp. In all cases, the regions of high sequence similarity between mouse and human *Hoxb5* genes extended further than when *Hoxb5* sequences from any other species were compared. Both J1 and J2 sequences are present in all compared *Hoxb5* loci. In mouse and human, J1 is located 1.7 kb downstream of the stop codon of *Hoxb5*. In *hoxbb* teleost clusters, the element is localized 1.1 kb away from the termination codon of the *hoxb5* gene. In *hoxba* clusters, J1 is found 1.3 kb and 1.9 kb downstream of *hoxb5a* in zebrafish and *Takifugu,* respectively. The percentage of sequence identity varied from 56% to 99% for the J1 element, depending on the species compared (Table 3.1).

Figure 3.3. Conserved non-coding elements in paralogous group 5 *Hox* genes of vertebrates.

PipMaker comparison of sequences of *Hoxb5* loci from mouse, human, zebrafish and *Takifugu*. The mouse *Hoxb5* locus is used as a reference and shown on the horizontal axis. The percentage identity to the corresponding sequences from human (Hsu and Look), zebrafish (Dr) and *Takifugu* (Tr) are compared and shown on the vertical axis. Positions of CNEs are indicated by shaded regions: yellow for the 5'-flanking region of *Hoxb5*; red for J1, blue for J2, green for J3, and gray for the exons and introns. A sequence similarity lower than 50% was observed between the J2 elements from *Takifugu hoxb5a* and mouse *Hoxb5* using a different algorithm. Therefore, it is not represented on this PIPMAKER analysis. Annotations corresponding to the reference sequence are indicated along the top of the plot: gene names and orientation, exons (black boxes), simple repeats (vertical rectangles), and CpG islands (white and gray rectangles).


Hoxb5
Hoxb4
1
2
1
2
HsHOXB
Drhoxba
Drhoxbb
Trhoxba
Trhoxbb
100%
50%
100%
50%
100%
50%
0k
2k
4k
6k
8k
10k
12k
14k
16k
18k
20k
21123

The J2 sequences are located further downstream, about 5.5 kb from the stop codon of mouse and human *Hoxb5* genes (Figure 3.3). In the zebrafish and *Takifugu hoxbb* clusters, the J2 element is localized ~2.1 kb past the termination codon of *hoxb5b* genes. The cognate sequences from *hoxba* loci are located 5.5 and 3.9 kb downstream of the genes in zebrafish and *Takifugu*, respectively. Pairwise comparisons of J2 sequences resulted in percentages of identity that varied between 46% and 90%. The J2 sequence from *Takifugu* was the most divergent (Table 3.1). Since the sequence similarity between the J2 sequences from the *Takifugu hoxb5a* and the mouse *Hoxb5* genes was below 50%, the *Takifugu* J2 element was not detected by the PIPMAKER software (Figure 3.3).

A third conserved sequence, named J3, is found further downstream in the 3'-flanking region of the mammalian *Hoxb5* gene and of *hoxb5a* genes of teleosts, but is not present in teleost *hoxb5b* loci (Figure 3.3). The J3 sequence is located approximately 9.5 kb downstream of the stop codon of mouse and human *Hoxb5*. In zebrafish and *Takifugu hoxba* clusters, the distance between *hoxb5a* and the J3 element was 8.2 kb and 8.8 kb, respectively. This element showed percentages of identity that varied from 47% to 91% in pairwise comparisons.

Finally, the immediate 5'-flanking region of *Hoxb5* genes was also well conserved between all species that were examined (69-94%, Table 3.1; Figure 3.3).

We also identified two other conserved regions in the non-coding sequences of *Hoxb5* loci. One of them, located between the J2 and J3 elements, is found in all the mammalian and teleost species examined. Interestingly, this sequence remained conserved in *hoxb5a* teleost loci while it appears to be randomized in *hoxb5b* loci (Figure 3.3). Given the small length of this conserved region (<50bp), we reasoned that it is unlikely to be one of the *Hoxb5* enhancers since enhancers can usually be traced as

Table 3.1. Percentage identity for 5' *Hoxb5* promoter region, J1, J2 and J3 in pairwise sequence comparisons

5' pr%, 129 bp / J1%, 255 bp	*MmHoxb5*	*HsHOXB5*	*Drhoxb5a*	*Drhoxb5b*	*Trhoxb5a*	*Trhoxb5b*
MmHoxb5		94	86.4	79.7	71.4	80.5
HsHOXB5	98.7		79.8	77.2	72.4	76.4
Drhoxb5a	80.8	80.3		93.5	70.1	91
Drhoxb5b	57.9	58.6	64.7		68.6	93.5
Trhoxb5a	73.6	73.9	85.5	61.4		72.8
Trhoxb5b	56.2	57.8	63	60	57.4	

J2%, 151 bp / J3%, 257 bp	*MmHoxb5*	*HsHOXB5*	*Drhoxb5a*	*Drhoxb5b*	*Trhoxb5a*	*Trhoxb5b*
MmHoxb5		90	65.8	65.8	45.7	58.3
HsHOXB5	91		68	68	47	63.8
Drhoxb5a	77.6	72.4		66.9	57.5	69.2
Drhoxb5b	-	-	-		51.7	46.7
Trhoxb5a	48.6	50.4	55.6	-		57
Trhoxb5b	-	-	-	-	-	

stretches of conservation between distantly related species that extends over at least 100 bp (Loots et al. 2000). However, we cannot exclude the possibility that this sequence is a regulatory element, other than enhancer, that is involved in *Hoxb5* regulation.

Another conserved region is located further downstream of the J3 element (Figure 3.3) and represent a previously identified microRNA gene (*mir10a*) that codes for a 22 nt non-coding RNA molecule and is suggested to be involved in regulation of *Hox* expression (Mansfield et al. 2004).

3.3. Generation of reporter constructs that carry large fragments and individual CNEs from *Hoxb5* loci

PIPMAKER analysis revealed that non-coding regions of zebrafish *hoxb5a* and *hoxb5b* genes contain three putative enhancer elements (J1, J2 and J3). Moreover, the immediate 5' flanking regions of the *Hoxb5* loci were found to be highly conserved, suggesting their potential importance in *hoxb5* gene regulation. We reasoned that testing combined and individual regulatory activity of these conserved regions from the *hoxb5a* and *hoxb5b* zebrafish loci should help elucidate changes in the non-coding sequences that mediated functional diversification of the duplicated *hoxb5* genes.

In order to test the collective regulatory activity of the identified CNEs, we generated reporter constructs carrying continuous sequences from the *hoxb5a* and *hoxb5b* loci that contain all the identified conserved non-coding elements. This was done by ET cloning, as described in the "Materials and Methods" section. Briefly, 18,495 bp and 12,169 bp DNA fragments from *hoxb5a* and *hoxb5b* loci, respectively, were first inserted

into a modified chloramphenicol resistance vector (Figure 2.1; Figure 2.2). At the second step of ET cloning, we generated reporter constructs by inserting *lacZ* and *egfp* reporter genes in frame with the coding sequences of *hoxb5a* and *hoxb5b* genes (Figure 3.4, panels: A, C). These constructs are referred to as *hoxb5alacZ* and *hoxb5aegfp,* and *hoxb5blacZ* and *hoxb5begfp* respectively (Figure 3.4, panels: A,C).

As the J3 element is only present in one of the duplicated *hoxb5* loci, *hoxb5a*, while it appears to have degenerated in *hoxb5b*, we hypothesized that it might be involved in unique aspects of *hoxb5a* regulation. To better understand the functional importance of the J3 element, we deleted J3 from reporter constructs containing the *hoxb5a* zebrafish locus by substituting the element with a neomycin resistance gene (Figure 3.4B). The resulting reporter constructs are referred to as *hoxb5a*ΔJ3*lacZ* and *hoxb5a*ΔJ3*egfp*.

In addition, we inserted the J3 element into the *hoxb5blacZ* and *hoxb5begfp* constructs to examine if such insertion would render *hoxb5b* expression more similar to that of *hoxb5a*. The resulting constructs were named *hoxb5b*insJ3*lacZ* and *hoxb5b*insJ3*egfp* (Figure 3.4D).

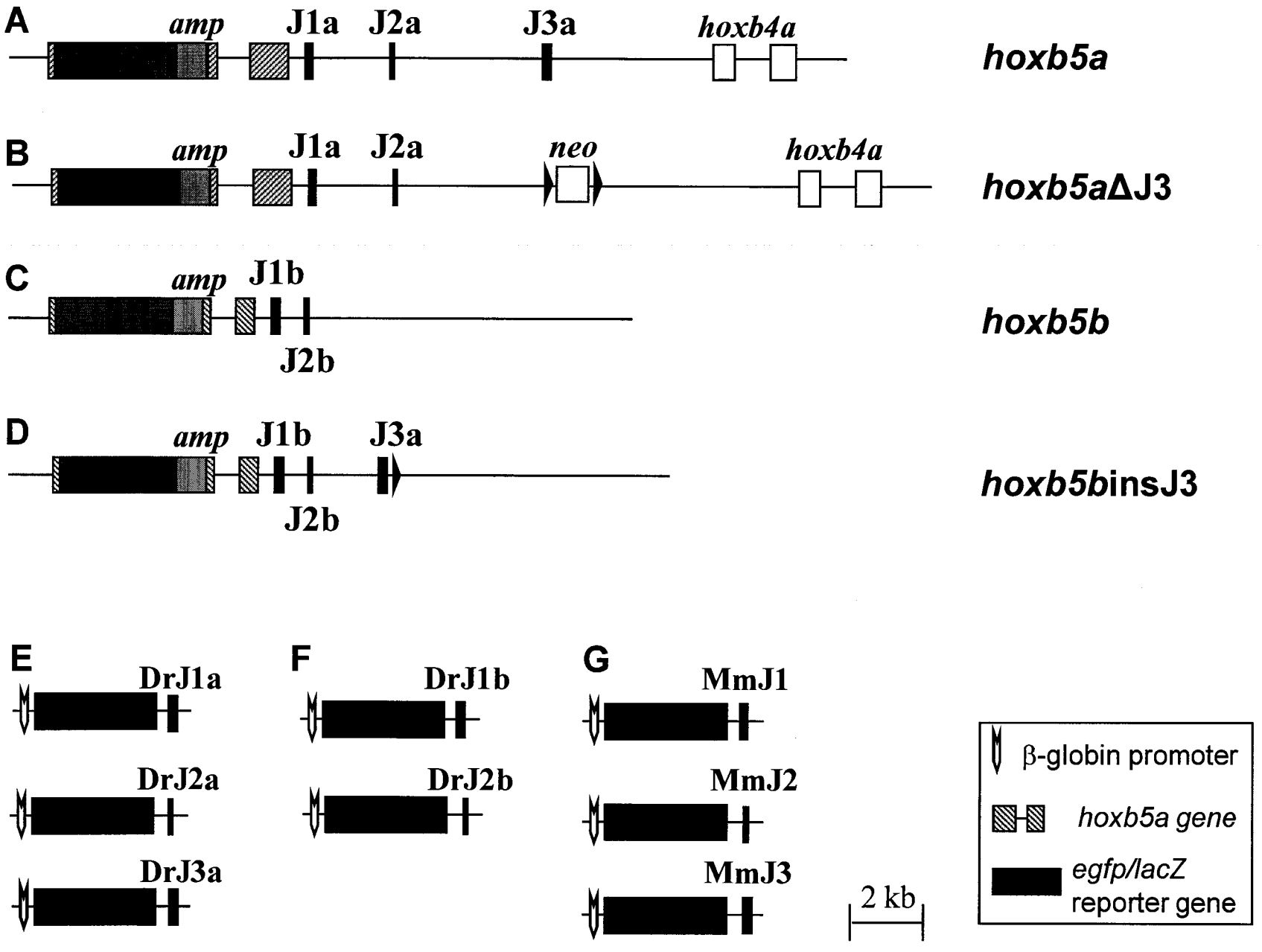
To assess the individual contributions of each of the CNEs to overall *Hoxb5* expression and examine if changes occurred in the paralogous zebrafish sequences were associated with differences in expression of the two zebrafish paralogs, we have also generated multiple reporter constructs carrying individual elements. To do that, we have inserted the individual J1, J2 and J3 elements from the duplicated zebrafish and a single mouse *Hoxb5* loci into the *egfp/lacZ* reporter constructs carrying β-globin minimal promoter. The resulting constructs are referred to as DrJ1a*lacZ/egfp*, DrJ1a*lacZ/egfp*,

Figure 3.4. Schematic representation of *Hoxb5* reporter constructs tested in transgenic assays.

Schematics (A) – (D) depict reporter constructs carrying large regions of zebrafish *hoxb5a* and *hoxb5b* loci. The *hoxb5a* (A) and *hoxb5b* (C) reporter constructs carry all the identified regulatory elements from the corresponding zebrafish loci; the *egfp/lacZ* reporter genes are in frame with the *hoxb5a* and *hoxb5b* coding sequences. Dashed rectangles illustrate exon sequences of the corresponding *hoxb5* genes. Schematic (B) depicts the reporter constructs containing the *hoxb5a* zebrafish locus from which the J3 element has been deleted. (D) shows the reporter constructs containing the *hoxb5b* zebrafish locus in which the J3 element has been inserted.

Schematics (E) – (D) illustrate transgenic constructs carrying individual CNEs from mouse and zebrafish *Hoxb5* loci coupled to a β-globin minimal promoter and the *egfp/lacZ* reporter.

amp – ampicillin resistance gene, *neo* – neomycin resistance gene, black triangles – *lox*P sites, dark rectangles - *lacZ/egfp* reporter genes. Letters a and b specify whether element is from the *hoxba* or from the *hoxbb* zebrafish complex. Mm, *Mus musculus*; Dr, *Danio rerio*. See text for details.


A
amp
J1a
J2a
J3a
hoxb4a
hoxb5a
B
amp
J1a
J2a
neo
hoxb4a
hoxb5aΔJ3
C
amp
J1b
J2b
hoxb5b
D
amp
J1b
J3a
J2b
hoxb5binsJ3
E
DrJ1a
DrJ2a
DrJ3a
F
DrJ1b
DrJ2b
G
MmJ1
MmJ2
MmJ3
2 kb
β-globin promoter
hoxb5a gene
egfp/lacZ
reporter gene

DrJ3a*lacZ/egfp* (Figure 3.4E); DrJ1b*lacZ/egfp*, DrJ1b*lacZ/egfp* (Figure 3.4F); MmJ1*lacZ/egfp*, MmJ2*lacZ/egfp*, MmJ3A*lacZ/egfp* (Figure 3.4G).

3.4. Reporter transgene expression driven by large fragments from the *hoxb5a* and *hoxb5b* zebrafish loci

The engineered reporter constructs carrying large fragments from the *hoxb5a* and *hoxb5b* loci (*hoxb5alacZ/egfp*, *hoxb5blacZ/egfp*, *hoxb5a*ΔJ3*lacZ/egfp* and *hoxb5*insJ3*lacZ/egfp*) were tested in transgenic mice and zebrafish.

3.4.1 Mouse transgenesis

We developed three stable mouse lines that carry the *hoxb5alacZ* transgene and two lines each carrying the *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenes. In each line, we examined reporter gene expression in embryos harvested at E10.5, E12.5 and E13.5. Embryos collected at E10.5 and E12.5 were analyzed via whole mount β-galactosidase staining, while *lacZ* expression in E13.5 embryos was examined on parasagittal sections and through staining of internal organs. The presence of a transgene in mouse founders was first detected by PCR on mouse ear tissue with transgene specific primers and then confirmed by X-gal staining on whole mount embryos harvested from a cross between transgenic males and wild type females. In addition to these lines, we collected two primary embryos transgenic for *hoxb5blacZ* and one primary transgenic embryo for

*hoxb5a*ΔJ3*lacZ*. The primary transgenic embryos were collected and stained with X-gal at E12.5.

All of the *hoxb5lacZ* transgenic embryos obtained (both primary and from a stable transgenic lines) showed *lacZ* expression in regions that coincide with domains of *Hoxb5* expression (compare Figure 3.2 to Figures 3.5 and 3.7). For instance, at E12.5, *hoxb5alacZ* and *hoxb5a*ΔJ3*lacZ* targeted reporter gene expression to prevertebra and the embryonic neural tube (Figure 3.6, panels: A, M) while the *hoxb5blacZ* construct drove *lacZ* expression in the neural tube and the neural crest-derived cranial and dorsal root ganglia (Hogan et al. 1988; Figure 3.6, panels: G,H ; Figure 3.7, panels: L, N).

At E10.5, *hoxb5alacZ* transgenic embryos exhibited strong β-galactosidase activity along the embryonic neural tube with the anterior border of the reporter gene expression reaching the caudal hindbrain (Figure 3.5, panels: A-F, arrows). By E12.5, the anterior boundary of *lacZ* expression shifted more rostrally extending to the caudal limit of the otic vesicle (Figure 3.6, panels: A-F, arrows). This rostral extension of the expression domain is likely mediated by the retinoic acid response element previously characterized by Sharpe and collegues (Sharpe et al. 1998) that resides within the J2 CNE identified in this study. Furthermore, the *hoxb5alacZ* construct consistently targeted the reporter gene expression to the paraxial mesoderm/anterior somites at all the developmental stages examined (Figure 3.5, panels: A-F; Figure 3.6, panels: A-F, asterisks). Detailed analysis of the *hoxb5alacZ* expression pattern on sagittal sections of E13.5 transgenic embryos revealed the presence of β-galactosidase activity in the cartilage primordia of embryonic ribs (Figure 3.7B, "cp").

The *hoxb5a*ΔJ3*lacZ* construct targeted expression of the *lacZ* reporter gene to

Figure 3.5. Expression of the *hoxb5alacZ*, *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenes in mouse embryos at embryonic day 10. 5.

Embryos were collected from crosses between transgenic males and wild type females at E10.5 and examined for β-galactosidase activity using X-gal as substrate. Numbers in the lower right of each panel denote the transgenic line examined. Stars indicate at somites and mesodermal derivities while arrows point to the transgene expression in the embryonic neural tube. "ng" marks staining in the nodose ganglion and "fg" – in the facio-acoustic ganglion. Scale bar: 1/2 mm. Panels (A), (C), (E), (G), (I), (K), (M) are lateral views with anterior to the left and panels (B), (D), (F), (H), (J), (L) are dorsal views. Panels (A) - (F) depict embryos collected from males transgenic for the *hoxb5alacZ* construct, panels (G) - (J) show *hoxb5blacZ* and panels (K) - (N) – *hoxb5a*ΔJ3*lacZ* transgenic embryos.

hoxb5aΔ J3lacZ
hoxb5blacZ
hoxb5alacZ
M
N
7620
7626

Figure 3.6. Expression of the *hoxb5alacZ, hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenes in mouse embryos at embryonic day 12. 5.

Embryos were collected from crosses between transgenic males and wild type females at E12.5 and examined for β–galactosidase activity using X-gal as substrate. Numbers in the lower right indicate the transgenic line numbers examined. Panels (K), (L) show two primary transgenic embryos generated with *hoxb5blacZ* and (Q) and (R) show lateral and dorsal views of a primary transgenic embryo for the *hoxb5a*ΔJ3*lacZ* construct. Scale bars are 1/2 mm. Panels (A), (C), (E), (G), (H), (J), (K), (L), (M), (O), (Q) show lateral views and panels (B), (D), (F), (I), (N), (P), (R) are dorsal views of the transgenic embryos. Stars point at somites and their derivatives while arrows point at the developing neural tube.

hoxb5a∆ J3lacZ
hoxb5blacZ
hoxb5alacZ
M
G
7603
H
7603
N
7629
I
7605
C
7624
J
7605
D
7624
O
pr1
K
pr1
E
7629
R
pr1
L
pr2
F
7626

domains of *Hoxb5* expression highly similar to those driven by its intact counterpart, *hoxb5alacZ*. Indeed, at E10.5, *hoxb5a*ΔJ3*lacZ* transgenic embryos showed β-galactosidase activity in the spinal cord with anterior limits of expression reaching the base of the hindbrain (Figure 3.5, panels: K, L, arrows). Furthermore, similar to the regulatory activity recorded for *hoxb5alacZ*, the *hoxb5a*ΔJ3*lacZ* construct drove expression of the reporter gene to developing somites and their derivatives in transgenic embryos collected at 10.5 and 12.5 days post coitum (Figure 3.5, panels: K , L; Figure 3.6, panels: M-R, asterisks).

While embryonic domains of the *hoxb5alacZ* and the *hoxb5a*ΔJ3*lacZ cis*-regulatory activity extensively overlap, levels of the reporter gene expression targeted by the *hoxb5a*ΔJ3*lacZ* appeared to be significantly diminished compared to that driven by the *hoxb5alacZ* construct (Figure 3.5, compare panels A-F to K-N; Figure 3.6, compare panels A-F to M-R).

At E10.5, mouse embryos transgenic for the *hoxb5blacZ* construct showed *lacZ* expression in the central nervous system with the anterior boundary of the reporter gene expression shifted more posteriorly compared to that observed in the *hoxb5alacZ* and *hoxb5a*ΔJ3*lacZ* transgenic embryos. Indeed, while in E10.5 embryos transgenic for the *hoxb5alacZ* and *hoxb5a*ΔJ3*lacZ* constructs, the most rostral boundary of *lacZ* expression in the embryonic neural tube reaches the caudal hindbrain, it only extends to the caudal limits of the embryonic forelimbs in the *hoxb5blacZ* embryos (Figure 3.5, compare panels A-F to panels G-J, arrows). By E12.5, the rostral boundary of reporter gene expression in the central nervous system driven by the *hoxb5blacZ* construct extends up to the caudal limits of the otic vesicle, similar to that detected in the *hoxb5alacZ* and *hoxb5a*ΔJ3*lacZ*

transgenic embryos at this stage (Figure 3.6, compare panels G-L to A-F and M-R, arrows). In addition to the regulatory activity detected in the CNS, the *hoxb5blacZ* transgenic embryos exhibit β-galactosidase activity in neural crest derivatives such as cranial and dorsal root ganglia and associated nerve fibres. Thus, at E10.5 β-galactosidase activity was detected in the nodose ganglion of the Xth cranial nerve, the facio-acoustic ganglion and dorsal root ganglia in *hoxb5blacZ* transgenic embryos (Figure 3.5, panels: G, H). Expression of *lacZ* in these domains persisted until at least E13.5 as detailed analysis of *hoxb5blacZ* transgenic lines on cryosections of E13.5 mouse embryos has also demonstrated the presence of β-galactosidase activity in X-th cranial nerve and dorsal root ganglia (Figure 3.7, panels: L, N). X-gal staining of internal organs dissected from *hoxb5alacZ* transgenic embryos at E13.5 showed the presence of β-galactosidase activity in the midgut where *lacZ* appeared to be expressed in dotlike fashion (Figure 3.7H; Figure 3.8, panels: A-C, g). The dotlike pattern of expression pattern may reflect *Hoxb5* expression in structures of the enteric nervous system of the intestine similar to that recorded for the *Hoxa5* gene in mouse (Larochelle et al. 1999). In addition, the *hoxb5alacZ* construct targeted reporter gene expression to stomach, meta- and mesonephros and adrenal gland. Similarly, an X-gal signal of lower intensity was detected in the stomach, midgut and adrenal gland in *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenic embryos (Figure 3.8, panels: D-I).

Embryos obtained from founder transgenic animals for the *hoxb5alacZ* construct (lines 7620, 7624, 7626) exhibited stable and consistent patterns of reporter gene expression at all embryonic stages examined (E10.5, E12.5, E13.5) (Figures 3.5-3.7). In contrast, embryos obtained from *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* (lines 7603, 7665, 7629,

Figure 3.7. Analysis of the *hoxb5alacZ*, *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgene expression on sagittal cryosections of mouse embryos.

Embryos were collected from crosses between transgenic males and wild type females at E13.5, sectioned and examined for β–galactosidase activity using X-gal as substrate. All panels show sagittal sections of E13.5 transgenic embryos. Numbers in the lower right indicate transgenic lines examined. Scale bars are 1 mm in panels (A), (C), (F), (I), (K), (N) and 1/3 mm in panels (B), (D), (E), (G), (L), (J), (H), (O). In all panels, arrows point at the neural tube, stars depict previously described mesodermal domain. Panels (A) through (H) show *lacZ* expression observed in the *hoxb5alacZ* transgenic embryos, panel (K) depicts β-galactosidase activity in the *hoxb5a*ΔJ3*lacZ*, panels (I), (J), (L), (M), (O) - in the *hoxb5blacZ* transgenic embryos. Sections of the *hoxb5alacZ* transgenic embryos revealed strong X-gal signal in the cartilage primordium ("cp"), kidneys ("k"), and the gut ("g"). The signal of lower intensity is observed in the lung tissues ("L"). At this stage of embryonic development, the *hoxb5blacZ* construct drives expression in the neural tube (arrow), gut ("g") and Xth cranial nerve exiting trigeminal nucleus ("X").

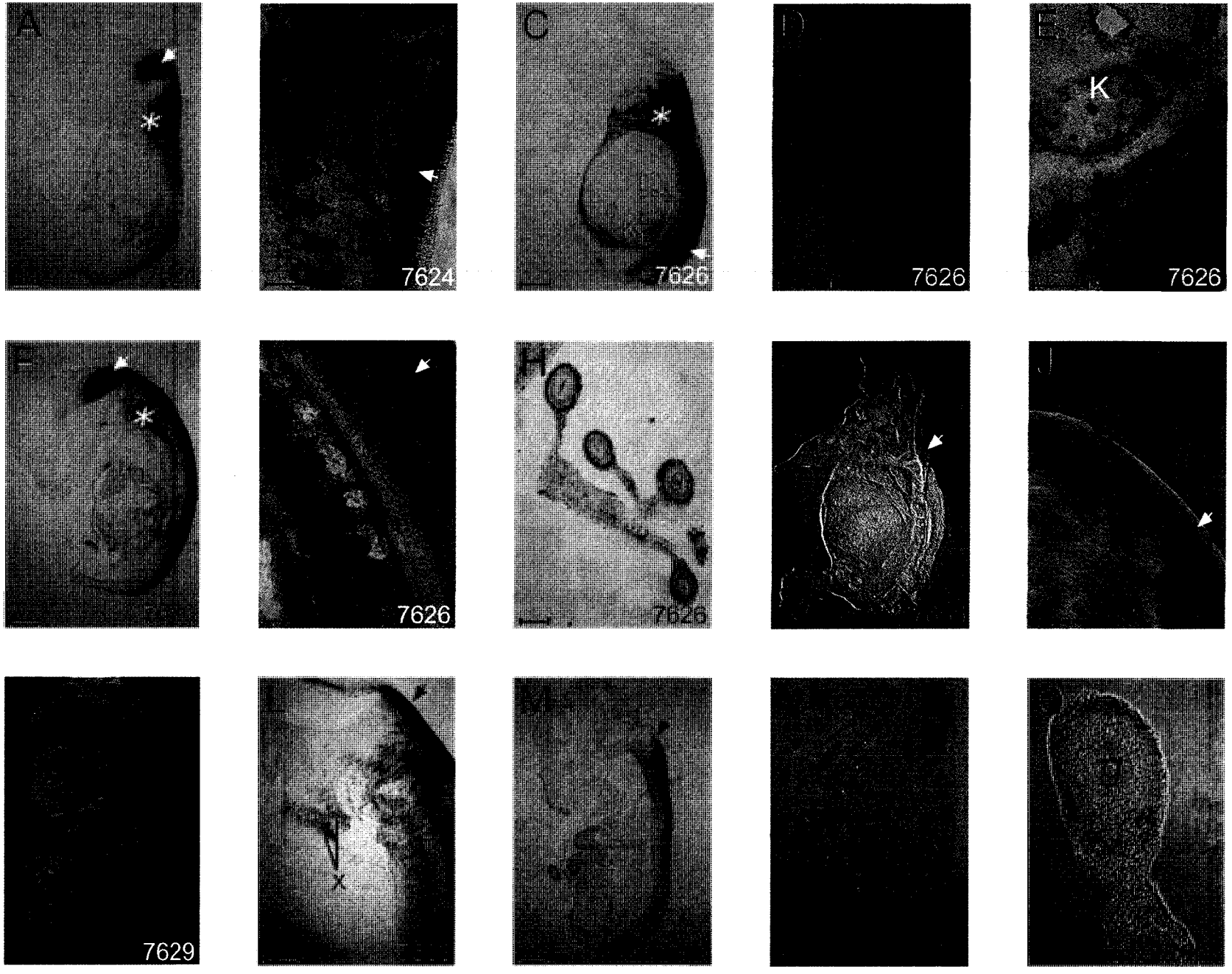

A
7624
C
7626
D
7626
E
K
7626
F
7626
H
7626
I
7629
X

Figure 3.8. Expression of the *hoxb5alacZ*, *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenes in the internal organs of mouse embryos.

Internal organs were dissected from embryos collected from crosses between transgenic males and wild type females and examined for β–galactosidase activity using X-gal substrate. All panels show internal organs extracted from E13.5 transgenic embryos. Scale bar: 1/8 mm in panels (A) – (E), (G), (H); 1/16 mm in panels (F) and (I). Numbers at the lower right corner indicate transgenic line examined. Panels (A), (B), (C) show internal organs isolated from the *hoxb5alacZ* embryos, (D), (E), (F) – from the *hoxb5blacZ* and (G), (H), (I) - from the *hoxb5a*ΔJ3*lacZ* transgenic embryos. "k" depicts expression in kidneys, "g" – the gut and "st" – the stomach of the transgenic embryos.

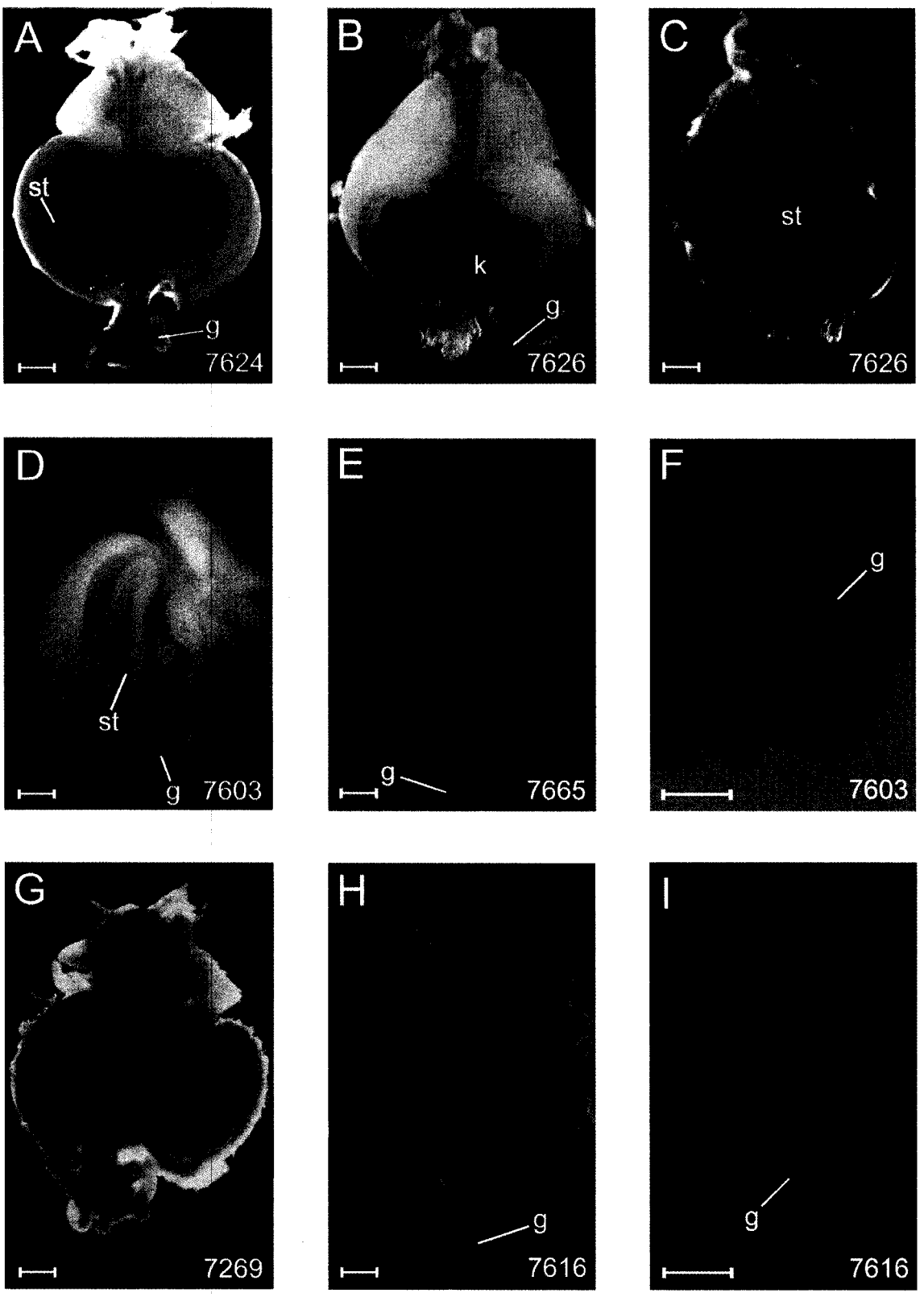

A
st
g
7624
B
k
g
7626
C
st
7626
D
st
g
7603
E
g
7665
F
g
7603
G
7269
H
g
7616
I
g
7616

Figure 3.9. Embryos obtained from the *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* lines show variegated patterns of transgene expression.

Embryos were collected from crosses between transgenic males and wild type females at E12.5 and examined for β-galactosidase activity using X-gal as substrate. Embryos obtained from the same matings of the *hoxb5blacZ* (first row) and the *hoxb5a*ΔJ3*lacZ* (second row) transgenic males with wild type females are shown. Scale bar: 1mm. Panels (A), (C), (E), (G), (I), (K) are lateral views and panels (B), (D), (F), (H), (J), (L) show dorsal views of corresponding transgenic embryos. Arrows in panels (A), (B), (G) and (H) depict weak transgene expression in the neural tube.

A
7603
B
7603
C
7603
D
7603
E
7603
F
7603
hoxb5blacZ
G
7629
H
7629
I
7629
J
7629
K
7629
L
7629
hoxb5aΔ J3lacZ

7616) males showed patterns of transgene expression that differ between embryos collected from the same mating (Figures 3.5-3.7 and Figure 3.9). For these constructs, only embryos with the strongest X-gal staining are shown in Figures 3.5 through 3.8. This phenomenon is meiotically transmissible because the variegated patterns of transgene expression are observed in F1 (obtained from an outcross of transgenic founders to wild type females) and F2 embryos (obtained from an outcross of F1 transgenic males to wild type females).

3.4.2 Zebrafish transgenesis

The *egfp* reporter constructs in which reporter genes were inserted in a region of the *hoxb5a* and *hoxb5b* loci containing the CNEs (*hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp*and *hoxb5*insJ3*egfp*) were injected in one-cell stage zebrafish embryos and transgene expression was monitored under the UV light at 24 hpf, 48hpf and 72 hpf, consistent with the time points previously chosen for *hoxb5a* and *hoxb5b* in situ hybridization experiments.

When tested in zebrafish embryos, *hoxb5aegfp*, *hoxb5begfp* and *hoxb5a*ΔJ3*egfp* constructs recapitulated aspects of the endogenous *hoxb5a* and *hoxb5b* patterns. In most cases, the anterior border of reporter gene expression driven by the *hoxb5aegfp*, *hoxb5begfp* and *hoxb5a*ΔJ3*egfp* constructs corresponded to the posterior hindbrain reflecting rostral restriction of endogenous *hoxb5* gene expression detected by in situ hybridization experiments (compare Figure 3.10, panels A-I to Figure 3,1, panels: A, B, K, L, O, P).

At 24 hours, the *hoxb5aegfp* construct targeted *egfp* expression to cells of the developing central nervous system (Figure 3.10A, arrows) and somites (Figure 3.10A, arrowheads) mirroring mesodermal and neuronal domains of *hoxb5a* expression (Figure 3.1, panels: A, G, I). At the later stages, strong reporter gene expression persisted in the developing CNS cells and was also observed in the muscle cells along the anterior-posterior axis of the transgenic embryos (Figure 3.10, panels: B, C). In addition, at 48 hpf, 10% of transgenic founders injected with the *hoxb5aegfp* construct expressed *egfp* in the heart and the cardinal veins (data not shown).

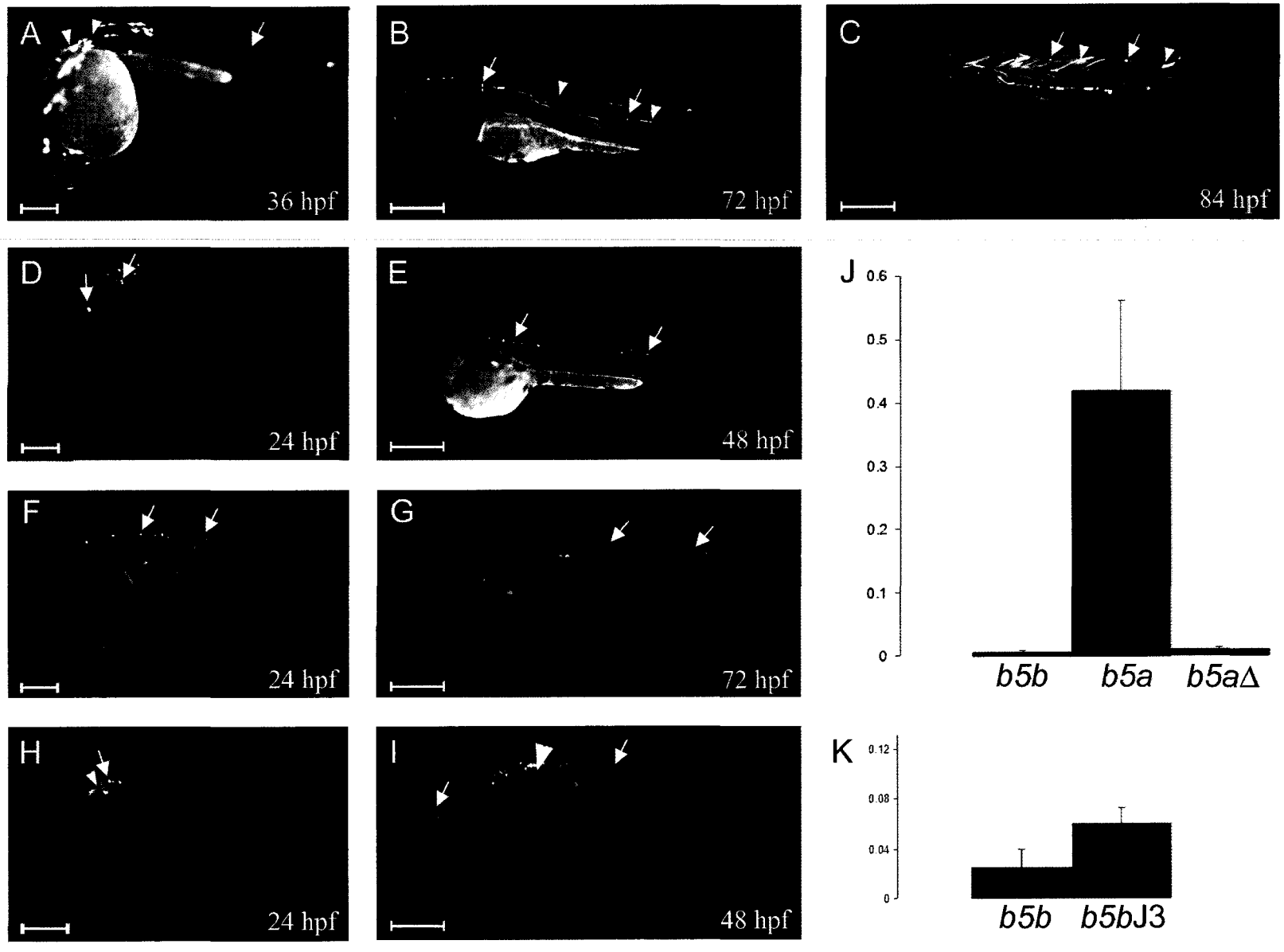
Similar to its regulatory activity in transgenic mouse embryos, the *hoxb5b* construct targeted reporter gene expression preferentially to cells of neuronal origin at all the developmental stages examined (Figure 3.10, panels: D, E). This contrasted with the strong mesodermal activity recorded for the paralogous *hoxb5a* locus. Occasionally, an egfp signal was also observed in cells of hematopoeitic origin in the zebrafish embryos transgenic for *hoxb5begfp* (data not shown).

The *hoxb5a*ΔJ3*egfp* construct exhibited both neuronal and mesodermal regulatory activity similar to that recorded for its intact counterpart, *hoxb5aegfp* (Figure 3.10, panels: F, G; data not shown). However, the reporter gene expression appeared significantly diminished in cells of mesodermal origin of the *hoxb5a*ΔJ3*egfp* transgenic founders compared to that observed in zebrafish embryos injected with the *hoxb5aegfp* construct (data not shown). Consistent with data obtained from mouse transgenesis, levels of reporter gene expression in *hoxb5aegfp* transgenic embryos generally appeared, upon visual inspection, to be higher than those seen in embryos injected with *hoxb5begfp* and *hoxb5a*ΔJ3*egfp* constructs. Indeed, embryos injected with the *hoxb5aegfp* construct had, on average, a greater number of egfp-positive cells than those injected with either

Figure 3.10. Expression of *hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp* and *hoxb5b*insJ3*egfp* transgenes in zebrafish embryos.

Reporter activity of the transgenic constructs was analyzed in primary zebrafish embryos using fluorescent microscopy and quantitative RT-PCR. Panels (A) through (I) show lateral views (anterior to the left) of primary transgenic zebrafish embryos injected with the following constructs: (A), (B), (C) - *hoxb5aegfp*; (D), (E) - *hoxb5begfp*; (F), (G) – *hoxb5a*ΔJ3*egfp*, (H), (I) - *hoxb5b*insJ3*egfp*. Panel A shows a zebrafish embryo at 36 hpf, (B) and (G) – at 72 hpf, (C) – at 84 hpf, (D), (F), (H) – at 24 hpf, (E) and (I) – at 48 hpf. Arrows point at the cells of the developing nervous system while arrowheads mark somites and muscles cells. Scale bars: 1/2 mm in panels (B), (C), (E), (G); 1/4 mm in panels (A), (D), (F), (H); 1/8 mm in panel (I).

Panels (J) and (K) show results of quantitative assessment of relative copy numbers of *egfp* transcripts detected in *hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp* and *hoxb5b*insJ3*egfp* transgenic embryos via RT-PCR. Column heights represent relative copy numbers of *egfp* transcripts detected in transgenic embryos as normalized for expression of the endogenous *hoxb5a* gene. Each bar represents the average of three groups of transgenics and error bars indicate standard errors on the mean. Panel (J) illustrates quantitative differences in relative *egfp* expression levels detected in *hoxb5aegfp* ***(b5a)*** primary transgenic zebrafish compared to that of embryos injected with *hoxb5begfp* ***(b5b)***, *hoxb5a*ΔJ3*egfp* (***b5a***Δ) constructs. Panel (K) shows comparison between relative levels of *egfp* expression detected in *hoxb5begfp* (***b5b***) and *hoxb5b*insJ3*egfp* (***b5b*J3**) primary transgenic zebrafish.


A
36 hpf
B
72 hpf
C
84 hpf
D
24 hpf
E
48 hpf
F
24 hpf
G
72 hpf
H
24 hpf
I
48 hpf
J
0.6
0.5
0.4
0.3
0.2
0.1
0
b5b
b5a
b5aΔ
K
0.12
0.08
0.04
0
b5b
b5bJ3

hoxb5begfp or *hoxb5a*ΔJ3*egfp* constructs. Furthermore, injections of the *hoxb5aegfp* construct in zebrafish embryos resulted in much larger proportion of *egfp*-positive embryos (defined as embryos containing two or more *egfp* – expressing cells) compared to percentage obtained with the *hoxb5begfp* and *hoxb5a*ΔJ3*egfp*constructs. Thus, approximately 60% of zebrafish embryos injected with the *hoxb5aegfp* construct showed *egfp* expression whereas only 30% of zebrafish embryos injected with *hoxb5begfp* and 37.5% injected with *hoxb5a*ΔJ3*egfp* expressed *egfp*. Intriguingly, injections of the *hoxb5b*insJ3*egfp* construct that carries the *hoxb5b* locus in which the J3 element was inserted, produced larger numbers of *egfp*- positive embryos (~50%) with visibly stronger expression levels of the reporter gene expression (Figure 3.10, panels: H, I).

Combined, the results obtained from mouse and zebrafish transgenesis experiments demonstrate the DNA sequences included in the constructs are able to recapitulate most, if not all, of the domains of the *hoxb5a* and *hoxb5b* expression. Furthermore, the differences in regulatory activities exhibited by the *hoxb5a* and *hoxb5b* transgenic constructs appear to be associated with those aspects of expression that differ between paralogs. For instance, one of the most striking differences observed between expression patterns of *hoxb5a* and *hoxb5b* expression in zebrafish is the exclusive expression of *hoxb5b* in structures of the peripheral nervous system, such as the nerve ganglia, and high expression levels of *hoxb5a* in the somites. Consistently, our data indicate that the *hoxb5blacZ* construct targeted reporter gene expression to the facio-acoustic ganglion, nodose ganglion, dorsal root ganglia and associated nerve fibres (Figure 3.5G; Figure 3.7, panels: L, N) while the *hoxb5aegfp* and *hoxb5alacZ* constructs drive reporter gene expression in the somites (Figure 3.5, panels A-F, asterisks; Figure 3.10A, arrowheads) and their derivatives (Figure 3.7B, "cp"; Figure 3.10C, arrowheads).

Combined, these data suggest that changes occurred within the DNA sequences included in our transgenic constructs are responsible for the differences in expression of the *hoxb5* paralogs. These results confirm the DDC model prediction which states that resolution of functional redundancy commonly occurs through subfunctionalization of duplicated genes through changes in their regulatory elements.

3.4.3. Quantitative RT-PCR

Our transgenesis experiments in both mouse and zebrafish gave us indications that levels of reporter gene expression could be consistently lower in animals carrying the *hoxb5b* and *hoxb5a*ΔJ3 transgenes compared to that those with the *hoxb5a* transgenes. The apparently lower reporter gene expression could be due to the absence of the J3 element in the *hoxb5b* and *hoxb5a*ΔJ3 loci as J3 insertion *hoxb5b* loci resulted in increase of *egfp* expression levels in *hoxb5b*insJ3 transgenic embryos. To test this hypothesis, we performed real-time quantitative RT-PCR on RNA extracted from 100 primary transgenic embryos for each of the *hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp* and *hoxb5b*insJ3*egfp* constructs. For these experiments, we generated primary transgenic embryos with each construct. The relative number of *egfp* transcripts in each group of injected embryos was estimated and normalized to the endogenous *hoxb5a* transcript levels. As shown in Figure 3.10J, *hoxb5aegfp* transgenics showed a 104-fold increase in relative copy number of *egfp* transcripts while only a 2.5 fold increase was detected in *hoxb5a*ΔJ3*egfp* transgenic embryos compared to that detected in *hoxb5begfp* transgenic embryos. The detected decrease in levels of the reporter gene expression driven by the *hoxb5a*ΔJ3*egfp* construct compared to that produced by the intact *hoxb5aegfp* locus is likely attributed to the

absence of the J3 element. Combined with data from mouse transgenesis, these results suggest that the presence of the J3 element is necessary for maintaining high *hoxb5* expression levels. We then tested the activity of the *hoxb5b*insJ3*egfp* construct to examine if insertion of this element in the *hoxb5begfp* locus would be sufficient to render higher levels of the reporter gene expression. We have detected 2.4-fold increase in reporter gene expression levels in the *hoxb5b*insJ3*egfp* primary transgenic embryos compared to that produced in embryos injected with the *hoxb5begfp* construct (Figure 3.10K). Taken together, these results suggest that the J3 element is both necessary and sufficient to maintain higher levels of *hoxb5* gene expression.

3.5. Reporter transgene analysis of individual non-coding elements near the *Hoxb5* genes

3.5.1. Mouse and zebrafish transgenesis

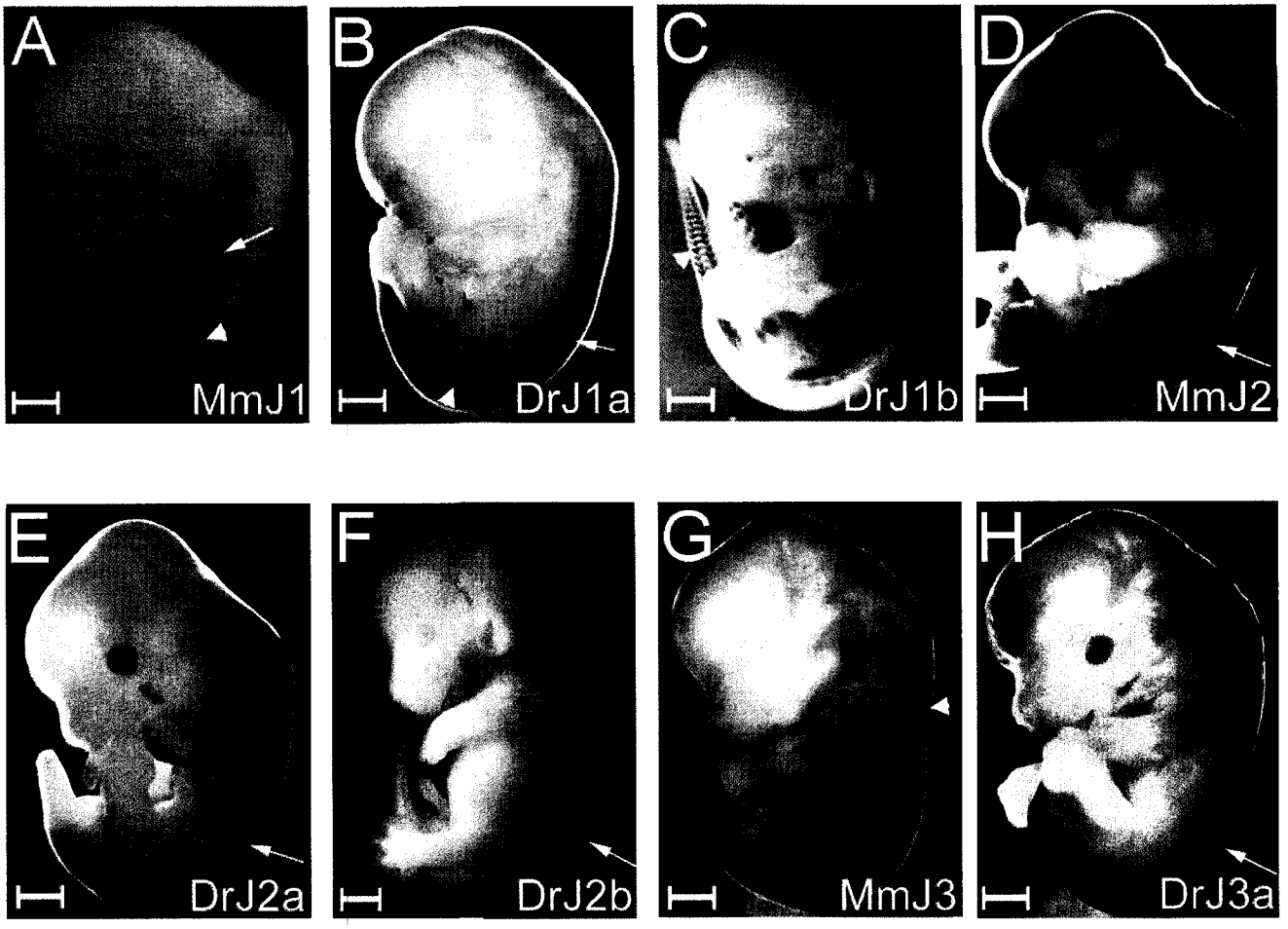
When tested in transgenic animals, large DNA fragments encompassing non-coding regions from the *hoxb5a* and *hoxb5b* loci were able to recapitulate both overlapping and unique aspects of expression of the *hoxb5a* and *hoxb5b* genes. These results suggest that changes at the level of regulatory elements may indeed be responsible for the expressional divergence of the two paralogs. Phylogenetic footprinting of *Hoxb5* loci revealed the presence of three highly conserved sequences, J1, J2 and J3, residing in the 3'-flanking region of the *Hoxb5* genes. Next, we wanted to assess the individual contributions of each of the CNEs to overall *Hoxb5* expression and examine if changes occurred in the paralogous zebrafish sequences were associated with differences in

expression of the two zebrafish paralogs. To do that, we generated multiple *egfp/lacZ* reporter constructs containing a β-globin minimal promoter coupled to the individual J1, J2 and J3 elements from mouse and from the duplicated zebrafish *Hoxb5* loci (Figure 3.4 E, F, G) and tested their activity in transgenic animals.

Transgenic F0 mouse embryos were examined at E13.5. Transgenic zebrafish embryos were systematically examined at 1, 2, 3 and, occasionally, 4 dpf to cover the main time points for which expression patterns of *hoxb5a* and *hoxb5b* genes were characterized (Bruce et al., 2001 and Figure 3.1). When tested in transgenic mouse embryos, each of the conserved sequences from the mouse *Hoxb5* locus was able to target reporter transgene expression to endogenous domains of *Hoxb5* expression (Sakach and Safaei 1996; Oosterveen et al. 2003). For instance, a transgene containing the mouse J1 element (MmJ1), targeted *lacZ* expression to structures of the paraxial mesoderm (Figure 3.11A) and neural tube (Table 3.2; data not shown). A detailed examination of the MmJ1 transgenic embryos also revealed ß-galactosidase activity in the notochord and dorsal root ganglia (Table 3.2, DRG; data not shown). A similar construct containing the mouse J2 element, MmJ2, induced strong reporter gene expression in the neural tube while a signal of lower intensity was observed in prevertebra along the entire anterior-posterior axis (Table 3.2; data not shown). Finally, the construct containing mouse J3 (MmJ3) targeted *lacZ* expression to the neural tube, developing vertebrae and posterior somites (Figure 3.11G; Table 3.2). The mouse J2 and J3 sequences were also able to drive *lacZ* expression in DRG but with lower intensity (Table 3.2). These data confirmed that the identified CNEs behaved as *cis*-acting regulatory elements and provided us with a means

Figure 3.11. Enhancer activity of conserved non-coding elements of *Hoxb5* loci in transgenic mouse embryos.

Reporter activity of the transgenic constructs was analyzed in primary transgenic mouse embryos. The age of the embryos varied from E12.5 to E14.5. Scale bar: 1 mm. For each panel, the name and origin of the CNE tested in the reporter construct is indicated at the bottom left corner. Arrowheads mark somite derivatives, while arrows point at the neural tube. Mm, *Mus musculus*; Dr, *Danio rerio*; letters a and b specify whether element is from the *hoxba* or from the *hoxbb* zebrafish complex.


A
MmJ1
B
DrJ1a
C
DrJ1b
D
MmJ2
E
DrJ2a
F
DrJ2b
G
MmJ3
H
DrJ3a

to address the question of whether changes at the level of regulatory elements contributed to functional divergence of teleost *hoxb5* genes.

Comparative functional studies of the CNEs from zebrafish and mouse *Hoxb5* loci were performed. All transgenic zebrafish embryos obtained with MmJ1 showed *egfp* expression in cells of mesodermal origin. Twenty percent of them also showed *egfp* expression in neural cells (Figure 3.12C; Table 3.2; data not shown) indicating a spectrum of activities comparable to that observed in mice (Figure 3.11A; Table 3.2). As well, MmJ2 drove *egfp* expression almost exclusively to cells of mesodermal origin in the transgenic fish obtained (Table 3.2; data not shown). This contrasted with the strong neural activity detected in transgenic mice (Figure 3.11D; Table 3.2). When tested in mouse embryos, the zebrafish J1 element from the *hoxba* cluster, (DrJ1a) induced transgene expression in prevertebrae and in the neural tube (Figure 3.11B; Table 3.2). ß-galactosidase activity was occasionally observed in DRG and/or in nerves issued from DRG (Table 3.2; data not shown). In parallel, DrJ1a transgenic zebrafish embryos expressed *egfp* in cells of mesodermal origin from 1 dpf, in some cases starting at the level of the most anterior somites (Figure 3.12A; data not shown), until 4 dpf, the last point examined (Table 3.2). At later stages, *egfp* expression was observed in muscle cells and lateral plate mesoderm cells (data not shown).

Transgenic mouse embryos produced with the zebrafish J1B (DrJ1b) construct showed β- galactosidase staining in segmented mesodermal structures and in the dorsal part of the neural tube (Figure 3.11C; Table 3.2; data not shown). Besides inducing *egfp* expression in somites of most zebrafish embryos, the DrJ1b element was also able to target reporter gene expression to the neural tube in 30% of transgenic embryos (Figure 3.12B; Table 3.2).

Transgenic mouse embryos generated with the constructs containing the J2 elements from the zebrafish *hoxb* clusters, DrJ2a and DrJ2b, expressed the *lacZ* reporter gene mainly in neural derivatives (Figure 3.11, panels: E, F). When tested in zebrafish embryos, Dr21a and DrJ2b sequences targeted *egfp* expression to muscle cells with a high frequency (Figure 3.12, panels: D, E; Table 3.2). A neural enhancer activity was also observed in transgenic fish: *egfp* expression was observed in the developing spinal cord of 13% and 71% of zebrafish embryos injected with the DrJ2a and DrJ2b constructs, respectively (Figure 3.12E; Table 3.2). The MmJ3 sequence mainly conferred mesodermal expression in transgenic zebrafish: all primary transgenic zebrafish embryos carrying the MmJ3 construct showed a strong *egfp* expression in muscle cells along the anterior-posterior axis while only 2% of them also expressed the transgene in neural tissues (Figure 3.12F; Table 3.2). Mouse embryos produced with the MmJ3 construct showed *lacZ* expression in prevertebrae and posterior somite derivatives (Figure 3.11G; Table 3.2). Transgene expression was also detected in neural structures: two mouse embryos showed β-galactosidase-positive cells in the neural tube and one mouse embryo expressed *lacZ* in the DRG (Table 3.2; data not shown). In mice, DrJ3a exhibited mainly neural enhancer activity (Figure 3.11H; Table 3.2). Similarly to its mouse counterpart, the DrJ3a construct directed transgene expression in muscle cells in most transgenic zebrafish embryos (Figure 3.12G; Table 3.2). ß-galactosidase activity was also detected in DRG and associated nerve fibers in two specimens (data not shown).

Thus, comparative analysis revealed that regulatory activity conferred by the cognate CNEs from duplicated zebrafish loci extensively overlapped and in general, corresponded to expression patterns common to both *hoxb5* orthologs.

Figure 3.12. Enhancer activity of conserved non-coding elements of *Hoxb5* loci in primary transgenic zebrafish.

Reporter activity of the transgenic constructs was analyzed in primary transgenic zebrafish using fluorescent microscopy. The name and origin of the tested CNEs are indicated at the bottom left corner of each panel. In panel (A) through (G), CNEs were coupled to the human ß-globin minimal promoter. Scale bars: 1/2 mm in panels (A)-(C) and 1/8 mm in panels (D) – (G). Cells of the developing central nervous system are indicated by arrows in panels (B), (C), (E). Arrowheads point at somites in panel (A) and at muscle cells in panels (D) - (G). All embryos are in lateral views, anterior to the left. Mm, *Mus musculus*; Dr, *Danio rerio*; letters a and b specify whether element is from the *hoxba* or from the *hoxbb* zebrafish complex.

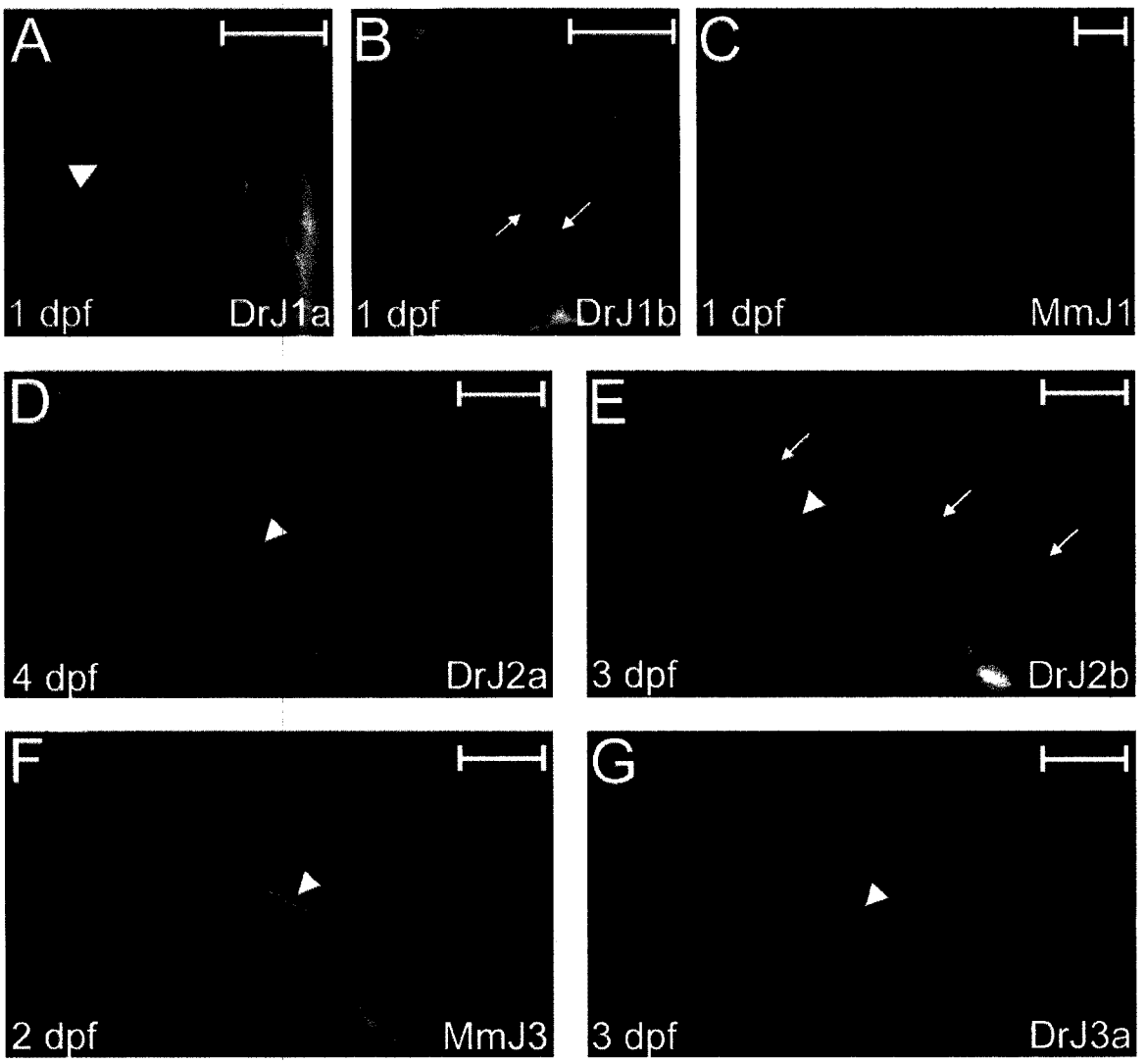

A
1 dpf
DrJ1a
B
1 dpf
DrJ1b
C
1 dpf
MmJ1
D
4 dpf
DrJ2a
E
3 dpf
DrJ2b
F
2 dpf
MmJ3
G
3 dpf
DrJ3a

Table 3.2. Expression of reporter constructs in primary transgenic mouse and zebrafish embryos.

Element tested	Mouse injections				Zebrafish injections	
		Expression sites				
	#expressing/ #Tg	Neural	DRG	somites/ prevertebrae	Neural	Mesodermal
MmJ1	5/5	4	2	4	20% (11/55)	100% (55/55)
DrJ1A	4/7	3	2	2	2% (1/48)	100% (48/48)
DrJ1B	3/10	2	1	2	30% (22/74)	80% (59/74)
MmJ2	3/6	3	2	3	0% (0/42)	100% (42/42)
DrJ2A	2/2	2	1	1	13% (7/54)	93% (50/54)
DrJ2B	4/8	3	1	1	71% (52/73)	67% (49/73)
MmJ3	4/8	2	1	2	2% (1/49)	100% (49/49)
DrJ3A	6/8	5	2	0	4% (2/54)	98% (53/54)

DRG, dorsal root ganglion

3.6. Phylogenetic analysis of *Hoxb5* regulatory sequences

The DDC model predicts that regulatory elements of duplicated genes will undergo rapid complementary degeneration (Force et al. 1999). Consistent with this notion, the J3 element is retained only in *hoxb5a* locus of teleosts. Degeneration of this element in the *hoxb5b* locus has likely contributed to functional divergence of *hoxb5* paralogs. In contrast, the J1, the J2 and the 5'-flanking sequences of *hoxb5* genes are found in both teleost *hoxb* clusters. Despite a few differences observed in the activities of the individual regulatory sequences from *hoxb5a* and *hoxb5b* loci when tested individually, the transgenic analysis of CNEs was not sufficient to reveal clear complementation between regulatory activities of elements from the two zebrafish loci compared to those from the single *Hoxb5* locus in mice. It is still possible that duplicated elements may have split the original regulatory function between them if separate functional units within the elements underwent complementary changes. Phylogenetic analysis of the CNEs was undertaken in order to obtain evidence for their divergence in duplicated teleost genes. We first aligned J1, J2 and 5' flanking region sequences from all *Hoxb5* loci and screened the alignments for the presence of short (6-12 bp) and highly-conserved (>91%) sequences. Such sequences that are evolving at an exceptionally slow rate are likely to be functionally important for the regulatory activity of the elements (Holloway et al. 2005). Although we identified short DNA sequences that were conserved in cognate elements of mammals and teleosts (Table 3), we were unable to find sequences that were present in elements from one of the teleost paralogs (e.g. *hoxb5a* from zebrafish or *Takifugu*), but were divergent in the other teleost paralogs (e.g *hoxb5b*; data not shown). These data argue

Figure 3.13. Phylogenetic analysis of *Hoxb5* loci from human, mouse, zebrafish, and *Takifugu*.

(A) Phylogenetic tree predicted by the DDC model. The tree topology is based on the assumption that *hoxba* and *hoxbb* teleost clusters resulted from a duplication event that happened after the separation of fish and tetrapod lineages and before the divergence of *Takifugu* and zebrafish lineages; (B) – (F) Phylogenetic trees built with MEGA3 based on the analysis of the following sequences: (B) coding sequences of *Hoxb5* genes (850 bp); (C) 5' upstream region of *Hoxb5* genes (130 bp); (D) J1 CNE (255 bp); (E) J2 CNE (150 bp); (F) CNEs present in all the *Hoxb5 loci* (535 bp). 5' upstream region, J1 and J2 were joined in a continuous single sequence for each *Hoxb5* locus. As the J3 CNE was only found in the *hoxb5a* complexes of zebrafish and *Takifugu*, it was not included in this analysis. Hs, human, Mm, mouse, Dr, zebrafish, Tr, *Takifugu*. The scale bar indicates evolutionary distance, calculated using Kimura's two-parameter algorithm.

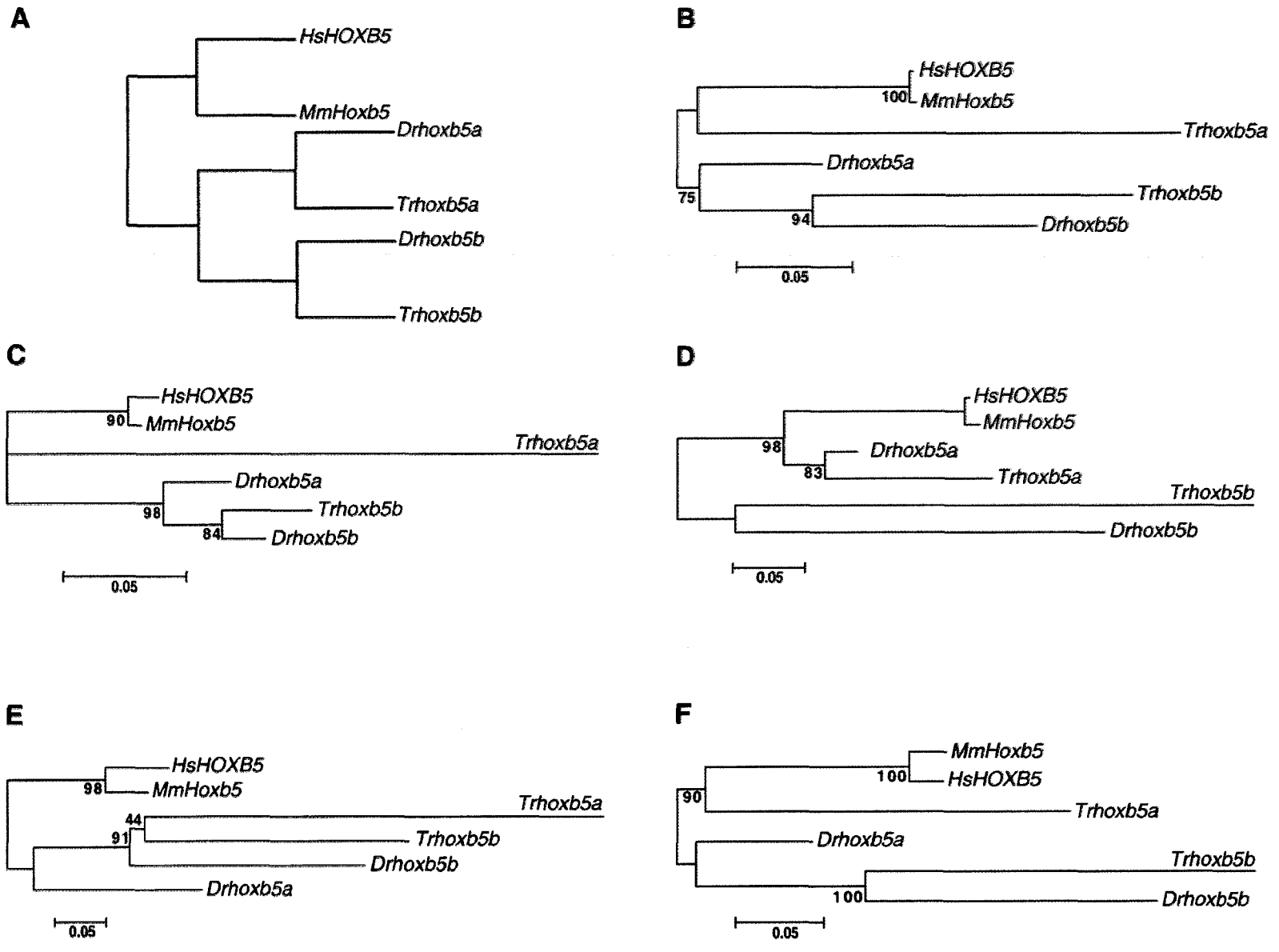

A
HsHOXB5
MmHoxb5
Drhoxb5a
Trhoxb5a
Drhoxb5b
Trhoxb5b
B
HsHOXB5
100
MmHoxb5
Trhoxb5a
Drhoxb5a
75
Trhoxb5b
94
Drhoxb5b
0.05
C
HsHOXB5
90
MmHoxb5
Trhoxb5a
Drhoxb5a
98
Trhoxb5b
84
Drhoxb5b
0.05
D
HsHOXB5
MmHoxb5
98
Drhoxb5a
83
Trhoxb5a
Trhoxb5b
Drhoxb5b
0.05
E
HsHOXB5
98
MmHoxb5
Trhoxb5a
44
91
Trhoxb5b
Drhoxb5b
Drhoxb5a
0.05
F
MmHoxb5
100
HsHOXB5
90
Trhoxb5a
Drhoxb5a
Trhoxb5b
100
Drhoxb5b
0.05

Table 3.3. Potential Transcription Factor Binding Sites identified in the conserved *Hoxb5* non- coding sequences.

Sequence	Transcription factor	CNS	Position on Mm *Hoxb* contig
TTACCT	RARβ	5' region of *hoxb5a*	9883-9888
AGAATTT	c/EBP	5' region of *hoxb5a*	9905-9911
TTTACGA	HoxA9	5' region of *hoxb5a*	9909-9915
CACGTG	Smad 3/c-Myc	5' region of *hoxb5a*	9930-9935
CCATATTTGG	MCM1/SRF	5' region of *hoxb5a*	9973-9982
TGACATT	HoxA9 / Meis-1a	J1	13393-13399
TCGTAAA	HoxA9 / Meis-1a	J1	13439-13445
CACGTG	c-Myc	J1	13492-13497
TTTATGG (-)	HoxA9 / Meis-1a	J1	13502-13508
CATAAAGTG	Pax-2.1	J1	13503-13511
TTTTATGGTTTA	Ubx	J1	13514-13525
TTTATGG	HoxA9 / Meis-1a	J1	13515-13521
TAACTG	c-Myb	J1	13582-13587
ATGAGA	XPF	J2	17207-17212
TGATCC	GR	J2	17232-17237
AGGTCA	RARα1	J2	17242-17248
GGGTGA	RXRα	J3	21094-21099
TGAACC	RXRα	J3	21097-21102
AGGTCA	RARα1	J3	21105-21110
TAAAAT	Pou1F1a	J3	21128-21133
AACAAAG	Lef-1, Sox3, Sox5	J3	21265-21171
AGATTA	GATA1/4	J3	21186-21191

The listed 6-12 bp sequences are highly conserved in all examined species (mouse, human, zebrafish and *Takifugu*) and coincide with known transcription factor binding sites.

against a simple complementary degeneration of regulatory subfunctions within the elements that are present in both teleost *hoxb5* loci.

Sequences common to all *Hoxb5* loci were then scanned using the Patch Search database for putative transcription factor binding sites. Among the transcription factors obtained, some corresponded to transcription factors known to be involved in *Hox* gene regulation. This suggests a likely role for them in enhancer activity of the J elements. For example, one of the conserved regions identified by phylogenetic filtering inside the J2 element corresponded to a RARE shown to be essential for correct anterior expression of both *Hoxb8* and *Hoxb5* genes in the neural tube (Oosterveen et al. 2003). This RARE is 17 bp-long and was found to be highly conserved in human, mouse and zebrafish (Oosterveen et al. 2003). Another putative RARE was present in one of the highly conserved regions of the J3 element (Table 3). The presence of highly conserved motifs within the enhancer sequences from mammalian and both duplicated teleost loci may indicate that some aspects of ancestral *Hoxb5* regulation remained preserved in both *hoxb5a* and *hoxb5b* loci after the duplication event. These data suggest that the "fate" of the *Hoxb5* regulatory elements from the duplicated teleost loci may be different from that predicted by the DDC model.

The DDC model predicts that degeneration of redundant subfunctions should occur within 4- 12.5 million years after the duplication event (Force et al. 1999). Functional specialization of the *hoxb5a* and *hoxb5b* genes in teleosts may have happened early in euteleost evolution and, thus, have preceded the divergence of closely related zebrafish and *Takifugu* lineages. In such a case, one may expect higher sequence conservation between orthologous functional modules (i.e. *hoxb5a* from zebrafish and *hoxb5a* from

Takifugu) than between paralogous modules (i.e. zebrafish *hoxb5a* and *hoxb5b*) (Figure 3.13A).

To test the model, we built phylogenies of *Hoxb5* loci based on the analysis of coding sequences of *Hoxb5* genes from human, mouse, zebrafish and *Takifugu* (Figure 3.13B), as well as sequences from the 5'-flanking regions of *Hoxb5* genes and of the J1 and J2 elements (Figure 3.13, panels: C,E). The tree built for the J1 CNE matched the branching pattern of the hypothetical model (Figure 3.13, compare panels A and D), whereas phylogeny for other functional modules differed (Figure 3.13, compare panel A to panels B, C, E, F). While topologies of trees constructed for the coding sequences (Figure 3.13B) and the 5' flanking regions of *Hoxb5* genes (Figure 3.13C) were similar, a tree built for the J2 CNE (Figure 3.13E) showed a unique branching pattern, suggesting this CNE may be subjected to a selective pressure distinct from that applied upon other functional modules of *Hoxb5* genes. In order to assess the composite pattern of divergence of non-coding functional units from *Hoxb5* loci, sequences of the CNEs were first joined in a continuous sequence for each locus and then analyzed with MEGA3. The topology of the resulting tree (Figure 3.13F) was similar to the one calculated for the coding sequences and the 5' flanking regions of *Hoxb5* genes suggesting that the coding region and most regulatory regions of *Hoxb5* loci are under common selective constraints. Overall, the dynamics of molecular changes involved in the evolution of *hoxb5a* and *hoxb5b* teleost genes differ from those anticipated for duplicated genes that undergo subfunctionalization (Force et al. 1999). Indeed the regulatory elements of *hoxb5a* and *hoxb5b* duplicates appear to diverge slower than anticipated.

In summary, phylogenetic analysis of sequences from the duplicated *hoxb5* loci uncovered significant deviations from dynamics of molecular changes predicted by the

DDC model for duplicated genes that undergo subfunctionalization (Force et al. 1999).

4. Discussion

Gene duplication has long been considered as a major source of evolutionary novelty. Paralogous genes resulting from duplication events have the opportunity to diverge and acquire new specific functions. Divergence in expression patterns of duplicated genes has been proposed to be the first step in the establishment of their functional divergence. The DDC model originally proposed by Force et al. states that subfunctionalization is a common mechanism of preservation of duplicated genes (Force et al. 1999). Mainly based on well reasoned mathematical equations, this model makes the following predictions with regards to molecular mechanisms of subfunctionalization:

1. Subfunctionalization happens by means of changes in the non-coding regions of duplicated genes;

2. Resolution of functional redundancy occurs through complementary degeneration of individual regulatory elements.

3. The process of subfunctionalization should be completed within 12.5 million years after a duplication event.

This model has received a lot of attention in the field of evolutionary developmental biology in the past few years. To date, the fates of several pairs of gene duplicates were investigated in an attempt to bridge the mathematical predictions with biological reality by a number of research groups. Amongst gene duplicates that are thought to undergo subsequent subfunctionalization after the duplication event, as inferred from comparative analysis of the expression patterns of the duplicated genes, are paralogous *Nudt10* and *Nudt11* genes in mammals, *sox9a/sox9b* and *mitf-m/mitf-b* genes in teleost (Altschmied et al. 2002); Hua et al. 2003; (Kluver et al. 2005). Combined, these data confirm that

complementary loss of independent subfunctions (subfunctionalization) may indeed constitute a common mechanism of resolution of functional redundancy between the duplicated genes. However, these studies were not sufficient to provide an insight into the molecular changes that occurred at the level of regulatory elements that instigated subfunctionalization.

In an attempt to clarify whether the dynamics of *cis*-regulatory changes support the DDC model prediction, two research groups, Santini et al. and Wolfe and Elgar, performed comparative sequence analysis of multiple gene loci that are duplicated in teleosts but are present as single copies in mammals (Santini et al. 2003; Woolfe and Elgar 2007). These studies demonstrated that paralogous genes in teleost often retain differential subsets of putative regulatory elements, consistent with the notion of regulatory subfunctionalization (Santini et al. 2003; Woolfe and Elgar 2007).

Recently, Tumpel and colleagues demonstrated that changes occurring in the duplicated regulatory elements may indeed be responsible for differential domains of expression of the two co-paralogous genes (Tumpel et al. 2006). The authors identified three conserved non-coding elements in the *Hoxa2* loci that each direct expression to distinct regions of the hindbrain. Tumpel et al. also tested the activity of cognate conserved regions from duplicated *Takifugu hoxa2a* and *hoxa2b* loci in chicken and mouse embryos. The results of this study indicated that subtle changes in the *cis*-regulatory elements from the duplicated *Takifugu hoxa2* loci may account for differential expression patterns of the co-paralogous genes.

Although these findings provided functional evidence for *cis*-regulatory changes being a driving force for subfunctionalization, this study suffered from a few experimental deficiencies. For instance, the duplicated regulatory modules were tested

outside of their original environment in transgenic constructs that encompass a β-globin minimal promoter rather than the endogenous promoters of the corresponding genes. As endogenous gene promoters often contribute to tissue-specific regulatory activity, usage of the exogenous β-globin minimal promoter in the transgenic constructs might have veiled important aspects of transcriptional regulation of *hoxa2* genes. Furthermore, the differential regulatory activity of elements from the duplicated *hoxa2a* and *hoxa2b* loci was mainly inferred from examination of chicken embryos electroporated with the transgenic constructs. *In vivo* electroporation often results in uneven mosaic distribution of injected DNA. It would certainly complicate the analysis of subtle differences in tissue specificity conferred by distinct sets of *hoxa2* regulatory elements. This suboptimal experimental design, in which multiple aspects of regulation of duplicated *hoxa2* genes were bound to be missed, significantly limited resolution of this study. Therefore, despite important insights provided by this work, our understanding of the nature of the molecular changes that drive functional diversification of duplicated genes remained largely incomplete.

Here, we carried out detailed analysis of the structure and function of regulatory elements from subfunctionalized zebrafish *hoxb5a* and *hoxb5b* genes to assess if the patterns of *hoxb5* *cis*-regulatory changes are in line with the predictions of the DDC model. Towards this goal, we first identified potential regulatory elements of the *Hoxb5* genes by phylogenetic footprinting. Next, we tested the individual and collective contribution of these elements to the overlapping and unique aspects of expression of the paralogous *hoxb5* genes in transgenic mice and zebrafish.

4.1. Divergence of the non–coding regions of the zebrafish *hoxb5a* and *hoxb5b* genes

When tested in transgenic animals, large DNA regions from the duplicated *hoxb5a* and *hoxb5b* loci were able target reporter gene expression to domains of endogenous *hoxb5* expression. Indeed, the *hoxb5aegfp/lacZ* and *hoxb5begfp/lacZ* constructs were transcribed in the neural tube and somite derivatives of transgenic mice and zebrafish (Figures 3.5 - 3.7 and Figure 3.10). In addition, *hoxb5* transgenic mouse embryos showed the presence of β-galactosidase activity in the other domains associated with *Hoxb5* expression, such as the stomach, gut and adrenal glands (Hershko et al. 2003) (Figure 3.7, panels: E, H, O; Figure 3.8). Consistent with the differences detected in expression patterns of the two paralogs (Figure 3.1), the regulatory elements present within the tested region of the *hoxb5a* locus drove reporter gene expression in the somites (Figure 3.5, panels: A-F, asterisks; Figure 3.10A, arrowheads) and their derivatives, such as muscle cells in zebrafish and the vertebral cartilage condensations in mice (Figure 3.10C; Figure 3.7, panels: B, G), while the corresponding region from the *hoxb5b* locus targeted *lacZ* expression to the nodose, facio-acoustic and dorsal root ganglia and associated nerve fibres (Figure 3.5G; Figure 3.7, panels: L, M). Moreover, the rostral boundary of *lacZ* expression in the embryonic neural tube of E10.5 mouse embryos was shifted posteriorly in *hoxb5blacZ* transgenic embryo when compared to that seen in *hoxb5alacZ* transgenic embryos (Figure 3.5, compare panels B and H, arrows). These data correlate with unique expression domains described for *hoxb5a* and *hoxb5b* genes in early zebrafish development (10-24 hours, Bruce et al., 2001; Figure 3.1). Combined, these results indicate that large fragments from duplicated *hoxb5a* and *hoxb5b* loci encompassing

conserved regulatory elements near the *hoxb5* transcription units are able to recapitulate overlapping and unique aspects of expression of the corresponding genes.

4.2. Execution of complementary subfunctions of the *hoxb5a* and *hoxb5b* genes may rely on interaction between multiple *cis*-regulatory elements

Under the DDC model, preservation of complementary sets of subfunctions in duplicated gene copies occurs through complementary degeneration of their individual regulatory elements (Force et al. 1999). For this scenario to be viable, regulation of an ancestral gene is required to be modular in nature: the gene subfunctions have to be independent from each other and carried out by individual regulatory elements.

We questioned if, consistent with the model predictions, complementary degenerative changes within the individual elements are responsible for distinct regulatory activity exhibited by large fragments from the *hoxb5a* and *hoxb5b* loci. As mentioned above, J1, J2 and the 5' flanking region of *hoxb5* are conserved in both duplicated clusters in teleosts. Our phylogenetic analysis and subsequent sequence comparison revealed the presence of short DNA sequences that were conserved in cognate elements in all the mammalian and teleost loci examined. However, we were unable to detect complementary blocks of sequence conservation between paralogous elements from the *hoxb5a* and *hoxb5b* teleost clusters. Unlike other CNEs preserved in both *hoxb5* loci in teleosts, the J3 element is only retained in the *hoxb5a* locus while degenerated in *hoxb5b*. Furthermore, the results of our transgenic experiments suggested

that the loss of J3 might have contributed to portioning of the expression pattern between zebrafish *hoxb5a* and *hoxb5b* (Figures 3.5-3.7). These findings confirm that, consistent with the DDC model predictions, joint preservation of the *hoxb5a* and *hoxb5b* genes in zebrafish was instigated by complementary loss of regulatory subfunctions that occurred through changes in the non-coding regions of the genes.

To further assess the dynamics of the molecular changes that directed functional diversification of the corresponding genes, we compared the activity of the individual regulatory elements from the duplicated zebrafish and single mouse *Hoxb5* loci.

We detected differential activities of DrJ1a and DrJ1b in transgenic animals that correlate with differences in the expression of the *hoxb5* paralogs. Indeed, transgene expression driven by DrJ1b often extended more caudally than that driven by DrJ1a. Thus, about a third of transgenic zebrafish embryos for DrJ1a expressed *egfp* along the anterior-posterior axis starting from somite 1 and 2, whereas less than 10% of DrJ1b transgenic embryos express *egfp* at the level of the most anterior somites. Similarly, only *hoxb5a* is expressed in the paraxial mesoderm in somites 2 and 3, whereas *hoxb5b* expression appeared to be more extensive caudally.

Combined, our results indicate that molecular changes in the CNEs from the duplicated teleost loci account for only some of the differences in the expression patterns of *hoxb5a* and *hoxb5b* genes (Bruce et al., 2001). For example, MmJ2, DrJ2a and DrJ2b targeted transgene expression to similar domains in the mouse neural tube starting from the correct anterior border of *Hoxb4/Hoxb5* expression. In addition, when tested individually, elements from *hoxb5a* locus occasionally showed regulatory activity associated with exclusive domains of the *hoxb5b* gene. For example, the DrJ3a and DrJ1a elements occasionally drove *lacZ* expression in the dorsal root ganglia and associated

nerve fibres (Table 3.2; data not shown), an expression domain correlative with *hoxb5b* rather than *hoxb5a* expression in zebrafish.

Overall, our functional tests of individual enhancer elements were not sufficient to reveal clear complementation in regulatory activities between elements from the duplicated zebrafish *hoxb5* loci compared to that of cognate mouse elements. Indeed, when tested individually, the enhancer activities of the cognate regulatory elements from mouse and zebrafish *Hoxb5* loci overlap extensively. In contrast, regulatory activity of large regions from the *hoxb5a* and *hoxb5b* loci encompassing all the identified CNEs coincided with exclusive expression domains of the corresponding genes. It is possible that some complementary *hoxb5* subfunctions may rely on interactions between several identified regulatory elements rather than being executed by individual enhancers. Alternatively, activity of J1, J2 and J3 may be fine-tuned by other regulatory elements that are included in the *hoxb5a* and *hoxb5b* reporter constructs but have not been identified in this study.

The conserved non-coding elements found in the 3' flanking region of *Hoxb5* fall within three large DNA fragments (3 - 4.5 kb) in the *Hoxb5/Hoxb4* intergenic region that were previously shown to drive *lacZ* expression in mouse embryos with patterns recapitulating aspects of endogenous *Hoxb5* and *Hoxb4* expression (Sharpe et al. 1998). The three fragments, named LPM/PV, E, and D, encompass the J1, J2, and J3 CNEs, respectively. Our transgenic experiments demonstrated that, when tested individually, the enhancer activities of the J1, J2, and J3 represent most, if not all, of the activities of the LPM/PV, E, and D, domains (Sharpe et al. 1998). For example, domain E or MmJ2 alone both produced a strong reporter gene expression in the neural tube. Similarly, the activity of MmJ3 in the somite derivatives resembled that of domain D that was previously

referred to as the "mesodermal enhancer region" (Sharpe et al. 1998). Sharpe et al. have demonstrated that regulatory elements residing within the LPM/PV, E, and D domains are shared between the *Hoxb5* and *Hoxb4* genes. That suggests that the CNEs identified in this study may meet not be "independently mutable" and therefore may not meet some of the model requirements.

The J1, J2, and J3 CNEs also correspond to three DNA sequences recently reported by Hadrys and colleagues (2004) in their comparative analysis of the *Hoxb* clusters in mammals and teleosts. This study also suggested the existence of a novel promoter for the zebrafish *hoxb3a* gene, located ~10 kb upstream of the adjacent *hoxb4a* gene (Hadrys et al. 2004). This additional *hoxb3a* promoter would be located in the J1 sequence described here. It is therefore possible that the J1 sequence encompasses elements involved in *hoxb3a* regulation; however, we consider this unlikely. The J1 sequences are conserved (Figure 3.3) and functional (Figure 3.11; Figure 3.12) in both the *hoxba* and *hoxbb* zebrafish clusters while there is no *hoxb3* gene in the *hoxbb* cluster (Amores et al. 2004). Furthermore, our results demonstrate that the J1 sequence can drive reporter gene expression with *hoxb5*-like patterns (Figure 3.11, panels: A-C). We propose that the J1 sequence is more likely to be involved in the regulation of *hoxb5* than in that of *hoxb3*.

4.3. In addition to its enhancer activity, the J3 element may be required for proper maintenance of *hoxb5* expression

Absence of regulatory elements such as J3 in one of the two duplicated *hoxb5* loci is anticipated by the DDC model and could be the result of an unequal crossing-over event

or of the accumulation of degenerative mutations in the CNE shortly after duplication (Patel and Prince 2000). In this study, the J3 element exhibited enhancer activity recapitulating aspects of *Hoxb5* expression in transgenic animals. Sharpe and colleagues have previously referred to a large part of the mouse *Hoxb5/Hoxb4* intergenic region encompassing the J3 element as a "mesodermal enhancer region" (Sharpe et al., 1998). Our data suggest that the mesodermal domain of *Hoxb5* expression is, at least in part, attributed to the presence of the J3 element. While expression in the neural tube was present in all the embryos transgenic for the *hoxb5alacZ, hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* constructs, transgene expression in the somites and their derivatives was detected consistently only in *hoxb5alacZ* transgenics. In contrast, reporter gene expression in the "mesodermal" domain of *Hoxb5* expression was not detected in *hoxb5blacZ* transgenic embryos and appeared to be significantly diminished in *hoxb5a*ΔJ3*lacZ* transgenic embryos (Figure 3.6, compare panels A-F to M-R, asterisks). Similarly, *hoxb5aegfp* was clearly expressed in the zebrafish neural tube and somites, where zebrafish *hoxb5a* was shown to be expressed (Figure 3.10, panels: A-C), while the *hoxb5a*ΔJ3*egfp* and *hoxb5begfp* transcripts were mainly found in cells of neuronal origin (Figure 3.10, panels D-G; data not shown). Therefore the J3 element is responsible for the observed differences between the *hoxb5* transgenic embryos as this element is only present in constructs carrying the endogenous *hoxb5a* locus (*hoxb5alacZ* and *hoxb5aegfp*). Consistent with this observation, only *hoxb5a* was found to be expressed in the somites of developing zebrafish embryos (Bruce et al. 2001).

Our transgenic experiments revealed that the J3 element is not only responsible for the mesodermal domain of *Hoxb5* expression but may also be required for maintaining high levels of *hoxb5* expression. Indeed, data obtained from mouse transgenesis suggests

the reporter constructs encompassing J3 (*hoxb5a* and *hoxb5b*insJ3) drove, generally, higher levels of transgene expression than those encompassing their J3-lacking counterparts (*hoxb5b* and *hoxb5a*ΔJ3) (Figure 3.5; Figure 3.6).

Consistent with the data obtained from mouse transgenesis, embryos injected with the *hoxb5aegfp* construct showed higher levels of *egfp* expression than ones injected with the *hoxb5begfp* construct (Figure 3.10, compare panels A-C to D, E). The low levels of reporter gene expression driven by the *hoxb5b* locus is likely associated with the degeneration of the J3 element as the deletion of the J3 element from the *hoxb5a* locus caused a similar effect (Figure 3.10, panels: F, G). Accordingly, insertion of J3 in the *hoxb5b* loci increased levels of reporter gene expression and appears to at least partially rescue *hoxb5a*-like regulatory activity (Figure 3.10, panels: H, I). These observations have been further confirmed by assessment of relative *egfp* transcript levels in primary zebrafish embryos injected with *hoxb5aegfp*, *hoxb5begfp*, *hoxb5a*ΔJ3*egfp* and *hoxb5b*insJ3 by quantitative real-time RT-PCR (Figure 3.10, panels J, K). Taken together, these results suggest that the pattern of reporter gene expression driven by the *hoxb5b* and *hoxb5a*ΔJ3 sequences reflect their inability to maintain high *hoxb5* transcription in the absence of the J3 element. We hypothesize that the presence of J3 is necessary for proper rates of *hoxb5* transcription.

An intriguing finding of this study was that, while *hoxb5alacZ* transgenic embryos exhibit a stable and consistent pattern of reporter gene expression, embryos obtained from *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* (lines 7603, 7665, 7629, 7616) males showed variegated patterns of transgene expression that differ between embryos collected from the same mating (Figure 3.9). Staining patterns of the primary transgenic embryos also showed such variability. We have looked for several possible explanations for this phenomenon.

First, we considered the possibility that the observed phenotype could be due to late integration of a transgene into the mouse genome. Late integration of a transgene into the mouse genome is known to generate "mosaic transgenics" that carry a transgene only in a subset of tissues that may result in a partial pattern of transgene expression. However, the variegated expression pattern seen in *hoxb5blacZ* and *hoxb5a*ΔJ3*lacZ* transgenic animals cannot be associated with mosaicism as this pattern is meiotically transmittable and is maintained in the progeny (F1 and F2) of transgenic founders.

In some cases, variable or partial expression of a transgene can be attributed to the regulatory influences at the integration site. However, we propose this is not a parsimonious explanation of our results. Indeed, similar variations in patterns of *lacZ* expression were observed in two independent lines and two founders transgenic for *hoxb5blacZ,* and two lines and one founder transgenic for the *hoxb5a*ΔJ3*lacZ* construct. As each instance represents an event of independent transgene insertion into a random genomic location, we propose that the data likely reflect intrinsic regulatory activity of the tested *hoxb5* loci.

The Polycomb (Pc) and Trithorax (Trx) proteins are known regulators of transcriptional states of *Hox* genes. It is thought that Pc and Trx maintain transcriptional status of target genes though modification of chromatin structure. These chromatin proteins share a SET domain that mediates lysine-directed methylation (Rea et al., 2000). A recent study has described generation and characterization of mutant mice homozygous for a SET domain truncated allele of *Mll*, the mammalian homologue of the *Drosophila trithorax* gene. ΔSET mutant mice exhibited an alteration in the maintenance of proper transcription of several *Hox* genes during development (Terranova et al., 2006). *Hox* transcription appeared to have initiated but failed to be maintained in knockout mice.

The variegated patterns of *lacZ* expression seen in *hoxb5blacZ* and *hoxb5a*Δ*J3lacZ* are similar to those observed in ΔSET mice and may be a result of distortion in epigenetic mechanisms of *Hoxb5* regulation. Thus, the J3 element, in addition to its enhancer activity, may be important for maintenance of open chromatin structure that is necessary for the proper maintenance of *hoxb5* expression. It is possible that reporter gene expression may have been correctly initiated in *hoxb5blacZ* and *hoxb5a*Δ*J3lacZ* transgenics but was abrogated at different time points by unknown factors. We hypothesize that the J3 element may harbour transcription factor binding sites for chromatin modification proteins, such as trithorax, that are involved in maintenance of optimal gene transcription (Brock and Fisher 2005).

Combined, our results indicate that the changes that occurred in the non-coding regions of the *hoxb5a* and *hoxb5b* genes after duplication are not only responsible for splitting of the ancestal *Hoxb5* expression pattern between the duplicates, but they may have also resulted in differences in general levels of *hoxb5a* and *hoxb5b* expression. Mouse transgenesis experiments demonstrated that reporter gene expression driven by *hoxb5b* regulatory elements, in general, occurs at seemingly lower levels when compared to that driven by the corresponding regions from the *hoxb5a* locus. It is possible that proper expression and function of the *hoxb5b* duplicate in its exclusive domains did not require the presence of the J3 element. Alternatively, regulatory activity of the J3 element may be necessary for proper function of other *cis*-regulatory elements in both the *hoxb5a* and *hoxb5b* loci and its degeneration in the *hoxb5b* locus may have put the *hoxb5b* duplicate on its way to nonfunctionalization. We hypothesize that, following the duplication event, the *hoxb5a* duplicate retained the most essential functions of the

ancestal *Hoxb5* gene while functions preserved by *hoxb5b* were less vital, instigating degeneration of the J3 element in the *hoxb5b* locus.

4.4. Dynamics of changes in the *cis*-acting regulatory elements of *hoxb5* duplicates are more complex than predicted by the DDC model

Our phylogenetic footprinting data demonstrated that while the J3 element has only been retained in the *hoxb5a* loci, J1, J2 and the immediate 5' flanking region of the *hoxb5* genes remained preserved in both teleost *hoxb5* loci. Moreover, functional tests of cognate regulatory elements from duplicated *hoxb5a* and *hoxb5b* zebrafish loci did not reveal clear differences in their activity. Interestingly, in their recent comparative genomic study, Woolfe and Elgar identified other duplicated genes, such as *pax1* co-orthologs in *Fugu*, that carry unportioned sets of putative regulatory elements (Woolfe and Elgar 2007). The preservation of redundant regulatory elements in both duplicated gene loci is surprising since the DDC model states that a complete degeneration of redundant subfunctions in the duplicates (resolution) should occur soon after the duplication event (Force et al. 1999).

Our data suggest that, while the DDC model assumes that regulatory elements have modular structure and act as individual subfunctions, J1, J2 and J3 elements may be inter-related and each involved in execution of more than one regulatory *hoxb5* subfunction.

The results of our phylogenetic analysis indicate that functional modules of duplicated *hoxb5* loci may diverge at a slower evolutionary rate than is anticipated by the DDC model. Depicted deviations from the predicted rate of divergence were reported for the duplicated *sox9a* and *sox9b* genes of teleost (Cresko et al., 2003). Differences in the expression patterns of *sox9a* and *sox9b* in zebrafish and stickleback led to the conclusion that, even though partitioning of most *sox9* subfunctions occurred before the divergence of the teleost lineages, some gene subfunctions may have sorted differently in the two teleost species. Similarly, our phylogenetic analyses suggest that orthologous functional modules of duplicated *hoxb5* genes in *Takifugu* and zebrafish are separated by a larger evolutionary distance than is expected if subfunctionalization happened before the divergence of the species (Figure 3.13). We propose that some aspects of functional diversification of the *hoxb5a* and *hoxb5b* genes might have occurred independently in the zebrafish and *Takifugu* lineages similar to that reported for the *sox9a* and *sox9b* genes. Independent partitioning of ancestral gene subfunctions is predicted to contribute to reproductive isolation and enhance speciation and would constitute a possible evolutionary advantage conferred by subfunctionalization (Lynch and Force, 2000).

5. Conclusions

The results of our transgenic experiments demonstrate that large fragments from the duplicated *hoxb5a* and *hoxb5b* loci are able to recapitulate overlapping and unique aspects of expression of the corresponding genes. This suggests that functional diversification of the *hoxb5a* and *hoxb5b* genes in zebrafish occurred by means of changes in the regulatory elements of these genes.

Our data suggest that patterns of changes in *cis*-acting regulatory elements of the *hoxb5* duplicates in teleosts may be more complex than ones taken into consideration by the DDC model (Force et al. 1999). Indeed, while the model predicts complete portioning of redundant regulatory elements between the gene duplicates, most of the identified CNEs, J1, J2 and the 5' flanking region were preserved in both *hoxb5* loci after the duplication event. Furthermore, phylogenetic filtering and functional tests of individual elements from the *hoxb5a* and *hoxb5b* loci did not reveal clear signs of complementation between the regulatory elements retained in both zebrafish loci. Consistently, the results of the phylogenetic analysis demonstrated that regulatory sequences from these duplicated loci diverged slower than anticipated by the DDC model, suggesting that the *hoxb5* regulatory elements are under stronger regulatory constraints than predicted by the DDC model.

It has to be considered that while the model requires regulatory elements to be "independently mutable", previous study indicated that the identified CNEs may be shared between *Hoxb5* and *Hoxb4* (Sharpe et al. 1998). Moreover, our results suggest that execution of complementary *hoxb5a* and *hoxb5b* subfunctions may rely on interaction between multiple regulatory elements. When tested in transgenic animals, large DNA

regions encompassing multiple regulatory elements from the *hoxb5a* and *hoxb5b* loci exhibited differential regulatory activity that correlate with the distinct expression domains of the corresponding genes while constructs containing individual elements from the duplicated loci failed to do so. Moreover, individual regulatory elements may be involved in the execution of more than one regulatory *hoxb5* subfunction. Indeed, the J3 element, in addition to its enhancer activity, may be involved in other regulatory mechanisms, such as maintenance of an appropriate chromatin structure that is necessary for proper maintenance of *hoxb5* expression.

6. References

Aburomia R, Khaner O, Sidow A (2003) Functional evolution in the ancestral lineage of vertebrates or when genomic complexity was wagging its morphological tail. J Struct Funct Genomics 3(1-4): 45-52.

Akin ZN, Nazarali AJ (2005) Hox genes and their candidate downstream targets in the developing central nervous system. Cell Mol Neurobiol 25(3-4): 697-741.

Altschmied J, Delfgaauw J, Wilde B, Duschl J, Bouneau L et al. (2002) Subfunctionalization of duplicate mitf genes associated with differential degeneration of alternative exons in fish. Genetics 161(1): 259-267.

Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A et al. (2004) Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish. Genome Res 14(1): 1-10.

Amores A, Force A, Yan YL, Joly L, Amemiya C et al. (1998) Zebrafish hox clusters and vertebrate genome evolution. Science 282(5394): 1711-1714.

Andrade RP, Palmeirim I, Bajanca F (2007) Molecular clocks underlying vertebrate embryo segmentation: A 10-year-old hairy-go-round. Birth Defects Res C Embryo Today 81(2): 65-83.

Bajusz I, Sipos L, Gyorgypal Z, Carrington EA, Jones RS et al. (2001) The Trithorax-mimic allele of Enhancer of zeste renders active domains of target genes accessible to polycomb-group-dependent silencing in Drosophila melanogaster. Genetics 159(3): 1135-1150.

Brend T, Gilthorpe J, Summerbell D, Rigby PW (2003) Multiple levels of transcriptional and post-transcriptional regulation are required to define the domain of Hoxb4 expression. Development 130(12): 2717-2728.

Brock HW, Fisher CL (2005) Maintenance of gene expression patterns. Dev Dyn 232(3): 633-655.

Bruce AE, Oates AC, Prince VE, Ho RK (2001) Additional hox clusters in the zebrafish: divergent expression patterns belie equivalent activities of duplicate hoxB5 genes. Evol Dev 3(3): 127-144.

Bull JJ, Springman R, Molineux IJ (2007) Compensatory evolution in response to a novel RNA polymerase: orthologous replacement of a central network gene. Mol Biol Evol 24(4): 900-908.

Burke AC, Nelson CE, Morgan BA, Tabin C (1995) Hox genes and the evolution of vertebrate axial morphology. Development 121(2): 333-346.

Busturia A, Morata G (1988) Ectopic expression of homeotic genes caused by the elimination of the Polycomb gene in Drosophila imaginal epidermis. Development 104(4): 713-720.

Castronovo V, Kusaka M, Chariot A, Gielen J, Sobel M (1994) Homeobox genes: potential candidates for the transcriptional control of the transformed and invasive phenotype. Biochem Pharmacol 47(1): 137-143.

Catron KM, Iler N, Abate C (1993) Nucleotides flanking a conserved TAAT core dictate the DNA binding specificity of three murine homeodomain proteins. Mol Cell Biol 13(4): 2354-2365.

Chambeyron S, Bickmore WA (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. Genes Dev 18(10): 1119-1130.

Chisaka O, Capecchi MR (1991) Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene hox-1.5. Nature 350(6318): 473-479.

Chiu CH, Dewar K, Wagner GP, Takahashi K, Ruddle F et al. (2004) Bichir HoxA cluster sequence reveals surprising trends in ray-finned fish genomic evolution. Genome Res 14(1): 11-17.

Ciruna B, Rossant J (2001) FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. Dev Cell 1(1): 37-49.

Cooper DN, Youssoufian H (1988) The CpG dinucleotide and human genetic disease. Hum Genet 78(2): 151-155.

Cooper DN, M. Krawczak , Antonorakis SE (1995) The nature and mechanisms of human gene mutation. Metabolic and Molecular Bases of Inherited Disease 7: 259–291.

Copeland NG, Jenkins NA, Court DL (2001) Recombineering: a powerful new tool for mouse functional genomics. Nat Rev Genet 2(10): 769-779.

Cormack BP, Valdivia RH, Falkow S (1996) FACS-optimized mutants of the green fluorescent protein (GFP). Gene 173(1 Spec No): 33-38.

Delaval K, Feil R (2004) Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev 14(2): 188-195.

Deschamps J, van Nes J (2005) Developmental regulation of the Hox genes during axial morphogenesis in the mouse. Development 132(13): 2931-2942.

Dockter J, Ordahl CP (2000) Dorsoventral axis determination in the somite: a re-examination. Development 127(10): 2201-2206.

Dockter JL, Ordahl CP (1998) Determination of sclerotome to the cartilage fate. Development 125(11): 2113-2124.

Dolle P, Dierich A, LeMeur M, Schimmang T, Schuhbaur B et al. (1993) Disruption of the Hoxd-13 gene induces localized heterochrony leading to mice with neotenic limbs. Cell 75(3): 431-441.

Duboule D, Deschamps J (2004) Colinearity loops out. Dev Cell 6(6): 738-740.

Favier B, Dolle P (1997) Developmental functions of mammalian Hox genes. Mol Hum Reprod 3(2): 115-131.

Fienberg AA, Utset MF, Bogarad LD, Hart CP, Awgulewitsch A et al. (1987) Homeo box genes in murine development. Curr Top Dev Biol 23: 233-256.

Fisher RA (1935) The sheltering of lethal. Am Nat 69: 446-455.

Force A, Lynch M, Pickett FB, Amores A, Yan YL et al. (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics 151(4): 1531-1545.

Forlani S, Lawson KA, Deschamps J (2003) Acquisition of Hox codes during gastrulation and axial elongation in the mouse embryo. Development 130(16): 3807-3819.

Fraidenraich D, Lang R, Basilico C (1998) Distinct regulatory elements govern Fgf4 gene expression in the mouse blastocyst, myotomes, and developing limb. Dev Biol 204(1): 197-209.

Gavalas A, Trainor P, Ariza-McNaughton L, Krumlauf R (2001) Synergy between Hoxa1 and Hoxb1: the relationship between arch patterning and the generation of cranial neural crest. Development 128(15): 3017-3027.

Gehring WJ, Hiromi Y (1986) Homeotic genes and the homeobox. Annu Rev Genet 20: 147-173.

Gilbert (2000) Developmental Biology

Gogarten JP, Olendzenski L (1999) Orthologs, paralogs and genome comparisons. Curr Opin Genet Dev 9(6): 630-636.

Gould A (1997) Functions of mammalian Polycomb group and trithorax group related genes. Curr Opin Genet Dev 7(4): 488-494.

Gould A, Itasaki N, Krumlauf R (1998) Initiation of rhombomeric Hoxb4 expression requires induction by somites and a retinoid pathway. Neuron 21(1): 39-51.

Hadrys T, Prince V, Hunter M, Baker R, Rinkwitz S (2004) Comparative genomic analysis of vertebrate Hox3 and Hox4 genes. J Exp Zoolog B Mol Dev Evol 302(2): 147-164.

Hadrys T, Punnamoottil B, Pieper M, Kikuta H, Pezeron G et al. (2006) Conserved co-regulation and promoter sharing of hoxb3a and hoxb4a in zebrafish. Dev Biol 297(1): 26-43.

Haldane JBS (1933) The part played by recurrent mutation in evolution. Am Nat 67: 5-9.

Hartmann C, Corre-Menguy F, Boualem A, Jovanovic M, Lelandais-Briere C (2004) [MicroRNAs: a new class of gene expression regulators]. Med Sci (Paris) 20(10): 894-898.

Heard E, Disteche CM (2006) Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes Dev 20(14): 1848-1867.

Herault Y, Beckers J, Kondo T, Fraudeau N, Duboule D (1998) Genetic analysis of a Hoxd-12 regulatory element reveals global versus local modes of controls in the HoxD complex. Development 125(9): 1669-1677.

Hershko AY, Kafri T, Fainsod A, Razin A (2003) Methylation of HoxA5 and HoxB5 and its relevance to expression during mouse development. Gene 302(1-2): 65-72.

Hoegg S, Meyer A (2005) Hox clusters as models for vertebrate genome evolution. Trends Genet 21(8): 421-424.

Hoegg S, Brinkmann H, Taylor JS, Meyer A (2004) Phylogenetic timing of the fish-specific genome duplication correlates with the diversification of teleost fish. J Mol Evol 59(2): 190-203.

Hoey T, Levine M (1988) Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature 332(6167): 858-861.

Hoey T, Warrior R, Manak J, Levine M (1988) DNA-binding activities of the Drosophila melanogaster even-skipped protein are mediated by its homeo domain and influenced by protein context. Mol Cell Biol 8(11): 4598-4607.

Hogan B, Constantini, F., Lacy, E. (1986) Manipulating the mouse embryo: a laboratory manual. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory.

Hogan BL, Holland PW, Lumsden A (1988) Expression of the homeobox gene, Hox 2.1, during mouse embryogenesis. Cell Differ Dev 25 Suppl: 39-44.

Holland PW, Garcia-Fernandez J, Williams NA, Sidow A (1994) Gene duplications and the origins of vertebrate development. Dev Suppl: 125-133.

Holloway DT, Kon M, DeLisi C (2005) Integrating genomic data to predict transcription factor binding. Genome Inform 16(1): 83-94.

Hsu K, Look AT (2003) Turning on a dimer: new insights into MLL chimeras. Cancer Cell 4(2): 81-83.

Iimura T, Pourquie O (2006) Collinear activation of Hoxb genes during gastrulation is linked to mesoderm cell ingression. Nature 442(7102): 568-571.

Jaffe L, Jeannotte L, Bikoff EK, Robertson EJ (1990) Analysis of beta 2-microglobulin gene expression in the developing mouse embryo and placenta. J Immunol 145(10): 3474-3482.

Kappen C, Ruddle FH (1993) Evolution of a regulatory gene family: HOM/HOX genes. Curr Opin Genet Dev 3(6): 931-938.

Kato N, Aoyama H (1998) Dermomyotomal origin of the ribs as revealed by extirpation and transplantation experiments in chick and quail embryos. Development 125(17): 3437-3443.

Kidwell MG, Lisch D (1997) Transposable elements as sources of variation in animals and plants. Proc Natl Acad Sci U S A 94(15): 7704-7711.

Kiecker C, Lumsden A (2005) Compartments and their boundaries in vertebrate brain development. Nat Rev Neurosci 6(7): 553-564.

Kim C, Nielsen HC (2000) Hoxa-5 in mouse developing lung: cell-specific expression and retinoic acid regulation. Am J Physiol Lung Cell Mol Physiol 279(5): L863-871.

Kluver N, Kondo M, Herpin A, Mitani H, Schartl M (2005) Divergent expression patterns of Sox9 duplicates in teleosts indicate a lineage specific subfunctionalization. Dev Genes Evol 215(6): 297-305.

Kondo T, Zakany J, Duboule D (1998) Control of colinearity in AbdB genes of the mouse HoxD complex. Mol Cell 1(2): 289-300.

Krumlauf R (1994) Hox genes in vertebrate development. Cell 78(2): 191-201.

Krumlauf R, Holland PW, McVey JH, Hogan BL (1987) Developmental and spatial patterns of expression of the mouse homeobox gene, Hox 2.1. Development 99(4): 603-617.

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17(12): 1244-1245.

Lappin TR, Grier DG, Thompson A, Halliday HL (2006) HOX genes: seductive science, mysterious mechanisms. Ulster Med J 75(1): 23-31.

Larochelle C, Tremblay M, Bernier D, Aubin J, Jeannotte L (1999) Multiple cis-acting regulatory regions are required for restricted spatio-temporal Hoxa5 gene expression. Dev Dyn 214(2): 127-140.

Le Mouellic H, Lallemand Y, Brulet P (1992) Homeosis in the mouse induced by a null mutation in the Hox-3.1 gene. Cell 69(2): 251-264.

Levine A, Cantoni GL, Razin A (1992) Methylation in the preinitiation domain suppresses gene transcription by an indirect mechanism. Proc Natl Acad Sci U S A 89(21): 10119-10123.

Lewin B (2000) Homeodomains bind related targets in DNA.

Oxford: Oxford University Press In Genes VII

660-662.

Lewis EB (1978) A gene complex controlling segmentation in Drosophila. Nature 276(5688): 565-570.

Li W-H (1980) Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fishes. Genetics 95: 237-258.

Lohnes D (2003) The Cdx1 homeodomain protein: an integrator of posterior signaling in the mouse. Bioessays 25(10): 971-980.

Loots GG, Locksley RM, Blankespoor CM, Wang ZE, Miller W et al. (2000) Identification of a coordinate regulator of interleukins 4, 13, and 5 by cross-species sequence comparisons. Science 288(5463): 136-140.

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154(1): 459-473.

Mann RS (1997) Why are Hox genes clustered? Bioessays 19(8): 661-664.

Mann RS, Affolter M (1998) Hox proteins meet more partners. Curr Opin Genet Dev 8(4): 423-429.

Manna D, Breier AM, Higgins NP (2004) Microarray analysis of transposition targets in Escherichia coli: the impact of transcription. Proc Natl Acad Sci U S A 101(26): 9780-9785.

Mansfield JH, Harfe BD, Nissen R, Obenauer J, Srineel J et al. (2004) MicroRNA-responsive 'sensor' transgenes uncover Hox-like and other developmentally regulated patterns of vertebrate microRNA expression. Nat Genet 36(10): 1079-1083.

Marin I, Baker BS (1998) The evolutionary dynamics of sex determination. Science 281(5385): 1990-1994.

Mark M, Rijli FM, Chambon P (1997) Homeobox genes in embryogenesis and pathogenesis. Pediatr Res 42(4): 421-429.

Marshall H, Studer M, Popperl H, Aparicio S, Kuroiwa A et al. (1994) A conserved retinoic acid response element required for early expression of the homeobox gene Hoxb-1. Nature 370(6490): 567-571.

McGinnis W, Krumlauf R (1992) Homeobox genes and axial patterning. Cell 68(2): 283-302.

Meyer A, Malaga-Trillo E (1999) Vertebrate genomics: More fishy tales about Hox genes. Curr Biol 9(6): R210-213.

Moyle WR, Campbell RK, Myers RV, Bernard MP, Han Y et al. (1994) Co-evolution of ligand-receptor pairs. Nature 368(6468): 251-255.

Nelson CE, Morgan BA, Burke AC, Laufer E, DiMambro E et al. (1996) Analysis of Hox gene expression in the chick limb bud. Development 122(5): 1449-1466.

Nijhout HF, Emlen DJ (1998) Competition among body parts in the development and evolution of insect morphology. Proc Natl Acad Sci U S A 95(7): 3685-3689.

Ohno S (1970) Evolution by Gene Duplication. Springer-Verlag, Heidelberg, Germany.

Oosterveen T, Niederreither K, Dolle P, Chambon P, Meijlink F et al. (2003) Retinoids regulate the anterior expression boundaries of 5' Hoxb genes in posterior hindbrain. Embo J 22(2): 262-269.

Ordahl CP, Williams BA (1998) Knowing chops from chuck: roasting myoD redundancy. Bioessays 20(5): 357-362.

Palmeirim I (1999) [Segmentation in vertebrates: a molecular clock linked to periodic somite formation]. J Soc Biol 193(3): 243-256.

Papalopulu N, Lovell-Badge R, Krumlauf R (1991) The expression of murine Hox-2 genes is dependent on the differentiation pathway and displays a collinear sensitivity to retinoic acid in F9 cells and Xenopus embryos. Nucleic Acids Res 19(20): 5497-5506.

Patel NH, Prince VE (2000) Beyond the Hox complex. Genome Biol 1(5): REVIEWS1027.

Pellerin I, Schnabel C, Catron KM, Abate C (1994) Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster. Mol Cell Biol 14(7): 4532-4545.

Pendleton JW, Nagai BK, Murtha MT, Ruddle FH (1993) Expansion of the Hox gene family and the evolution of chordates. Proc Natl Acad Sci U S A 90(13): 6300-6304.

Piatigorsky J, Wistow G (1991) The recruitment of crystallins: new functions precede gene duplication. Science 252(5010): 1078-1079.

Pirrotta V (1997) PcG complexes and chromatin silencing. Curr Opin Genet Dev 7(2): 249-258.

Pollard KS, Salama SR, King B, Kern AD, Dreszer T et al. (2006) Forces shaping the fastest evolving regions in the human genome. PLoS Genet 2(10): e168.

Prince V (1998) Hox Genes and Segmental Patterning of the Vertebrate Hindbrain. AMER ZOOL 38: 634-646.

Prince V (2002) The Hox Paradox: More complex(es) than imagined. Dev Biol 249(1): 1-15.

Prince VE, Pickett FB (2002) Splitting pairs: the diverging fates of duplicated genes. Nat Rev Genet 3(11): 827-837.

Prince VE, Joly L, Ekker M, Ho RK (1998) Zebrafish hox genes: genomic organization and modified colinear expression patterns in the trunk. Development 125(3): 407-420.

Rauskolb C, Wieschaus E (1994) Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins. Embo J 13(15): 3561-3569.

Razin A, Cedar H (1991) DNA methylation and gene expression. Microbiol Rev 55(3): 451-458.

Rieckhof GE, Casares F, Ryoo HD, Abu-Shaar M, Mann RS (1997) Nuclear translocation of extradenticle requires homothorax, which encodes an extradenticle-related homeodomain protein. Cell 91(2): 171-183.

Ruddle FH, Bentley KL, Murtha MT, Risch N (1994) Gene loss and gain in the evolution of the vertebrates. Dev Suppl: 155-161.

Sakach M, Safaei R (1996) Localization of the HoxB5 protein in the developing CNS of late gestational mouse embryos. Int J Dev Neurosci 14(5): 567-573.

Sanogo YO, Dobson SL, Bordenstein SR, Novak RJ (2007) Disruption of the Wolbachia surface protein gene wspB by a transposable element in mosquitoes of the Culex pipiens complex (Diptera, Culicidae). Insect Mol Biol 16(2): 143-154.

Santini S, Boore JL, Meyer A (2003) Evolutionary conservation of regulatory elements in vertebrate Hox gene clusters. Genome Res 13(6A): 1111-1122.

Schwartz YB, Pirrotta V (2007) Polycomb silencing mechanisms and the management of genomic programmes. Nat Rev Genet 8(1): 9-22.

Scott MP (1992) Vertebrate homeobox gene nomenclature. Cell 71(4): 551-553.

Scott MP, Tamkun JW, Hartzell GW, 3rd (1989) The structure and function of the homeodomain. Biochim Biophys Acta 989(1): 25-48.

Semon M, Wolfe KH (2007) Reciprocal gene loss between Tetraodon and zebrafish after whole genome duplication in their ancestor. Trends Genet 23(3): 108-112.

Sharpe J, Nonchev S, Gould A, Whiting J, Krumlauf R (1998) Selectivity, sharing and competitive interactions in the regulation of Hoxb genes. Embo J 17(6): 1788-1798.

Small KM, Potter SS (1993) Homeotic transformations and limb defects in Hox A11 mutant mice. Genes Dev 7(12A): 2318-2328.

Spitz F, Gonzalez F, Duboule D (2003) A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell 113(3): 405-417.

Studer M, Gavalas A, Marshall H, Ariza-McNaughton L, Rijli FM et al. (1998) Genetic interactions between Hoxa1 and Hoxb1 reveal new roles in regulation of early hindbrain patterning. Development 125(6): 1025-1036.

Suemori H, Noguchi S (2000) Hox C cluster genes are dispensable for overall body plan of mouse embryonic development. Dev Biol 220(2): 333-342.

Taylor JS, Van de Peer Y, Braasch I, Meyer A (2001) Comparative genomics provides evidence for an ancient genome duplication event in fish. Philos Trans R Soc Lond B Biol Sci 356(1414): 1661-1679.

Terranova R, Agherbi H, Boned A, Meresse S, Djabali M (2006) Histone and DNA methylation defects at Hox genes in mice expressing a SET domain-truncated form of Mll. Proc Natl Acad Sci U S A 103(17): 6629-6634.

Thisse C, Thisse B, Schilling TF, Postlethwait JH (1993) Structure of the zebrafish snail1 gene and its expression in wild-type, spadetail and no tail mutant embryos. Development 119(4): 1203-1215.

Tumpel S, Cambronero F, Wiedemann LM, Krumlauf R (2006) Evolution of cis elements in the differential expression of two Hoxa2 coparalogous genes in pufferfish (Takifugu rubripes). Proc Natl Acad Sci U S A 103(14): 5419-5424.

Turner FR, Mahowald AP (1979) Scanning electron microscopy of Drosophila melanogaster embryogenesis. III. Formation of the head and caudal segments. Dev Biol 68(1): 96-109.

Tvrdik P, Capecchi MR (2006) Reversal of Hox1 gene subfunctionalization in the mouse. Dev Cell 11(2): 239-250.

van den Akker E, Fromental-Ramain C, de Graaff W, Le Mouellic H, Brulet P et al. (2001) Axial skeletal patterning in mice lacking all paralogous group 8 Hox genes. Development 128(10): 1911-1921.

van der Hoeven F, Sordino P, Fraudeau N, Izpisua-Belmonte JC, Duboule D (1996) Teleost HoxD and HoxA genes: comparison with tetrapods and functional evolution of the HOXD complex. Mech Dev 54(1): 9-21.

Wagner GP, Amemiya C, Ruddle F (2003) Hox cluster duplications and the opportunity for evolutionary novelties. Proc Natl Acad Sci U S A 100(25): 14603-14606.

Watterson GA (1983) On the time for gene silencing at duplicate loci. Genetics 105: 745-766.

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

Whalen AM, Steward R (1993) Dissociation of the dorsal-cactus complex and phosphorylation of the dorsal protein correlate with the nuclear localization of dorsal. J Cell Biol 123(3): 523-534.

Wing LY, Chuang PC, Wu MH, Chen HM, Tsai SJ (2003) Expression and mitogenic effect of fibroblast growth factor-9 in human endometriotic implant is regulated by aberrant production of estrogen. J Clin Endocrinol Metab 88(11): 5547-5554.

Wittkopp PJ (2006) Evolution of cis-regulatory sequence and function in Diptera. Heredity 97(3): 139-147.

Wolpert L (1998) Principles of Development.

Woolfe A, Elgar G (2007) Comparative genomics using Fugu reveals insights into regulatory subfunctionalization. Genome Biol 8(4): R53.

Wright CV, Cho KW, Oliver G, De Robertis EM (1989) Vertebrate homeodomain proteins: families of region-specific transcription factors. Trends Biochem Sci 14(2): 52-56.

Yee SP, Rigby PW (1993) The regulation of myogenin gene expression during the embryonic development of the mouse. Genes Dev 7(7A): 1277-1289.

Zerucha T, Stuhmer T, Hatch G, Park BK, Long Q 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(2): 709-721.